Estimation de la fréquence des formants basée sur une transformée en ondelettes complexes

Laurence Cnockaert[*]*, Jean Schoentgen*[†]*et Francis Grenez*

Université Libre de Bruxelles
Faculté des Sciences Appliquées, Service Ondes et Signaux
Av. F.D. Roosevelt 50, 1050 Bruxelles, Belgique
lcnockae@ulb.ac.be

ABSTRACT

The objective of this paper is to evaluate the performances of a formant estimation method in tracking variations due to the vocal tract movement during the production of sustained vowel. The formant frequency estimation is based on the instantaneous frequency obtained by means of a complex wavelet transform and is synchronised with the glottal cycle. Results for synthetic speech signals show that the precision of the formant frequency estimation is high. However, the estimated results are influenced by variations of the vocal frequency and variations of close formants. The method is illustrated for real speech.

1. INTRODUCTION

L'estimation des caractéristiques du conduit vocal à partir du signal de parole est un domaine de recherche important, notamment à cause de son utilité pour la compréhension et la modélisation du mécanisme de production de la parole. Pour décrire le conduit vocal, on mesure généralement les caractéristiques des formants qui sont les pics observés dans le spectre du signal vocal et qui correspondent aux résonances libres dans le conduit vocal. Le but de cet article est de caractériser les variations involontaires de la position des articulateurs lors de la production de voyelles soutenues, par l'intermédiaire des fréquences des formants. Une technique précise de mesure de la fréquence des formants est nécessaire à cet effet.

Les propriétés du conduit vocal varient dans le temps. D'une part, la forme du conduit vocal varie durant la production de la parole à cause des mouvements des articulateurs. D'autre part, des variations apparaissent au rythme du cycle glottique, à cause de la vibration des cordes vocales [6]. En effet, les cordes vocales oscillent entre une phase fermée et une phase ouverte, ce qui modifie les caractéristiques du système : pendant la phase fermée, le conduit vocal est fermé à la glotte et le signal de parole résulte des résonances libres dans le conduit, tandis que pendant la phase ouverte, le conduit vocal est couplé acoustiquement avec la glotte et la trachée, ce qui modifie les résonances du conduit.

Pour obtenir les meilleures performances dans le suivi des variations temporelles des paramètres des formants, il faut donc que les fenêtres d'analyse aient une longueur effective plus courte que le cycle glottique et soient synchronisées sur celui-ci [7]. Pour étudier les variations du conduit vocal supra-glottique, les caractéristiques pertinentes sont les valeurs des formants pendant la phase fermée de la glotte. Notons que, comme nous nous intéressons aux variations des formants pour des voyelles soutenues, le critère de performance de l'estimation des formants est basé sur la qualité du suivi des mouvements des formants, et non sur la proximité de la fréquence estimée des formants par rapport à la consigne.

Dans cet article, nous étudions une méthode d'estimation non-stationnaire des fréquences des formants, basée sur la fréquence instantanée obtenue au moyen d'une transformée en ondelettes complexes, avec synchronisation des mesures par rapport à la phase fermée du cycle glottique. Les performances de la méthode sont illustrées sur des signaux de parole synthétiques. L'effet des variations de la fréquence fondamentale et des fréquences des formants sur les estimations des fréquences des formants est étudié. Finalement, quelques résultats sont présentés pour des signaux de parole réels.

2. ESTIMATION DES FORMANTS

2.1. Transformée en ondelettes continue

La fréquence instantanée $FI(t)$ d'un signal passe-bande $s(t)$ est généralement définie au moyen de sa transformée de Hilbert $H[s(t)]$[1].

$$\Phi(t) = arg[s(t) + jH[s(t)]] \quad (1)$$

$$FI(t) = \frac{1}{2\pi}\frac{d\Phi(t)}{dt} \quad (2)$$

La transformée en ondelettes continue $CWT(\lambda, t)$ permet également de définir la notion de fréquence instantanée, lorsqu'on utilise une ondelette analytique [3].

