



HAL
open science

Déconvolution de signaux vocaux

Francis Goulamhousen Satsou

► **To cite this version:**

Francis Goulamhousen Satsou. Déconvolution de signaux vocaux. Traitement du signal et de l'image [eess.SP]. 2017. hal-01627891

HAL Id: hal-01627891

<https://inria.hal.science/hal-01627891>

Submitted on 2 Nov 2017

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

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

Déconvolution de signaux vocaux

Stage de l'Université de Bordeaux
(spécialité Traitement du signal)

par

GOULAMHOUSSEN SATSOU Francis

Encadrant :

M. DAOUDI Khalid

Directeur de recherche à INRIA

Déconvolution de signaux vocaux

Stage de Master 1

Goulamhousen Satsou Francis

Stage de Traitement de Signal
Septembre 2017

ABSTRACT

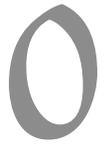
La déconvolution est une technique largement étudiée et utilisée en traitement du signal lorsqu'il s'agit d'estimer un signal à partir d'observations filtrées. Dans le cas de la parole, un problème classique est d'estimer le signal source (flux glottal) à partir du signal acoustique. L'objectif du stage est d'étudier et tester une méthode de déconvolution

Mots clés : Déconvolution, flux glottal.

INRIA Bordeaux sud Ouest
200, Avenue de la Vieille Tour, 33405 Talence

CONTENTS

0	INTRODUCTION	1
0.1	Résumé du stage	1
0.2	Introduction	2
1	MODÈLE LF	7
1.1	Modèle Flux Glottal	7
1.2	Modèle LF	8
1.3	Model LF code-1 sous matlab	10
1.4	Soft-matlab basé sur le model LF	11
2	ESTIMATION SIGNAL SOURCE	15
2.1	Enjeux et Applications	15
2.2	Closed phase inverse filtering	17
2.3	Iterative and inverse filtering	18
2.4	Mixed-phase decomposition	18
2.4.1	Description de la Méthode de Drugman	18
2.4.2	Description du code	19
3	EXPERIMENTATION	23
3.1	Génération du signal source	23
3.1.1	Génération de flux avec Code-1 matlab	24
3.1.2	Génération de flux avec Code-2 soft	24
3.2	Estimation sur une trame	25
3.3	Estimation sur une voyelle soutenu	26
4	CONCLUSION	35



INTRODUCTION

0.1 RÉSUMÉ DU STAGE

Lors de ce stage, nous avons étudié une méthode de déconvolution pour les signaux vocaux. la méthode en question estime un signal source[3]. Cette source a été convoluer par un filtre. Ce filtre représente le conduit vocal. Donc la méthode étudié propose de retrouver le signal source.

L'enjeu du stage était de comprendre la méthode, tester ses limites. On peu découper ce stage en 3 parties. il n'y a aucune chronologie entres ses parties.

PARTIE 1

La premiers fut l'étude bibliographique de la discipline qu'est l'estimation du signal source(Estimation Glottal). Cette étude bibliographie c'est fait sur deux socles : [1] et [2]. Puis nous avons pris soins d'étudier la source et sa modélisation. Et spécialement le Model LF du flux dérivative[4][6][7][9][8]. Nous avons aussi vu une façon de synthétiser la parole humain avec le modèle source-filtre

PARTIE 2

Le second partie fut l'étude de la méthode. Nous avons décortiqué et étudier les principes. Puis nous avons réfléchi à comment on peut évaluer sa méthodes.

PARTIE 3

Cette partie consistait à tester la méthodes. Nous avons construit les données pour tester la méthode. Pour tester la méthode on a essayé de générer de la parole partir une source connus. Puis on a comparer la source estimer avec la vrai source.

OUTIL

Les testes et les méthodes utilisés pendant ce stage on été fait et élaboré dans un environnement matlab sous un système d'exploitation *linux Ubuntu 16.04*

ENVIRONNEMENT DE TRAVAIL

Ce stage c'est déroulé à L'Inria de bordeaux Dans l'équipe GEOSTAT.L'Inria est une institut de recherche en sciences appliquées. Son domaine d'application est l'informatique et l'automatique. C'est une institue implantée partout en France. l'institut de bordeaux à été construit en 2008.

L'ÉQUIPE GÉOSTAT

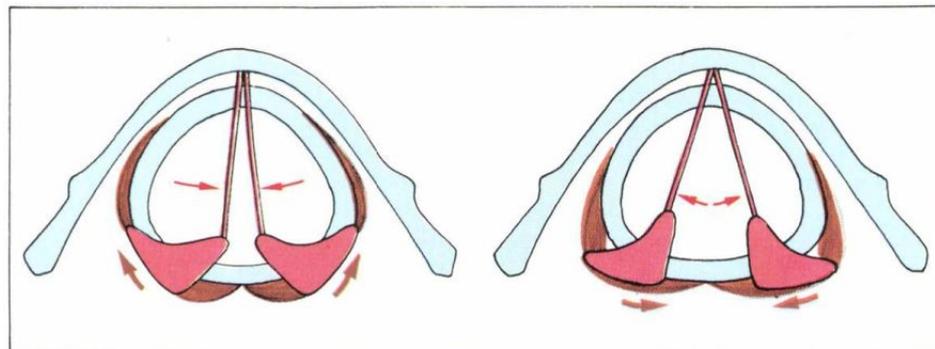
L'équipe GeoStat entreprend des recherches fondamentales et appliquées sur des nouvelles méthodes émergentes en analyse non-linéaire des signaux et systèmes com-

plexes, en utilisant des paradigmes liés aux notions d'invariance d'échelle, de prédictibilité, ainsi que dans l'utilisation de méthodes provenant de la physique statistique.

0.2 INTRODUCTION

COMMENT UN SON EST PRODUIT ?

Si on remonte à l'origine du mécanisme, on observe que de l'air est expulsé par les poumons, cette air passe par la trachée. Sous la pression de l'air les cordes vocales se mettent à vibrer s'en suit un cycle d'ouverture et de fermeture de la glotte (cf Figure 1). C'est ainsi qu'un son est produit.



Larynx : coupes horizontales passant par les cordes vocales et les cartilages aryténoïdes.

A gauche : fermeture de la glotte (voix);
à droite : ouverture de la glotte (respiration).

FIGURE 1 – Mécanisme de phonation

QU'EST CE QUE LE SIGNAL SOURCE ?

le flux glottal ou communément signal source, est la vélocité de l'air passant par la trachée quand un son est produit(cf Figure 2). Le signal source est un élément important et présent dans tout les études liées de près ou de loin à la parole. Toute les technologies liées à la parole incorporent cette idée de signal source. C'est encore plus vrai quand on regarde le système de production comme un modèle **Source-Filtre**.

LE MODÈLE SOURCE-FILTRE

La production des sons de parole peut être modélisée selon la théorie source-filtre (Fant, 1960) (cf figure 3). Cette théorie consiste à mettre à dire que la source des sons provient du souffle pulmonaire (flux glottal) qui fait entrer en vibrations les cordes vocales (CV) et le filtre est représenté par les cavités supra-glottiques(cf figure 3). Ce filtre joue un rôle d'amplificateur de certaines fréquences.

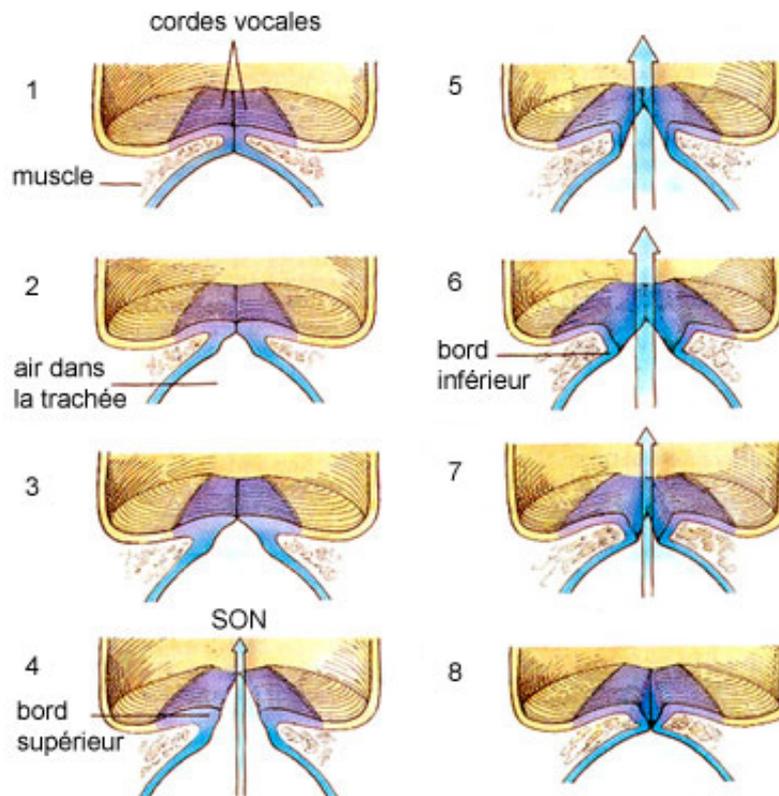


FIGURE 2 – Mécanisme de phonation

VOISEMENT

On peut caractériser les sons de la parole en deux classes, les sons *voisés* et *non-voisés* (sourds). Physiologiquement, un son voisé est produit directement par les cordes vocales. Le son est le résultat de leur vibration. Les voyelles sont naturellement voisées. Mathématiquement, c'est traduit par un signal quasi-périodique. En effet, on observe le signal, sur les parties voisées, on observe des motifs qui se répètent (cf figure 4).

Inversement aux sons voisés (*voiced*), les non-voisés (*unvoiced*) n'ont aucune structure de périodicité. Ils s'apparentent à du bruit. Pendant leur production, la glotte reste ouverte, laissant l'air convoyer jusqu'à la bouche. Majoritairement, ces sons résultent des mouvements des lèvres et de la configuration de la langue.

PLAN DU RAPPORT

Voici comment se déroulera le rapport. Dans un premier chapitre, nous parlerons du signal source. Nous exposerons en outre le modèle LF du flux dérivatif. Dans un second chapitre, nous parlerons de l'estimation du signal source, ses enjeux et ses applications. Puis nous présenterons la méthode de Drugman pour estimer la source. Dans le dernier, nous ferons des expérimentations sur l'algorithme de Drugman.