La transformée en ondelettes continue d'un signal $x(t)$ est définie comme

$$CWT(\lambda, t) = \int_{-\infty}^{+\infty} x(u)\,\frac{1}{\sqrt{\lambda}}\psi^*\left(\frac{u-t}{\lambda}\right)du, \quad (3)$$

où $\psi(t)$ est l'ondelette-mère, et où $CWT(\lambda, t)$ est le coefficient de la transformée en ondelettes pour un facteur d'échelle λ, à l'instant t.

L'amplitude et la phase des coefficients $CWT(\lambda, t)$ complexes, obtenus à partir d'une ondelette-mère complexe, sont respectivement l'enveloppe et la phase instantanée des composantes spectrales du signal dans la bande de fréquence centrée autour de la fréquence centrale f_c de l'on-

[*]Le premier auteur est boursière du *Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture* (Belgique).

[†]Le deuxième auteur est *Maître de Recherches* du *Fonds National pour la Recherche Scientifique* (Belgique).

delette [4]. La dérivée temporelle de la phase des coefficients $CWT(\lambda, t)$ est donc une estimation de la fréquence instantanée du signal dans cette bande de fréquences. Par conséquent, on peut étudier l'évolution de la fréquence instantanée dans différentes bandes de fréquence du signal au moyen des coefficients de la transformée en ondelettes.

Ici, l'ondelette complexe de Morlet a été utilisée [5] :

$$\psi_{\omega_c}(t) = C\, e^{-i\omega_c t}\left[e^{-\frac{t^2}{2\sigma_t^2}} - \sqrt{2}e^{-\frac{\omega_c^2\sigma_t^2}{4}} e^{-\frac{t^2}{\sigma_t^2}}\right]. \quad (4)$$

L'échelle λ de l'ondelette est déterminée par la fréquence centrale $f_c = \frac{\omega_c}{2\pi}$, qui est la fréquence d'oscillation de l'ondelette. Le produit $\omega_c\sigma_t$ fixe le lien entre la largeur de l'enveloppe gaussienne de l'ondelette et sa fréquence d'oscillation f_c. Pour avoir une famille d'ondelettes, le produit $\omega_c\sigma_t$ doit être constant. Le facteur C normalise l'énergie. La durée effective de l'ondelette peut être définie comme $2\sigma_t$. La forme gaussienne de l'enveloppe de l'ondelette de Morlet minimise le produit des résolutions temporelle et fréquentielle de l'ondelette et permet par conséquent d'optimiser la précision des résultats.

2.2. Application à l'estimation des formants

La transformée en ondelettes continue permet donc de calculer la fréquence instantanée pour différentes bandes de fréquences du signal de parole. Au voisinage des fréquences centrales d'ondelettes dont la cyclicité correspond bien à celle du signal, l'amplitude de la transformée en ondelettes présente un maximum. La fréquence instantanée obtenue à partir de la phase des coefficients de la transformée en ondelettes est alors très proche de la cyclicité du signal et permet d'obtenir la fréquence fondamentale F_0 du signal [2]. De même, pour de plus petites échelles, si la fréquence d'un formant se situe dans la bande passante d'une ondelette, la fréquence instantanée résultante sera très proche de la fréquence du formant. La fréquence du formant obtenue à partir de la fréquence instantanée présente une meilleure résolution fréquentielle que le pas de calcul fréquentiel de la transformée en ondelettes [2]. Elle sera donc utilisée ici.

Pour optimiser les résultats, les fréquences des formants sont préalablement estimées par une méthode traditionnelle. Différentes valeurs de $\omega_c\sigma_t$ sont alors utilisées pour calculer une transformée en ondelette distincte et adaptée autour de l'estimation de chaque formant. Pour le premier formant, on veille à ce que la durée effective des ondelettes soit plus courte qu'un cycle glottique. Pour le deuxième formant, il faut que la résolution fréquentielle soit suffisamment fine pour dissocier le deuxième du premier formant. Pour le troisième formant, on choisit la bande passante de l'ondelette égale à 300Hz.