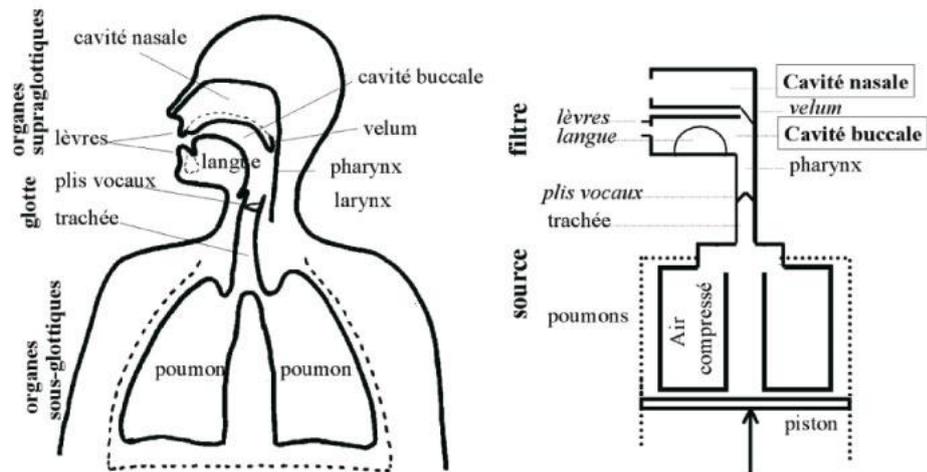
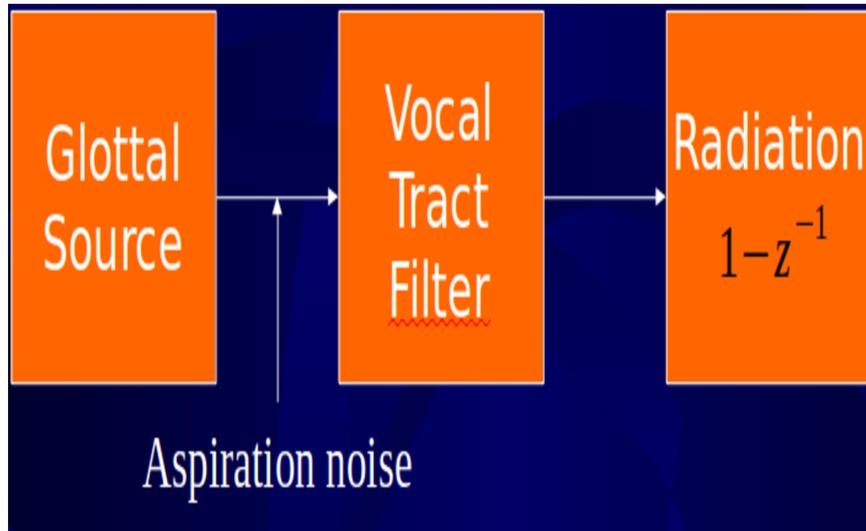


FIGURE 3 – Modele Source-Filtre

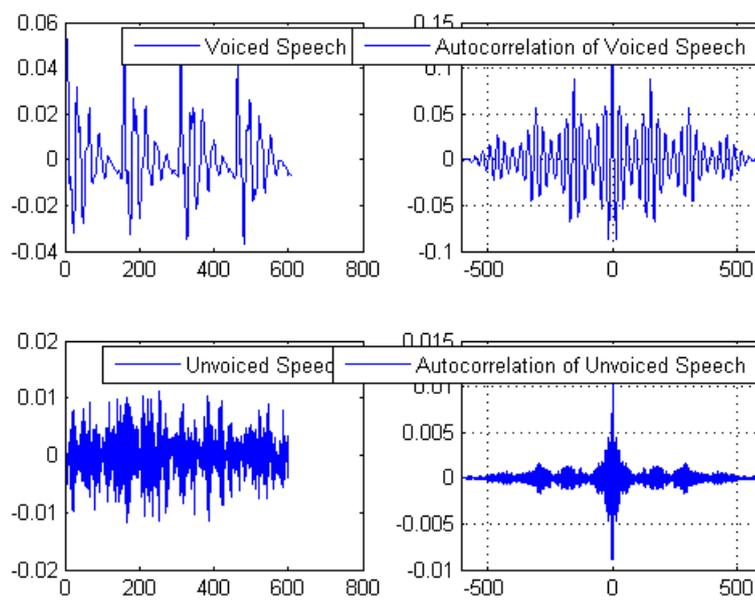


FIGURE 4 – Voisé et non-voisé

MODÈLE LF

Contents

1.1	Modèle Flux Glottal	7
1.2	Modèle LF	8
1.3	Model LF code-1 sous matlab	10
1.4	Soft-matlab basé sur le model LF	11

1.1 MODÈLE FLUX GLOTTAL

Dans les technologies se basant sur le modèle source-filtre, il existe plusieurs modèles plus ou moins réaliste qui s'accorde avec les faits physiologiques. Dans la littérature les plus connus sont les suivantes :

- Le modèle KLGOTTT88 (Klatt, 1990) : le flux glottal est modélisé par un polynôme d'ordre 3, dont les paramètres sont A_v , T_0 , $O_q = \frac{T_e}{T_0}$ et T_L .
- Le modèle R++ (veldhuis, 1998) : le flux glottal est modéliser par un polynôme d'ordre 4 pour la phase d'ouverture, et une courbe exponentielle pour la phase de retour. il est paramétrisé par 5 paramètres, qui sont T_0 , T_e , T_p , et T_a
- Le Modèle Rosenberg C (Rosenberg, 1971) : le flux glottal est modélisé par deux sinusoides. il est paramétrisé par A_v , T_0 , T_p , $T_n = T_e - T_p$
- Le modèle LF(Fant, Liljencrants, 1985[4]) : Le model LF modélise le flux dérivative et non le flux glottal.

Tous ces modèles ont ces propriétés en communs :

- le flux glottal est toujours positive
- le flux et sa dérivé sont quasi periodique
- ce sont des courbes en cloche
- ils ont tous en communs les paramètres T_0 , T_p , T_e , T_p

En ce qui nous concerne c'est le modèle LF que nous utiliseront, c'est l'un des modèles les plus aboutis avec le modèle KLGLOTT88.

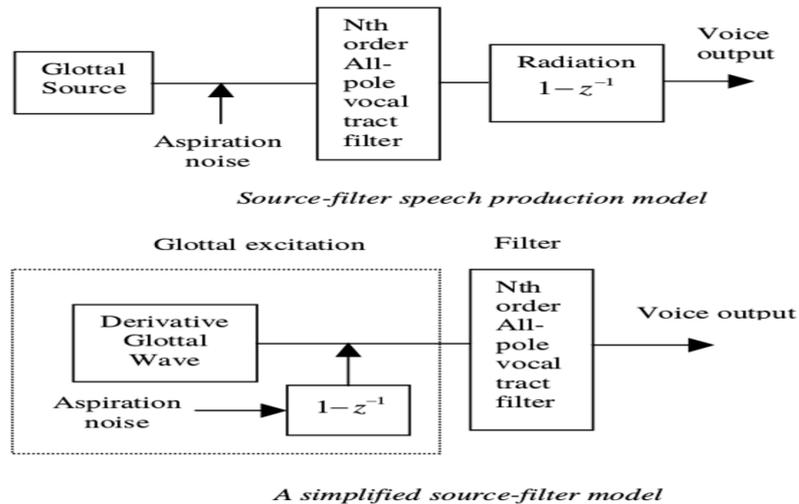


FIGURE 5 – Modèle Source-filtre simplification

1.2 MODÈLE LF

Le modèle Liljencrants-Fant[4] ou Modèle LF est un modèle mathématique décrivant la dérivée du flux glottal.

POURQUOI LA DÉRIVÉ PLUTÔT QUE LE FLUX ?

On peut simplifier le modèle Source-Filtre en utilisant la dérivée de la source, plutôt que la source. Quand la vélocité de l'air arrive au niveau de la bouche, le mouvement des lèvres produit un effet dit de radiation (*lips radiation effect* cf figure 3). L'effet de radiation se traduit comme un opérateur de dérivation. La dérivation est linéaire, donc sa position importe peu sur la chaîne de production. Sur le modèle source-filtre utiliser le flux dérivé revient à placer cet opérateur avant le filtre (cf figure 5).

POURQUOI SIMPLIFIER LE MODÈLE ?

D'un point de vue pratique c'est plus simple et avantageux. Simple car entre le flux et sa dérivée le signal porte les mêmes informations. Les informations sont même plus flagrantes sur la dérivée. C'est à dire, le flux dérivé donne visuellement plus d'information.

Exemple : On a un pic négatif quand débute la phase de fermeture, ou encore le flux qui s'annule en T_e .

Avantageux car on a qu'une seule opération à faire. On fait un filtrage pour produire un son et une déconvolution pour estimer le flux.

PARAMÉTRISATION

le flux dérivé est décrit par les équations suivantes :

$$E(t) = E_0 e^{\alpha t} \sin \omega_g t \quad (t < t_e) \quad (1)$$

$$E(t) = \frac{-E_e}{\beta t_a} \left[e^{-\beta(t-t_e)} - e^{-\beta(t_c-t_e)} \right] \quad (t_e \leq t < t_c) \quad (2)$$

Sa représentation graphique est deux courbes exponentiels.

La première équation représente la phase d'ouverture (*open phase*).

La seconde représente la phase de fermeture (*closing phase*) de glotte.

Pendant la phase d'ouverture (*open phase*) les parois des cordes vocales s'ouvrent et l'air est expulsé. Ensuite elle se ferme (*closing phase*) pour atteindre sa position de repos pendant la phase de fermeture. (cf figure 6)

Le modèle LF du flux dérivé est défini par ses ensembles de paramètres : $E_e,$

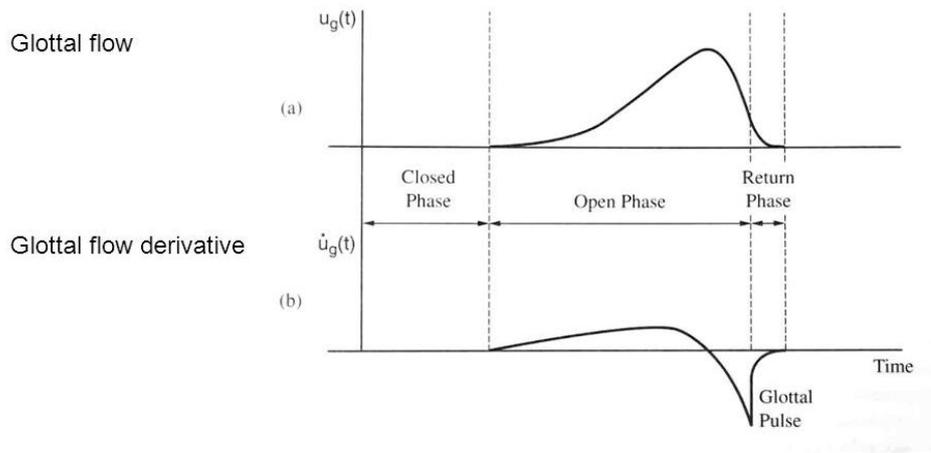


FIGURE 6 – Flux glottal et sa dérivée

$(E_0, \alpha, \omega, \epsilon), (T_p, T_e, T_a, T_c),$. Explicitons les :

PARAMÈTRE TEMPOREL

- T_p représente en temps le moment où la vitesse de l'air atteint son maximum. À T_p le flux dérivé est nul.
- T_c représente en temps la fermeture de la glotte. Où encore la fin de la phase de retour (*return phase* cf figure 6)
- T_0 représente en temps la durée d'un cycle. Pendant T_0 on a un cycle d'ouverture et de fermeture de la glotte. C'est aussi l'inverse de la fréquence fondamentale
- T_e représente en temps le début de la phase de fermeture.
- T_a représente en temps la durée effective de la phase retour.
-

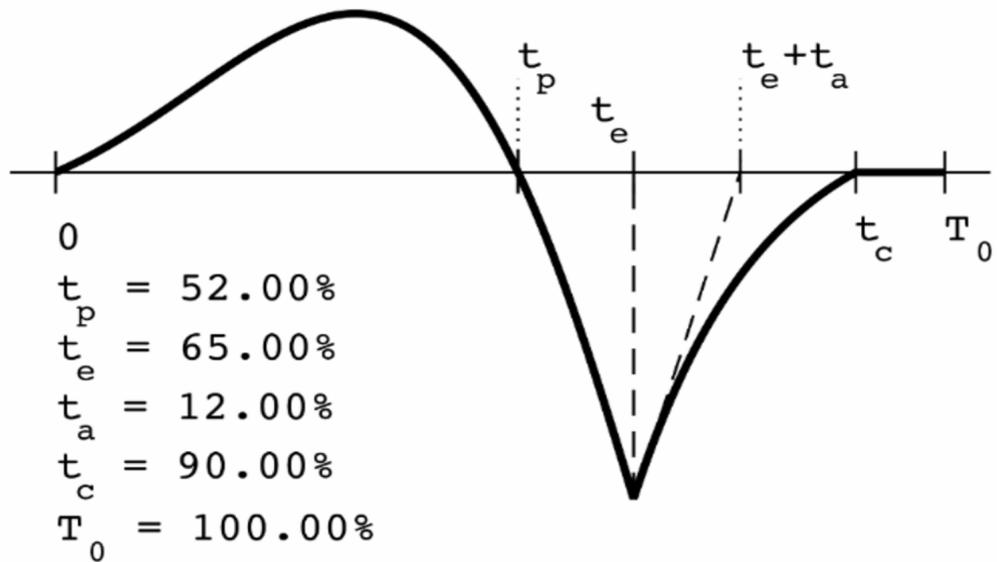


FIGURE 7 – Model LF flux dérivative

Sur cette figure on a en temps normaliser par T_0 un cycle du flux dérivative.

LES AUTRES PARAMÈTRES

- E_e est l'amplitude atteints en T_e .
- α est un paramètre qui module l'amplitude max du flux pendant la phase d'ouverture
- β détermine à quel vitesse la courbe retourne en zero pendant la phase retour.
- w_g coefficient qui détermine la courbure du flux pendant la phase d'ouverture. Il détermine aussi la longueur de la phase d'ouverture

1.3 MODEL LF CODE-1 SOUS MATLAB

Nous allons présenter notre premier code matlab utilisé pour générer le flux. Nous allons utiliser une programme sous matlab de mike Brookes. l'algorithme utilise 3 paramètres pour générer un flux dérivative.

Nous avons T_e , E_e et $P(3)$. $p(3)$ ce paramètre est un ratio qui contrôle la convergence en zero du flux de la phase retour. Donc pour contrôler le flux nous avons que ces 3 paramètres. Nous les avons moduler pour obtenir toute les variations possibles.

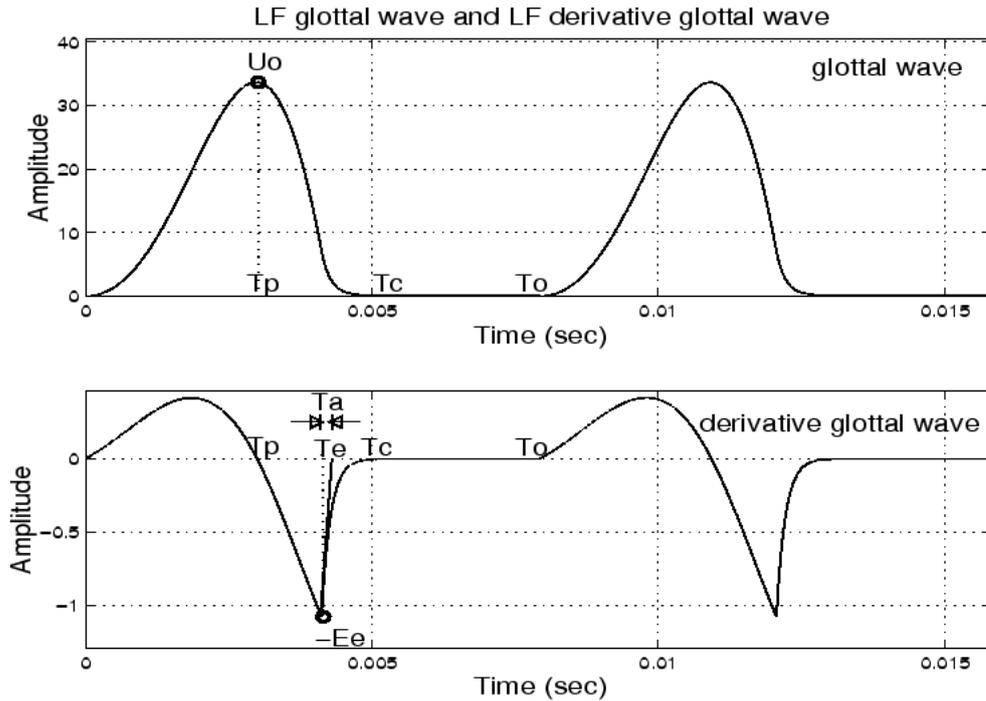


FIGURE 8 – Model LF et son flux dérivative

1.4 SOFT-MATLAB BASÉ SUR LE MODEL LF

Nous allons présenter un autre programme permettant de générer le flux dérivative. Ce soft fonctionne sous matlab en indépendant(*embed*). Il permet de générer des voyelles synthétiques selon le modèle-source filtre. Il utilise le modèle LF du flux dérivative pour générer le signal source. Nous avons isoler cette partie du programme pour générer des trains de flux dérivative.

Présentation

C'est un soft à but pédagogique de Hideki Kawahara. Il comporte deux parties. une partie filtre et une partie source. ces parties sont séparées par deux fenêtres. Dans la partie filtre on peu moduler, configurer les paramètre du filtre. Mais aussi l'ouverture entre les lèvres. On peut aussi choisir le sexe, le vibrato ou encore F_0 la fréquence fondamentale du flux.

Dans la partie source on peut moduler la forme de la source. la partie 2 est accessible que si on choisie le mode *Disign*. c'est un mode permettant de configurer le flux à la main et d'observer le résultat.

Fonctionnement du soft

Il faut charger une configuration du filtre vocal et du flux. Cette configuration est une structure (*struct*) matlab, avec en données les coefficients du filtre vocal et les paramètres du flux dérivative. Le soft propose une configuration par défaut. cette configuration est un [a].

Le soft a deux modes de fonctionnement.

Mode Normal

On peut générer des voyelles en modulant les coefficients du filtre. On peut aussi influencer sur la source au travers, avec la fréquence fondamentale F_0 et le sexe

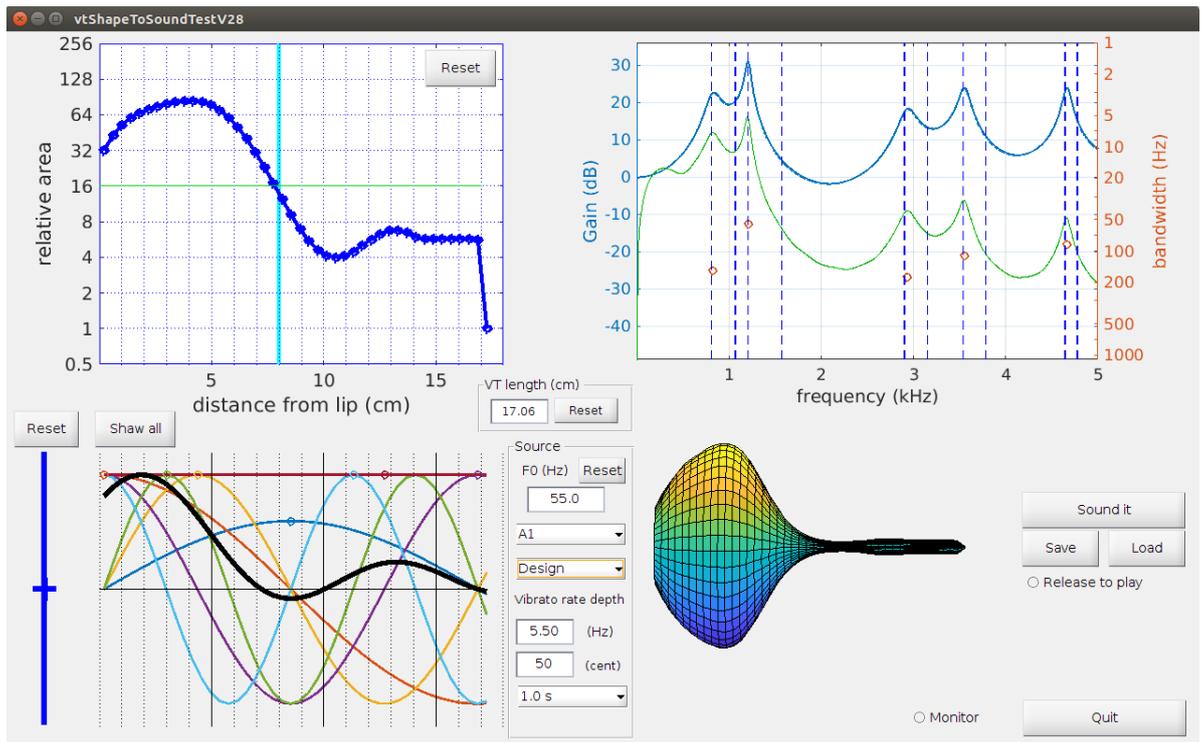


FIGURE 9 – partie 1 du soft

Mode Design

Dans ce mode on peut générer des voyelles soutenues en modulant les paramètres du flux dérivative.

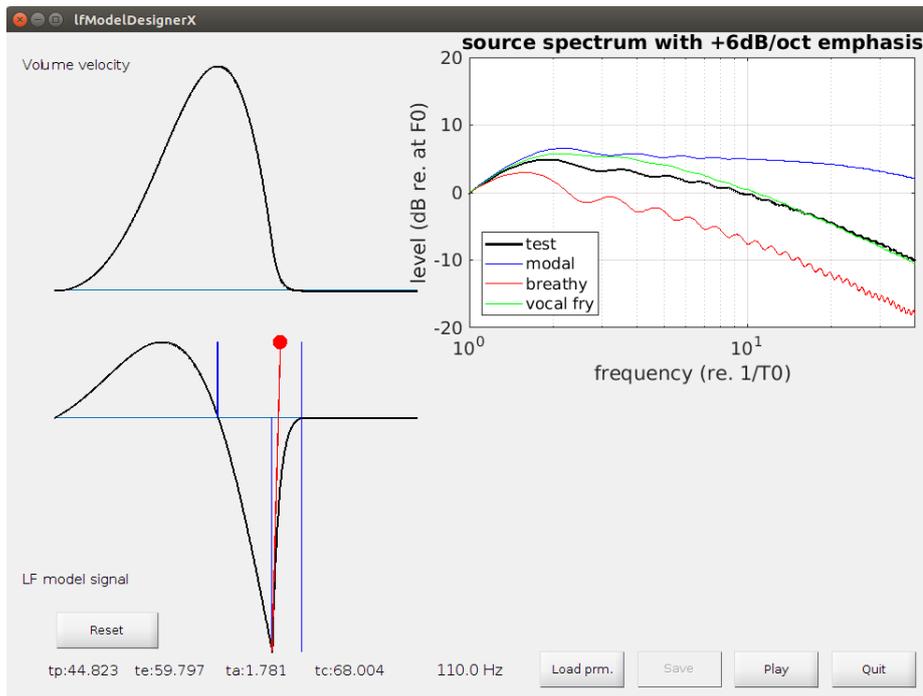


FIGURE 10 – partie 2 du soft

Contents

2.1	Enjeux et Applications	15
2.2	Closed phase inverse filtering	17
2.3	Iterative and inverse filtering	18
2.4	Mixed-phase decomposition	18
2.4.1	Description de la Méthode de Drugman	18
2.4.2	Description du code	19

2.1 ENJEUX ET APPLICATIONS

Pourquoi on veut estimer le signal source il y a plusieurs raison. Tout part du fait suivant qui est que chaque personne a une voix différente. car nos cordes vocal ont une unique configuration. cette unicité se traduit naturellement sur le flux glottal. Cette unicité du flux peu aider dans les domaines techniques

SYNTHÈSE VOCAL

En synthèse vocal, on développe des technologies reproduisant la voix humaine. La voix humaine a une multitude de configuration. Ces configurations sont liées à la forme du signal source, Donc il est important pour un modèle de savoir caractériser toute les subtilités de la voix(cf figure 11).