Instants de mesure Obtenir la fréquence instantanée le long des maxima d'amplitude de la transformée en ondelettes ne suffit pas pour obtenir le tracé des fréquences des formants. En effet, la variation au rythme du cycle glottique est encore présente. Il faut donc échantillonner la fréquence instantanée des formants pour en extraire une valeur caractéristique de la phase fermée de chaque cycle glottique. La transformée en ondelettes permet de détecter l'instant de fermeture glottique, qui se caractérise par un maximum d'énergie instantanée dans la transformée en ondelette. L'instant de mesure est donc choisi légèrement après le maximum d'énergie (la moitié de la longueur effective de l'ondelette), afin que l'ondelette d'analyse correspondante se situe dans la phase fermée de la glotte.

3. SIMULATIONS SUR DES SIGNAUX SYNTHÉTIQUES

Dans cette section, nous présentons des résultats illustrant le comportement de la méthode d'extraction des formants sur des signaux synthétiques. Le but de ces simulations est de mettre en évidence et de comprendre la précision et les limites de la méthode lorsqu'elle est appliquée à des signaux de parole réalistes.

3.1. Signaux synthétiques

Les signaux synthétiques sont basés sur un modèle source - conduit. Le signal de source est donné par la dérivée temporelle du modèle de débit glottique de Liljencrants et Fant. Le conduit est obtenu par une cascade de filtres IIR du second ordre variables dans le temps, modélisant chacun un formant. Pour modéliser l'interaction source-conduit, la bande passante des formants est modulée de façon synchronisée avec la source. Deux valeurs différentes de bandes passantes caractérisent donc la phase ouverte et la phase fermée de la glotte.

3.2. Résultats

Pour étudier l'influence des paramètres du signal synthétique sur les fréquences de formants mesurées, les cas présentés sont les suivants :
– F_0 fixe, fréquences des formants fixes,
– F_0 fixe, fréquences des formants variables,
– F_0 variable, fréquences des formants fixes.
Les performances sont évaluées sur base de la qualité du suivi des mouvements des formants, et non de la proximité entre la fréquence estimée des formants et la consigne.

Fréquence fondamentale fixe, fréquences des formants fixes Les figures 1 à 3 illustrent l'estimation des fréquences des formants pour un signal synthétique dont la fréquence fondamentale F_0 est de $120Hz$, et les fréquences des formants de $700Hz, 1200Hz$ et $2500Hz$. Les bandes passantes de tous les formants sont de $100Hz$ et de $150Hz$, pour les phases fermées et ouvertes de la glotte.

La figure 1 montre le signal synthétique, ainsi que l'amplitude de sa transformée en ondelettes, pour $\omega_c\sigma_t = 10$. Les étoiles blanches marquent les fréquences estimées des formants. Les pics aux fréquences des formants apparaissent clairement, ainsi que les moment d'excitation où de l'énergie est présente à toutes les fréquences.

La figure 2 montre une coupe de l'amplitude de la transformée en ondelette et de la fréquence instantanée en fonction de la fréquence centrale des ondelettes, pour un instant donné. Les traits pointillés marquent les fréquences de consigne des trois formants. Les lignes verticales montrent les fréquence centrales des ondelettes pour lesquelles il y a un maximum d'amplitude de la transformée en ondelettes. On peut voir les plateaux de la fréquence instantanée aux fréquences des formants, qui permettent d'obtenir la précision fréquentielle des mesures.

Le choix des instants de mesure est illustré à la figure 3. On y voit l'évolution des énergies des trois formants,

qui présentent un maximum par cycle glottique, ainsi que l'évolution des formants instantanés. Les losanges rouges marquent les fréquences estimées des formants. L'amplitude de la variation de celles-ci est inférieure à $0.2Hz$. Les variations obtenues pour d'autres valeurs de F_0 et de formants ont le même ordre de grandeur.