DÉTECTION DE LOCUTEUR

comme on l'a dit précédemment chaque personne a une unique signature vocale, c'est encore vrai pour le signal. Donc une bonne modélisation de ce signal permettrai de mieux caractériser la voix. Un exemple d'application serait lors d'une conférence, on aimerai séparer 3 personnes qui discute à tours de rôles, une bonne manière de le faire serait de caractériser leur signal source. Vu sous cette angle Il est claire que pouvoir caractériser le signal est un enjeux non négligeable.

BIOMÉDICALE

Dans le biomédicale on étudie les mécanismes de phonations. On a remarqué que le dysfonctionnement de ses mécanismes étaient liée à certaine maladie. On peut citer la diplophonie ou l'aphonie. Ce sont des pathologies liée à l'appareil vocal, mais plus récemment on a constaté qu'il y a avait un lien entre ces dysfonctionnements et les maladies dégénératif comme Parkinson. Les technologies actuelle font appelle à des machines performantes mais coûteuses, grand nombre sont intrusive et gênant pour

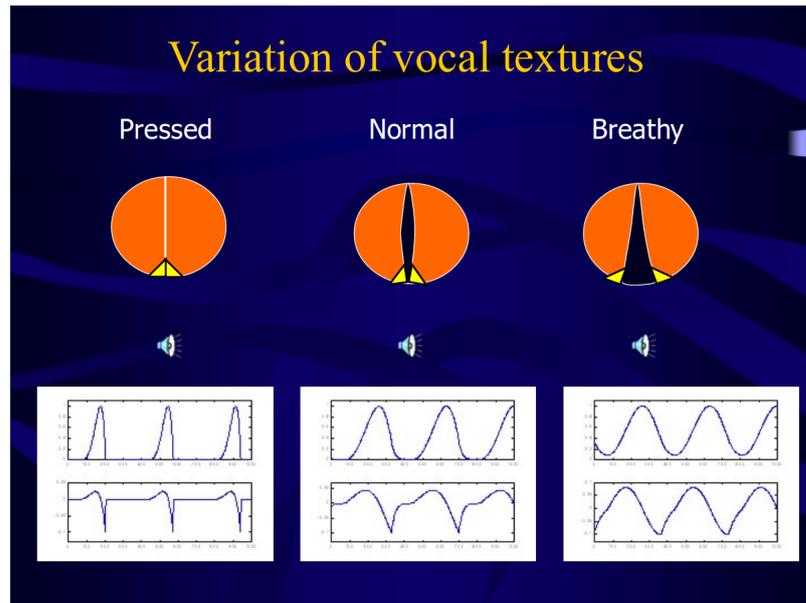


FIGURE 11 – Modèle LF et Texture vocal

le patient. Puis certaine maladie ne peuvent être détecter que dans une conditions réel c'est a dire sans machine ou camera interne. Donc l'idéal, serait de pouvoir détecter ses maladies autrement. Puisque le signal source est liée au informations sur la glotte qui lui mêmes révèle des informations physiologique. Développer des technologie se basant que sur la parole apporterai une avancé considérable dans ce milieu.

ESTIMATION DU SIGNAL SOURCE

Il y a plusieurs approches pour estimer le signal source ou flux glottal. il y a l'approche direct qui utilise des instruments ou capteurs sur l'organe vocal. On a une information en temps réel de l'activité organique. Cette méthode est contraignant en pratique(méthode intrusive) Mais elle est de loin la plus riche car prélevé à la source. Puis il y a la méthode indirect où on part du signal parole, on remonte l'information. On parle de déconvolution ou filtrage inverse. nous discuterons et utiliserons que cette dernier approche dans notre travail.

Estimer le signal source est un problème difficile,c'est devenu possible car il existe des algorithmes performant pour repérer les GCI/GOI.

Les GCI et GOI sont les instants d'ouverture et fermeture de la glotte.

GCI/GOI ET MODEL LF

Sur le model LF du flux dérivative les GCIs sont représentés par le paramètre T_e . A T_e on a un pic négative, à cette instant débute la fermeture de la glotte. le GOI est représenté par le paramètre t_0 , à cette instant débute la phase d'ouverture cf figure 12.

Notre étude se base sur un modèle source-filtre. Le signal parole produit de ce modèle résulte d'une convolution, ces convolutions sont appliquer et synchroniser sur

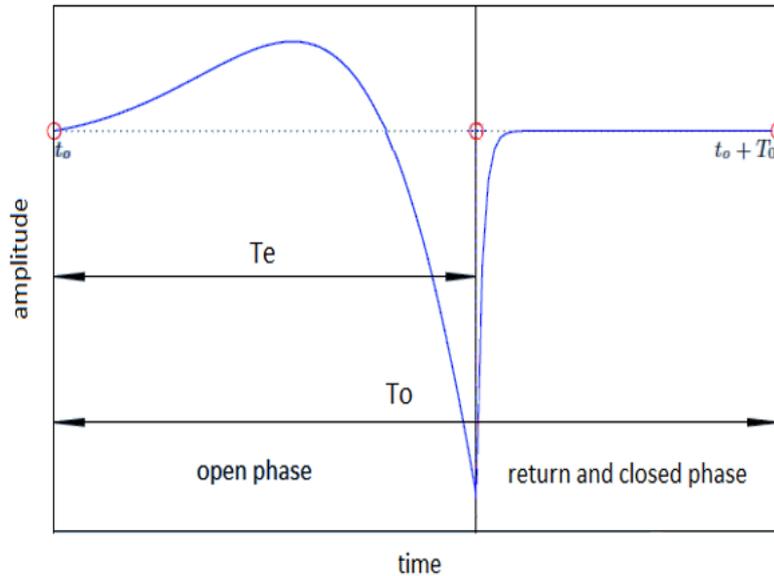


FIGURE 12 – GCI-GOI flux dérivative

chaque train de flux glottal(signal source). Pour faire une déconvolution et retrouver le signal original, on a besoin de retrouver la position des ses flux. On parle de synchronisation. D'où l'importance des GCI/GOI. elles déterminent le début et la fin d'un train de flux glottal. Si on connaît le début et la fin d'un flux on sait où appliquer la déconvolution. Si on ne connaît pas ces positions il est impossible de faire de l'estimation de source avec le signal parole.

Dans la littérature, il existe 3 classes de méthodes on a les *closed-phase inverse filtering*, *Closed phase inverse filtering* et *Mixed-phase decomposition*. Ses méthodes ont leurs avantages et inconvénients. Aujourd'hui il n'existe pas d'études comparatives entre ses méthodes. On peut seulement dire qu'une méthode est meilleure qu'une autre selon la situation.

Exemple : les méthodes itératives ont de meilleurs résultats sur des données bruitées alors que les deux autres méthodes sont meilleures en précision sur les données propres sans perturbation(*clean speech*)

2.2 CLOSED PHASE INVERSE FILTERING

Cette méthode se focalise sur les instants de fermeture de la glotte. On veut faire du filtrage inverse pour récupérer le signal source. Pendant cette phase de fermeture l'activité subglottale est minimale(région des cordes vocales), ce qui a pour conséquence d'avoir une meilleure estimation du filtre vocal. Une fois qu'on a récupéré les coefficients, on fait une déconvolution pour obtenir le signal source. Pour que cette méthode fonctionne il est fondamental de connaître les GCI/GOI. ce sont les instants de fermeture et les instants de fermeture de la glotte. ce sont des marqueurs temporels. elle nous indique à quel instant du signal il faut appliquer la déconvolution

2.3 ITERATIVE AND INVERSE FILTERING

Dans cette méthode le signal est estimé récursivement. On commence par une première estimation, à mesure que l'algorithme avance dans le temps, on a une meilleure estimation du signal source, avec un critère d'arrêt. Cette méthode est de loin la plus efficace quand le signal de départ est soumis à du bruit ou perturbations.

La méthode la plus connue est celle de Kalu (Kalu et al., 1992)

2.4 MIXED-PHASE DECOMPOSITION

Dans cette méthode le signal parole est assimilé à un système à phase mixte (Bozkurt, Dutoit, 2003). Un système est dit à phase mixte si il est stable et les zéros de la fonction transfert sont localisés à l'intérieur et à l'extérieur du cercle unité.

La phase minimum est la partie du système dont les zéros sont à l'intérieur du cercle. La phase maximum c'est l'inverse. La phase est maximum est ici le signal source tandis que la phase minimum la réponse du filtre vocal. Cette méthode vise à séparer ces deux contributions. Dans la littérature il existe plusieurs méthodes. Mais nous, nous allons discuter de la méthode de Drugman pour faire cette séparation, *complex cepstrum decomposition* [3]

2.4.1 Description de la Méthode de Drugman

Nous allons parler de l'algorithme de Drugman. Sa méthode s'appuie sur la décomposition de phase expliquée plus haut. Donc la parole est un système à phase mixte. La phase maximum est le signal source et la phase minimum le filtre de la cavité buccale. On souhaite faire une séparation. Cette séparation est possible dans le domaine cepstrum

Analyse Cepstrum

Le cepstrum est une technique permettant de décomposer des signaux non linéaires, on parle alors d'analyse homomorphique [5]. Dans le cepstre la convolution se transforme en addition. Le cepstre va séparer la phase maximum et la phase minimum la contribution du filtre vocal.

Soit $x(n)$ un signal discret et $\hat{x}(n)$ sa transformée de cepstre, alors :

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \quad (3)$$

$$\log [X(\omega)] = \log(|X(\omega)|) + j\angle X(\omega) \quad (4)$$

$$\hat{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log [X(\omega)] e^{j\omega n} d\omega \quad (5)$$

Comment fait-on la séparation ?

un système à phase mixte, une fois discrétisé dans le domaine cepstre la partie anti-causal c'est exactement la partie du système dont les zéros sont à l'extérieur du cercle unité. Et la partie causal les zéros a l'intérieur du cercle unité. Donc si on isole la partie causal et la partie anti-causal du signal discret on a réussi à faire notre séparation.

Où appliquer cette séparation ?

Cette séparation n'est pas n'importe comment. Pour produire un son on applique le filtre sur un train de flux glottal. il est évident que pour l'inverse ou il faut l'appliquer sur une portion où on est sûr d'avoir un train de flux glottal. d'où l'intérêt de repérer les GCI. si on a les GCI du signal on connaît la position des flux glottal et donc on sait où appliquer le cepstre pour faire la décomposition.

2.4.2 Description du code

ÉTAPE 1

Dans l'introduction on a parlé de son voisé et non-voisé. Et il est difficile de faire de l'estimation glottal sur tout un signal. Car il y a des parties non voisées. Sur ces parties, il n'y a pas de structure périodique donc impossible de trouver les GCI/GOI. Donc une première étape l'algorithme va repérer les parties voisées et non voisées du signal. À l'aide d'un algorithme dit de *pitchtracking*.

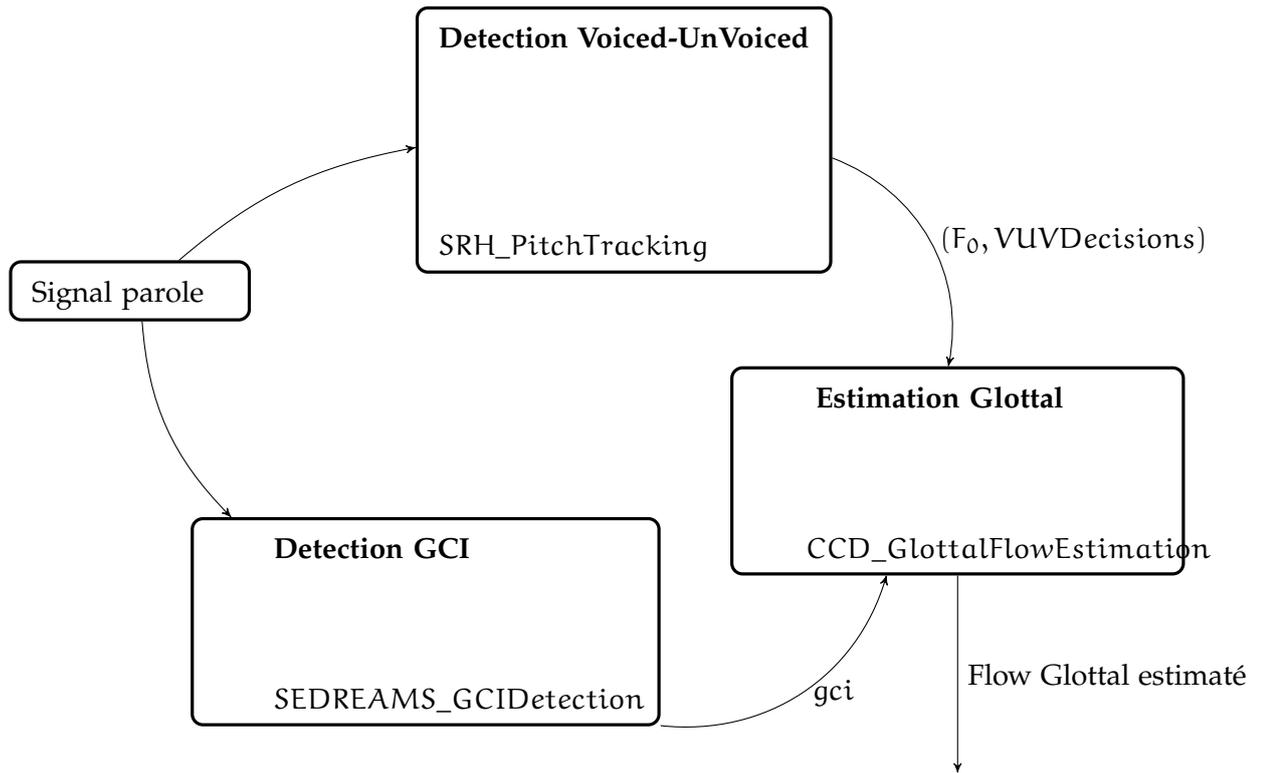
En parallèle, à l'aide d'un second algorithme on va détecter les GCIs. On obtiendra un vecteur qui sur tout le signal va donner les instants de fermeture de la glotte, ainsi que

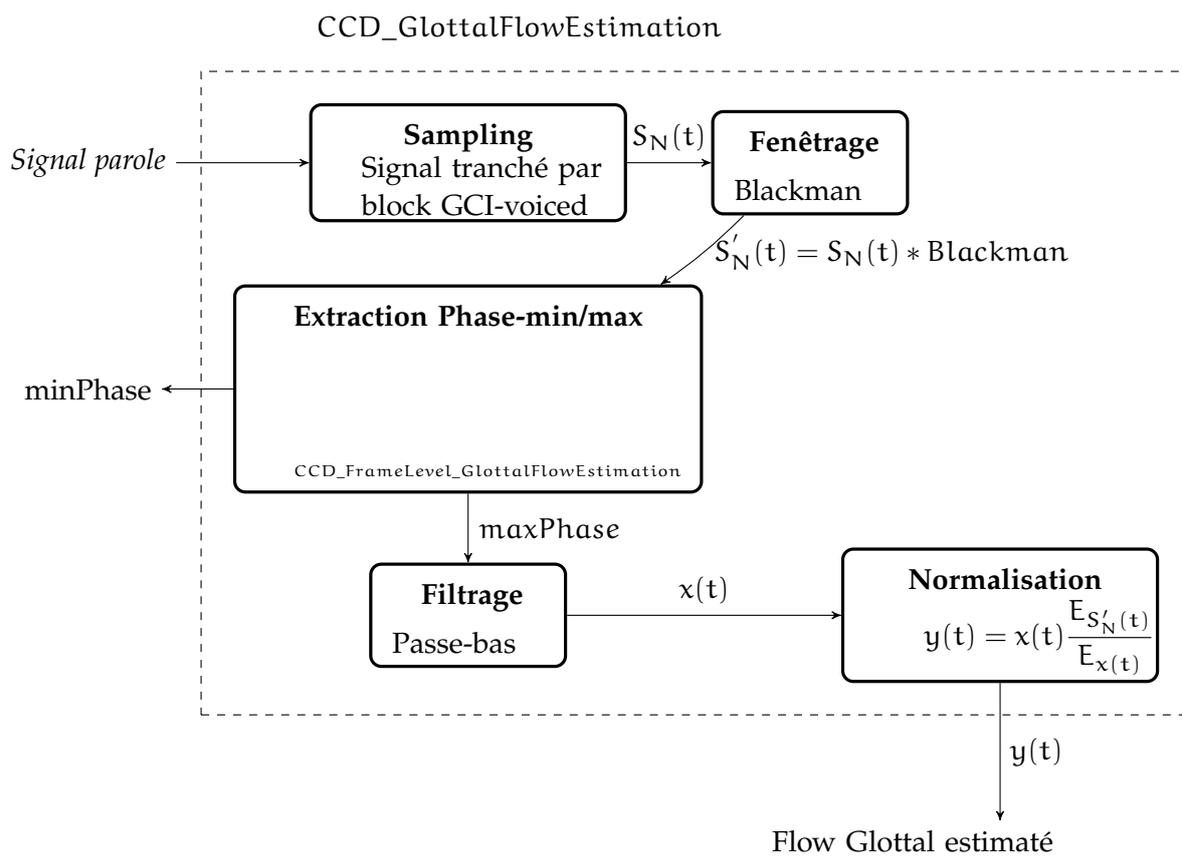
ÉTAPE 2

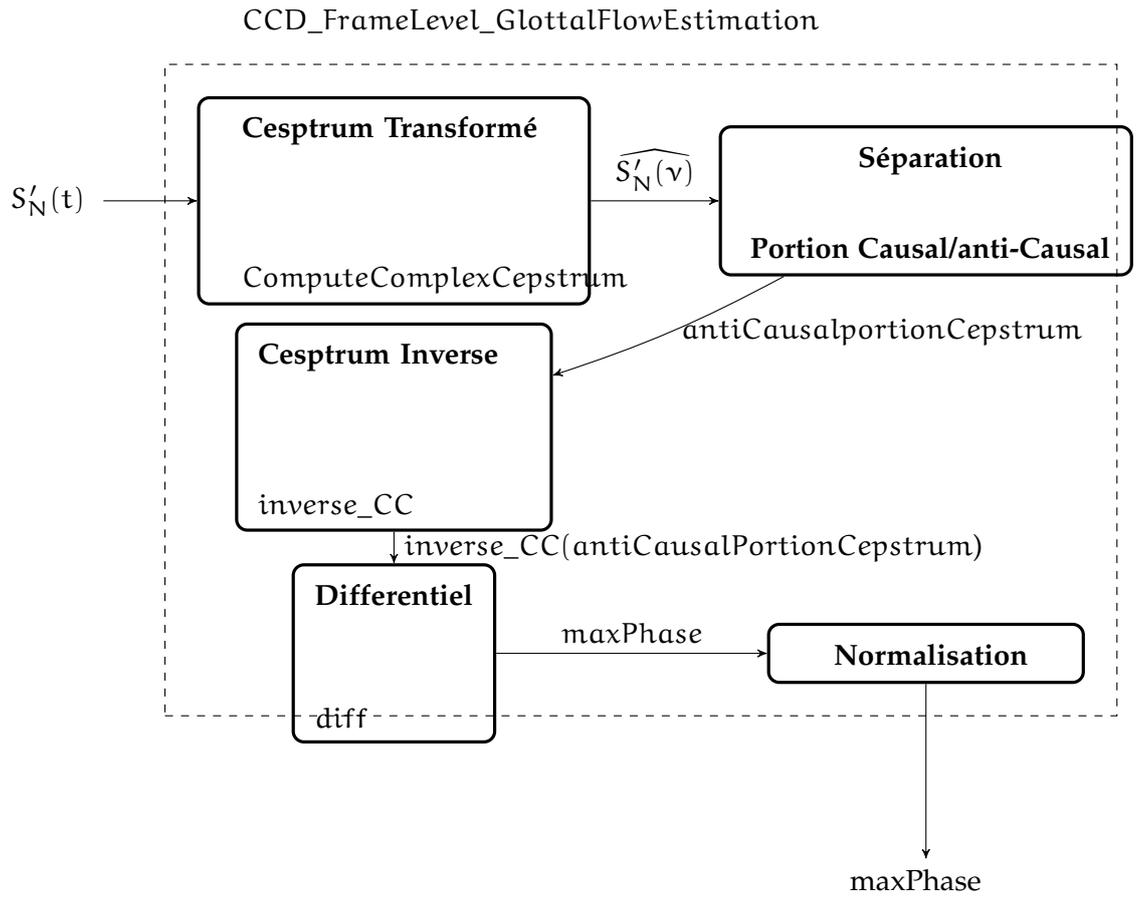
On a obtenu les parties voisées et non-voisées du signal. Ainsi que les positions des GCI. On va faire des opérations, cette opération consiste à garder que les GCIs des parties voisées du signal.

ÉTAPE 3

On va appliquer le cepstre pour faire la séparation. On a les GCI, entre deux GCIs successifs il y a un train de flux glottal. On applique le cepstre sur ces portions pour obtenir le flux glottal.







Contents

3.1	Génération du signal source	23
3.1.1	Génération de flux avec Code-1 matlab	24
3.1.2	Génération de flux avec Code-2 soft	24
3.2	Estimation sur une trame	25
3.3	Estimation sur une voyelle soutenu	26

Nous allons faire des tests sur l'algorithme de Drugman. Comme C'est impossible sur un signal réel de connaître le signal source, on ne peut vérifier la pertinence de sa méthode. On a besoin de connaître la source pour juger la méthode. Pour contourner cette difficulté on va construire de toute pièces un signal parole. on s'appuiera sur le modèle source-filtre 3 et le modèle LF du flux dérivative.

Dans un premier temps nous ferons des tests visuel. On va générer des cycles de flux glottal. A l'aide d'une méthode LPC on va extraire les coefficients du filtre sur une voyelles soutenue. On filtrera ses coefficients sur les sources générés. Ensuite on utilisera l'algorithme de Drugman pour retrouver les sources de départ.