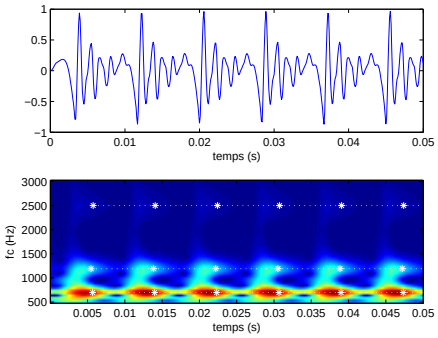

FIG. 1: Signal synthétique et amplitude de sa transformée en ondelettes pour $\omega_c\sigma_t = 10$. Les étoiles blanches marquent les fréquences estimées des formants. Les grandes amplitudes sont représentées en rouge, les faibles amplitudes en bleu.

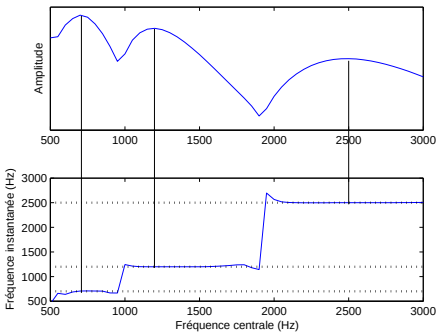

FIG. 2: Coupe de l'amplitude de la transformée en ondelette et de la fréquence instantanée en fonction de la fréquence centrale des ondelettes, pour un instant donné. Les pointillés marquent les fréquences des trois formants. Les lignes verticales montrent les fréquence centrales des ondelettes pour lesquelles il y a un maximum d'amplitude de la transformée en ondelettes.

Fréquence des formants variables Pour tester les performances de la méthode lorsque la fréquence des formants varie, des signaux synthétiques ont été générés avec une fréquence de formant variant linéairement.

Le tableau 1 montre les résultats pour des signaux synthétiques dont la fréquence de F_1 varie, pour deux valeurs de F_0 et deux valeurs de la fréquence de F_2. La fréquence de F_1 varie entre $700Hz$ et $725Hz$, la fréquence de F_2 est de $1100Hz$ ou $1200Hz$, et la fréquence de F_3 est de $2500Hz$. La fréquence fondamentale F_0 est de $100Hz$ ou $125Hz$. Dans le tableau 1, la première partie donne les variations maximales des écarts entre la fréquence mesurée de F_1 et sa consigne. La suite du tableau donne les variations maximales des estimations de F_2 et F_3.

La première partie du tableau 1, montre que, dans tous les

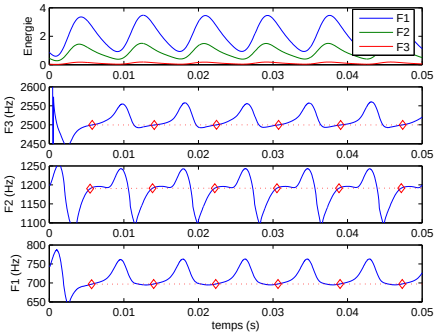

FIG. 3: Evolution de l'énergie des formants et des fréquences instantanées des trois premiers formants. Les losanges rouges marquent les fréquences estimées des formants.

cas, la fréquence estimée de F_1 suit bien la consigne. L'erreur est de l'ordre de $1Hz$, ce qui dépasse légèrement l'erreur obtenue pour un signal à formants fixes. La deuxième partie du tableau 1 montre que la mesure de F_2 est influencée par la variation de F_1, d'autant plus que l'écart fréquentiel entre les deux formants est petit et d'autant plus que F_0 est grande. La mesure de F_3 n'est pas influencé par les variations de F_1.

Les résultats de simulations similaires avec F_2 ou F_3 variable donnent les mêmes conclusions : Premièrement, plus F_0 est élevée, plus l'écart entre le formant variable et la consigne varie. Deuxièmement, plus le formant variable est proche du formant estimé, plus celui-ci varie également.