Dans une seconde partie on fera ses teste sur des voyelles entières. A l'aide du soft on va générer des voyelles synthétiques dont le flux dérivative est connu. Dans une dernier partie on fera des testes numériques

3.1 GÉNÉRATION DU SIGNAL SOURCE

Pendant ce stage nous avons trouvé un premier code matlab pour générer la source(*code mike brookes*). On a fait nos premiers testes avec ce code. Avec ce code on ne peut influencer que sur 3 paramètres. ses paramètres sont t_e , E_e et $p(3)$. $p(3)$ contrôle la pente de la phase de retour du flux. avec 3 paramètres on a généré des flux qu'on a ensuite filtrer par les coefficient LPC d'une voyelles. Nous avons estimé le flux sur les trames filtrés. Dans la suite du stage on a trouvé un soft qui permet a partir d'un flux défini générer des voyelles soutenus. On a pu comparer visuellement les vrais sources avec la source estimé par Drugman.

COMMENT AVOIR LE FILTRE DU VOCAL ?

Une des façons qu'on peut faire pour extraire les coefficients du filtre vocal est la méthode LPC. la méthode LPC est une technique permettant d'extraire les coefficient du tractus vocal d'un signal parole.

COMMENT UTILISER LA MÉTHODE LPC ?

Sur un signal audio il est possible d'extraire les coefficients du filtre vocal. Mais on

ne peut le faire n'importe comment. En effet un signal audio est une onde dont les caractéristiques varient dans le temps, sur un signal donnée les coefficients du filtre varient. d'une portions à un autres du signal elles sont différentes. De plus on a besoin d'une conditions de stationnarité pour extraire les coefficients. Comme un signal n'est jamais stationnaire sur toute sa durée, la solution serait de prendre des petites portions du signales. 25 ms est suffisant petit pour garantir toutes ses contraintes

LES VOYELLES SOUTENUS ?

Sur un signal parole il y a les sons voisés et non voisés. Il est facile de caractériser les sons voisés. Les voyelles soutenues sont des sons voisés, donc les signaux sont quasi periodiques. Ce qui veut dire qu'elles ne changent peu dans le temps. Extraire les LPC sur ses signaux est plus facile.

3.1.1 Génération de flux avec Code-1 matlab

Avec ce premier code nous avons généré nos premier flux(cf figure 13) et fait nos premières expériences (cf figure 15). Si on a pu estimer le flux sur sur nos trame généré et filtré(cf figure 16), Faire varier le flux était une tache difficile les 3 seuls paramètres proposé par le codes. On avait besoins de contrôler tout les paramètres du flux glottal. De plus les résultats obtenue ne permettait de vraiment trancher sur la méthode (cf figure 17).

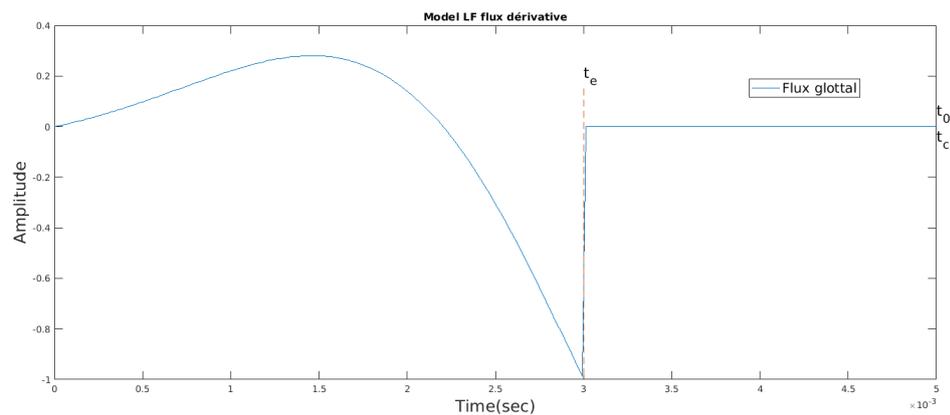


FIGURE 13 – flux dérivative généré par code-1 matlab

3.1.2 Génération de flux avec Code-2 soft

Nous avons extraits la partie du code du soft pour générer des train de flux glottal. Nous avons décider de faire varier le flux selon les critères suivantes :

T_0	O_q	α_m	Voyelles
90 : 5 : 240	0.3 : 0.05 : 0.9	0.55 : 0.05 : 0.8	[a]

Avec

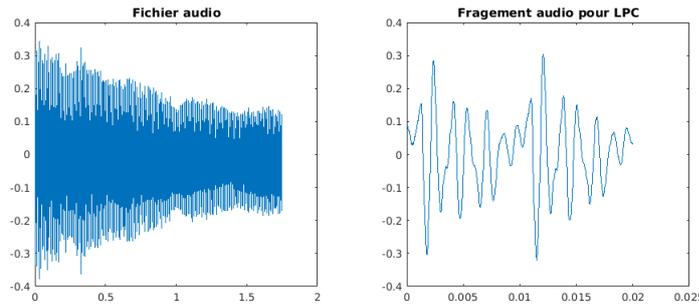


FIGURE 14 – Voyelle soutenu [a]

- $O_q = \frac{T_p}{T_0}$ le coefficient d'ouverture.
- $\alpha_m = \frac{T_p}{T_e}$ le coefficient d'asymétrie.

Cette façon de faire, permet de balayer toutes les formes possibles du flux(cf figure 18).

Ensuite nous avons modulé le paramètre T_a la durée de la phase retour, En fixant les paramètres T_0 , T_p , T_e (cf figure 19).

3.2 ESTIMATION SUR UNE TRAME

Nous allons juste parler de l'estimation du flux sur les trames générées avec le second code matlab. Contrairement au premier code, le second donne accès à tout les paramètres flux. Nous avons générés des trains de source contenant au plus 6 flux. Les flux contenus dans chaque trains ont des paramètres identiques. On a filtré les trains avec les coefficients LPC d'une voyelles [a] soutenue.(cf figure 16) Nous allons estimer le flux glottal sur un signal d'environ 25 ms 16

Resultat

Quand on connait les GCIs

Nous avons fait des tests. Pendant ces test nous avons acquisitionné sur l'algorithme la position exacte des GCIs. Après résultat nous avons conclu qu'il fallait deux conditions pour que la méthode soit optimale :

- la premier est que le paramètre T_a soit de l'ordre de 0.1% de la taille de T_0 (cf figure 20).
- La seconde est que la dérivée du flux soit discontinue en T_e . c'est à dire, en T_e quand on observe flux, on a un pic.

En dehors de ces conditions nous avons été confronté à plusieurs problèmes :

- Nous avons relevé un problème d'amplitude entre l'estimé et le vrai flux.(cf figure 21)

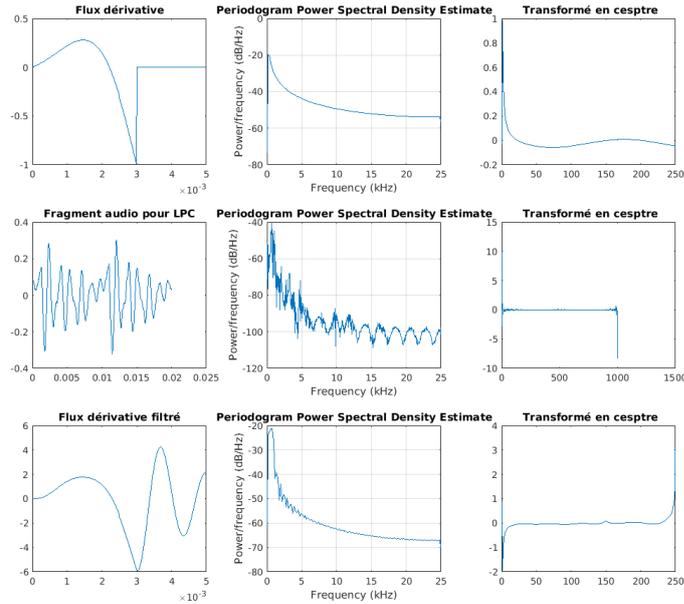


FIGURE 15 – Flux filtré avec coefficients LPC [a]

- Nous avons relevé un décalage entre le vrai flux et l'estimé. ce décalage est d'autant plus important quand $T_a > 1\%$ de la taille de T_0 .
- Nous avons relevé que la méthode a tendance à mal estimer la forme de flux sur la phase d'ouverture. Avant chaque début d'ouverture du flux, on a la présence d'artefact. Quand la fréquence augmente ces artefacts se transforme en petites oscillations. La méthode dégénère quand F_0 augmente (cf figure 21)
- Nous avons relevé que la méthode a du mal estimer la pente de la phase retour quand $T_a > 1\%$. (cf figure 21)

Quand les GCIs sont détectés automatiquement

Nous avons fait le même test. Mais cette fois on a utilisé l'algorithme de détection automatique des GCIs. Nous arrivons au même résultat que précédemment. Nos meilleurs estimation ont été quand T_a est de l'ordre de 0.1% de T_0 et la dérivée du flux discontinue en T_e . (cf figure 22 et 23)

3.3 ESTIMATION SUR UNE VOYELLE SOUTENU

Nous avons estimé le flux sur une voyelle synthétique. Ces voyelles ont été générées par le soft. Lors de ces tests nous avons fait varier la forme du flux, comme précédemment. Puis nous avons comparé le vrai flux et l'estimé. On a observé les faits suivants :

- Nous avons trouvé que la méthode fonctionne relativement bien sous les mêmes conditions annoncées dans les tests précédents. La forme de l'estimé et le vrai flux

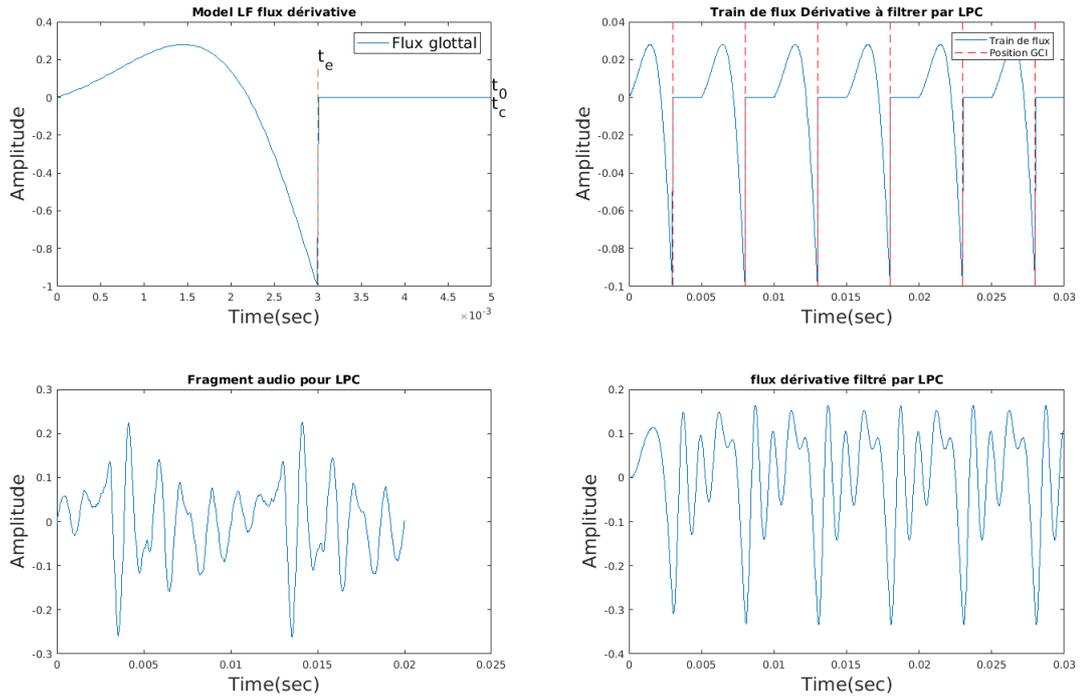


FIGURE 16 – Donnée a estimer

se superpose. A la différence que le flux estimé comporte maintenant des oscillations. (cf figure 24

- A la différence qu'à partir de $F_0 = 150\text{Hz}$ la méthode dégénère. On observe un décalage croissant entre le vrai flux et l'estimé.(figure 25, figure 26)

En dehors ces conditions la méthode dégénère.(figure 27)

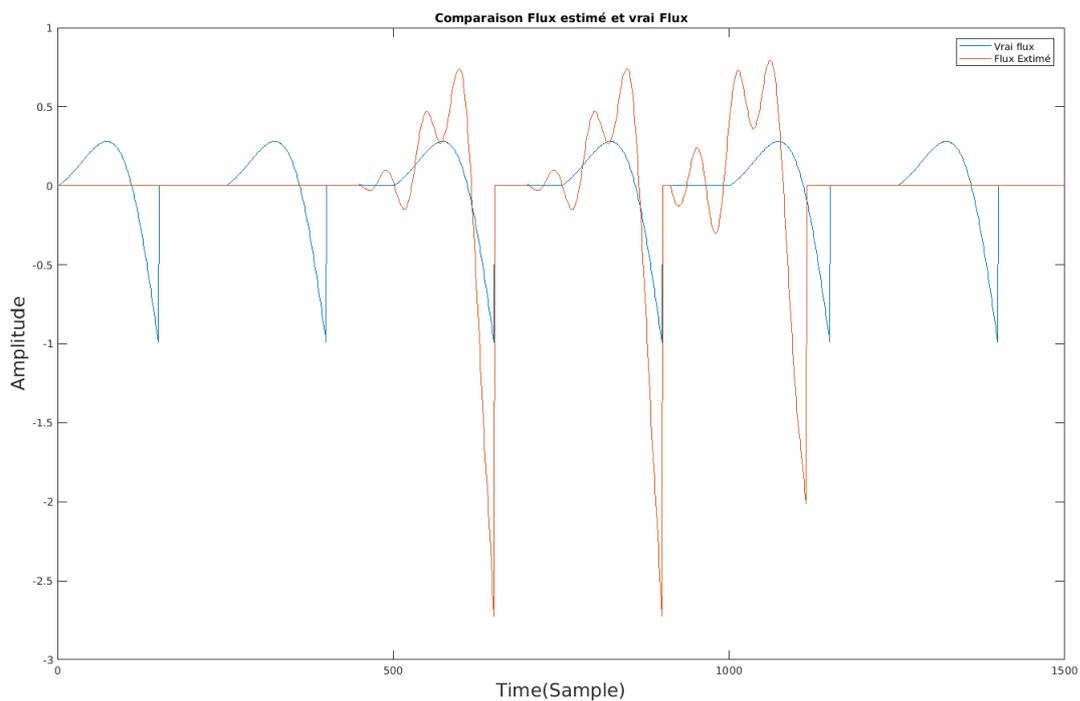


FIGURE 17 – Flux glottal estimé/flux généré par code-1

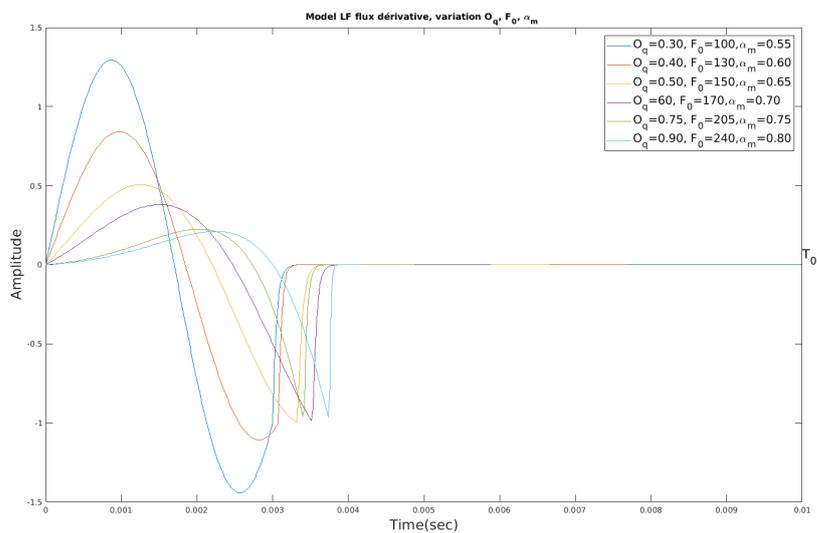


FIGURE 18 – Variation paramètre du Flux

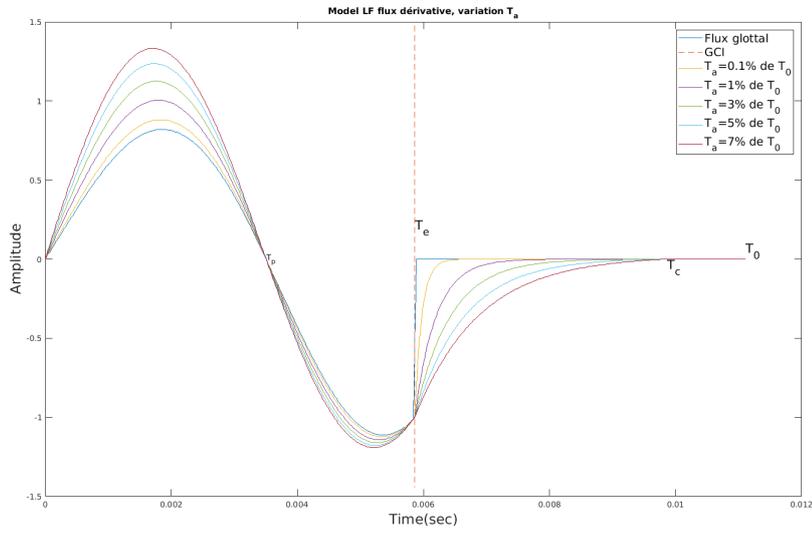


FIGURE 19 – Variation paramètre du Flux

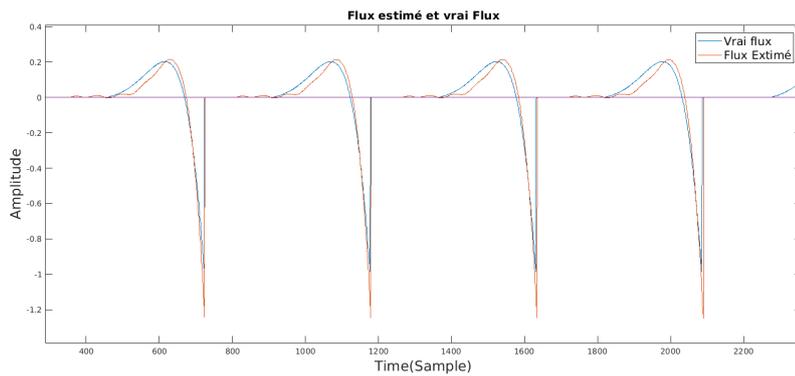


FIGURE 20 – $T_0 = T_c$, $T_p = 0.48 \times T_0$, $T_e = 0.59 \times T_0$, $T_a = 0.0002 \times T_0$

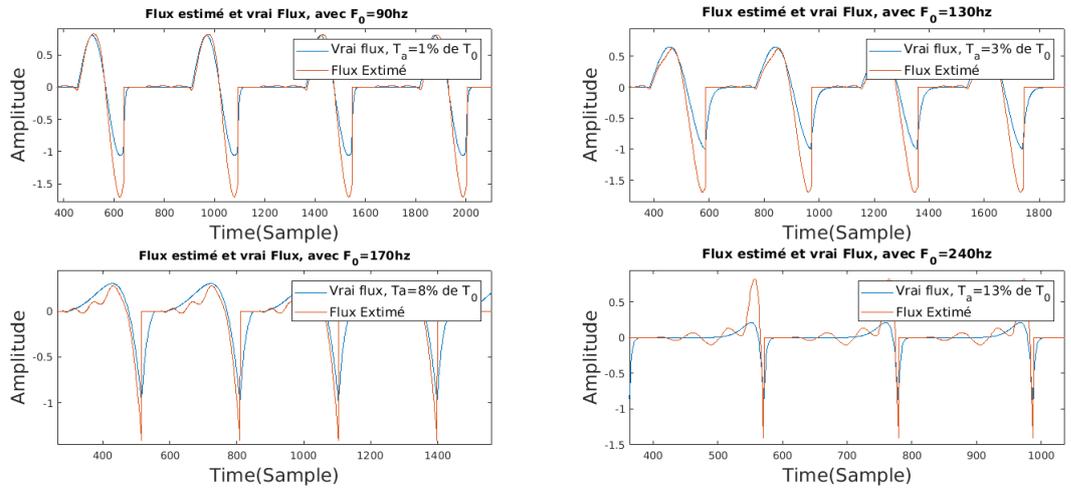


FIGURE 21 – Variation des paramètres du Flux

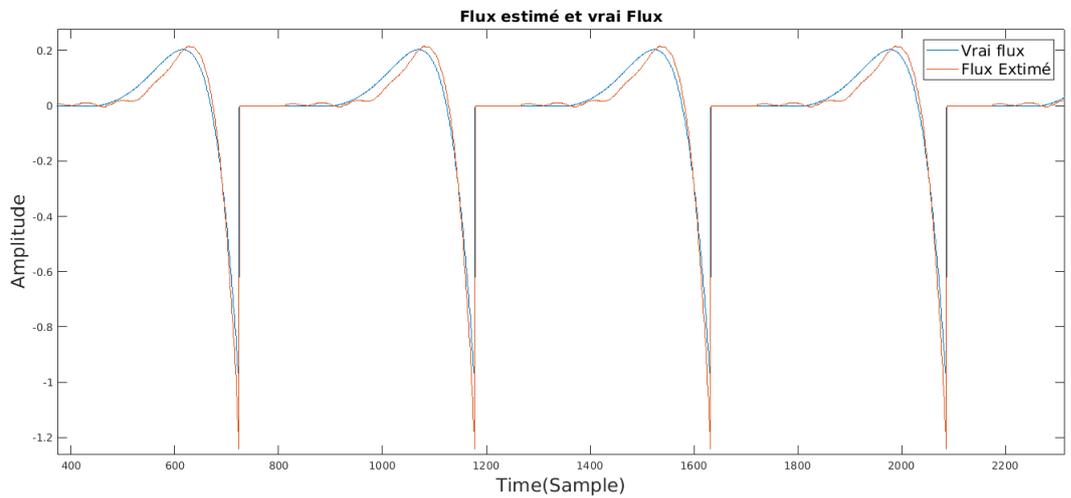


FIGURE 22 – Résultat avec détection automatique des GCIs

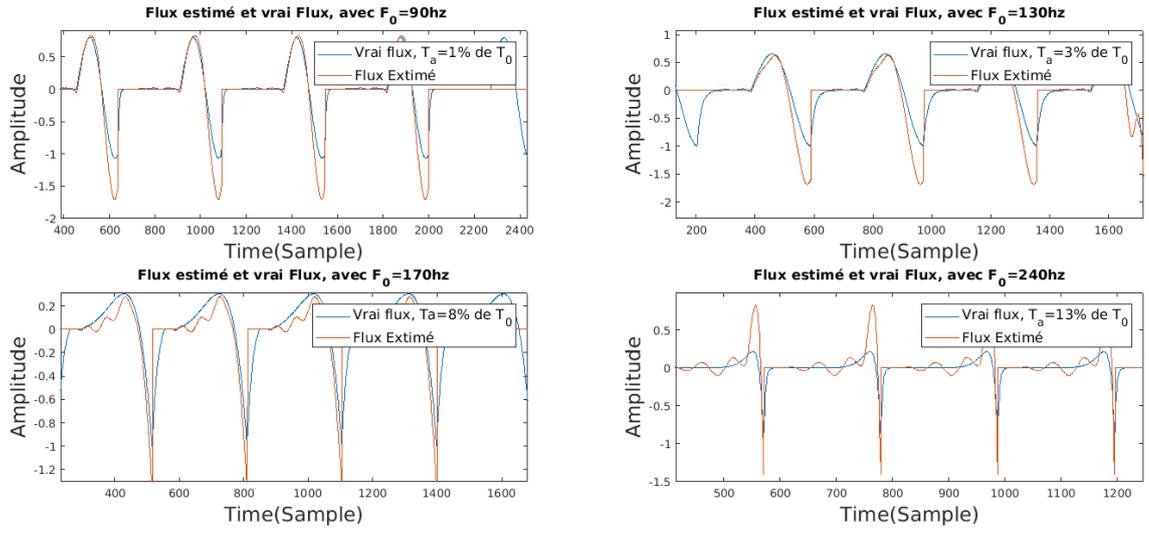


FIGURE 23 – Résultat avec détection automatique des GCIs

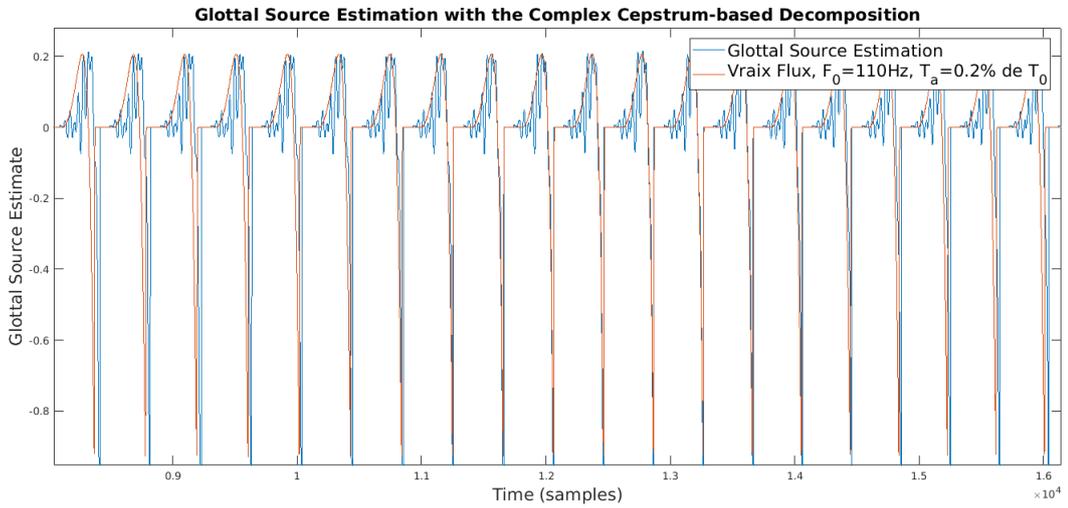


FIGURE 24 – Résultat sur voyelle avec détection automatique des GCIs

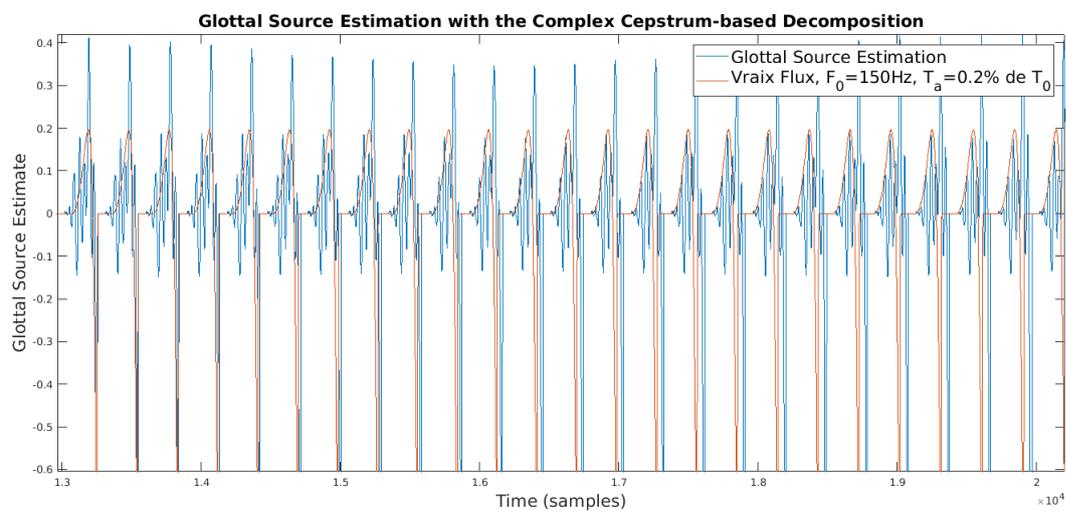


FIGURE 25 – Résultat sur voyelle avec détection automatique des GCIs

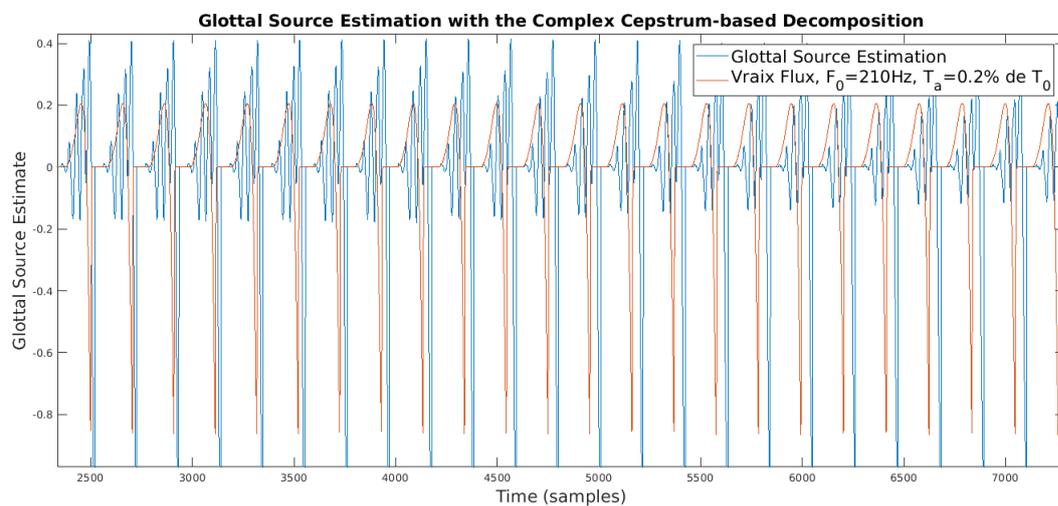


FIGURE 26 – Résultat sur voyelle avec détection automatique des GCIs

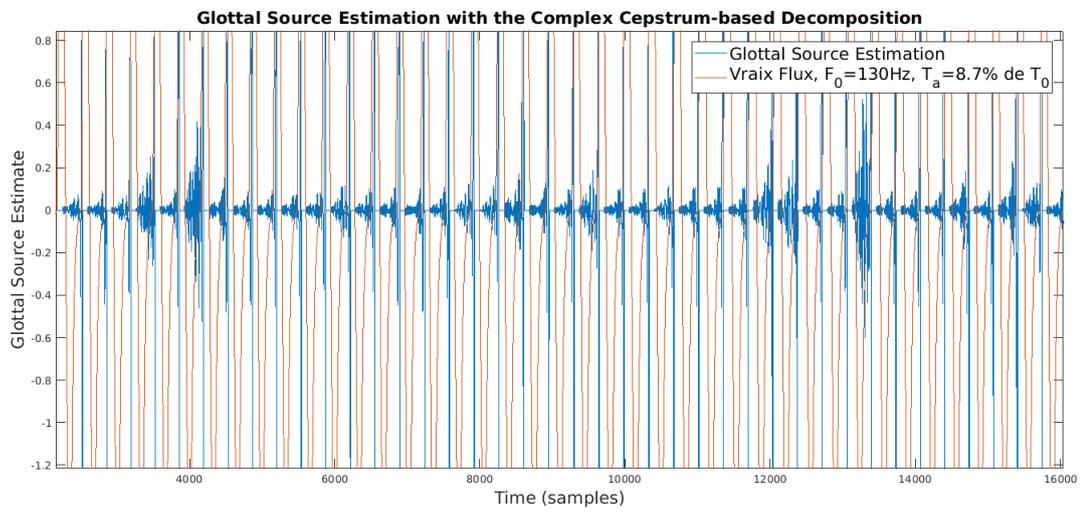


FIGURE 27 – Résultat sur voyelle avec détection automatique des GCIs

CONCLUSION

Pendant ce stage nous avons testé une méthode resente de déconvolution sur des signaux vocaux. A l'aide d'un soft nous avons pu généré des flux puis tester la méthode sur des trames. Nous avons fait varier les paramètres du flux, Et nous avons constaté que la méthode fonctionne que sous certaines conditions. Ces conditions sont liées à la forme du flux et au paramètres T_a . Puis à l'aide du soft nous avons généré une voyelle synthétique. Ensuite nous avons testé la méthodes sur la voyelle. Nous avons aussi constaté que la méthode fonctionne que sous des conditions bien particuliers. Comme la méthode ne fonctionne que sous conditions. C'est a dire, la méthode ne peut estimer toute les variétés possible du flux. Donc On peut dire que d'une façon générale la méthode ne fonctionne pas. Une suite éventuelle de ce stage serait de trouver une amélioration à la méthode.

BIBLIOGRAPHIE

- [1] T. DRUGMAN, B. BOZKURT, T. DUTOIT, *A Comparative Study of Glottal Source Estimation Technique*, *comput. Speech lang.* (2010), doi :10.1016/j.csl.2011.03.003.
- [2] T. DRUGMAN, P. ALKU, A. ALWAN, B. YEGNANARAYANA, *Glottal Source Processing : from Analysis to Applications*, *Comput. Speech Lang.* **28** (2014), 1117-1138.
- [3] T. DRUGMAN, B. BOZKURT, T. DUTOIT, *Complex cepstrum-based decomposition of speech for glottal source estimation*, *interspeech*,. (2009b), 116-119.
- [4] G. FANT, J LILJENCRANTS, Q. LIN, *A four-parameter of parameter model of glottal flow*, *STL-QPSR*, 4 :1-13 (1985).
- [5] A. OPPENHEIM, R. SCHAFER, *Discrete-time signal processing*, Prentice-Hall, chap. 12 (1989).
- [6] H. LU, J. O. SMITH III, *Glottal source modeling for singing voice synthesis*, center for computer Research in music and Acoustics, ICMC (2000).
- [7] S. CHUENWATTANAPRANITHI, S. MANNEWONGVATANA, B. THIPAKORN, , *The Deviation of Glottal Flow Derivative Model due to Emotional Speech*, center for computer Research in music and Acoustics, ICMC (2000).
- [8] C. ALESSANDRO, N. HENRICH, *The Spectrum of Glottal Flow models*, *Acta Acustica United with Acustica* vol.92(2006)1-1.
- [9] D. VINCENT , *Analyse et contrôle du signal glottique en synthèse de la parole*, PhD thèse, Université de Rennes 1 (2007).