TAB. 1: Précision de l'estimation des formants pour des signaux synthétiques dont F_1 varie linéairement, pour deux valeurs de F_0 et pour deux valeurs de F_2.

F1 : variation maximale de l'écart entre la mesure et la consigne		
	F0=100Hz	F0=125Hz
F2=1100Hz	0.35Hz	0.56Hz
F2=1200Hz	0.57Hz	1.20Hz
F2 : variation maximale		
	F0=100Hz	F0=125Hz
F2=1100Hz	1.99Hz	8.60Hz
F2=1200Hz	0.27Hz	0.39Hz
F3 : variation maximale		
	F0=100Hz	F0=125Hz
F2=1100Hz	0.004Hz	0.03Hz
F2=1200Hz	0.004Hz	0.03Hz

Fréquence fondamentale variable Pour tester l'effet de la variation de la fréquence fondamentale F_0, des signaux synthétiques ont été générés avec F_0 variant linéairement.

La figure 4 montre l'évolution de F_0 et des fréquences estimées des formants pour un signal synthétique dont F_0 varie linéairement entre $95Hz$ et $105Hz$. Les fréquences de consigne des formants sont de $700Hz$, $1200Hz$ et $2500Hz$. On peut voir que la fréquence des formants n'est pas parfaitement stable et varie en fonction de la proximité de la fréquence du formant avec les harmoniques de

F_0. L'effet est plus marqué pour les fréquences de formant plus faibles, mais reste néanmoins inférieur à $2Hz$.

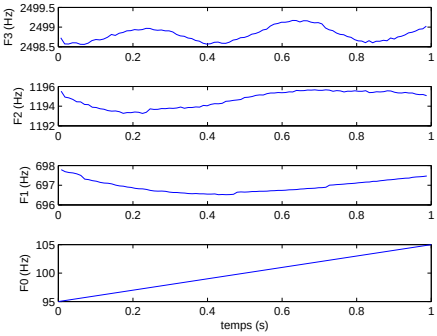

FIG. 4: Evolution de F_0 et des fréquences estimées des trois premiers formants, pour un signal synthétique dont la fréquence fondamentale varie linéairement.

4. APPLICATION À DES SIGNAUX RÉELS

La figure 5 montre une voyelle [a] soutenue et l'amplitude de sa transformée en ondelettes avec le paramètre $\omega_c\sigma_t = 10$. Les étoiles blanches correspondent aux fréquences estimées des formants. La figure 6 montre l'énergie instantanée des formants et les fréquences instantanées des trois premiers formants. Les losanges rouges marquent les fréquences estimées des formants. La figure 7 montre la fréquence fondamentale et les fréquences des formants obtenus pour la même voyelle soutenue, pour une durée plus longue. On constate que la méthode permet de détecter et de suivre convenablement les formants.

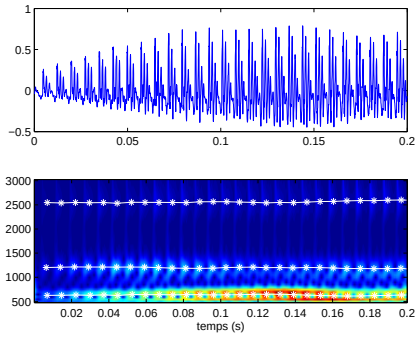

FIG. 5: Signal réel et amplitude de sa transformée en ondelettes pour $\omega_c\sigma_t = 10$. Les étoiles blanches marquent les fréquences estimées des formants. Les grandes amplitudes sont représentées en rouge, les faibles en bleu.

5. CONCLUSION

Une méthode d'estimation des fréquences des formants a été proposée. Elle est basée sur la fréquence instantanée obtenue au moyen d'une transformée en ondelettes complexes et est synchronisée par rapport au cycle glottique. Les performances de la méthode d'estimation des formants ont été évaluées pour le suivi des variations dues au mouvement du conduit vocal. Les résultats obtenus sur des signaux synthétiques montrent que la précision de l'estimation de la mesure des formants est très bonne. On

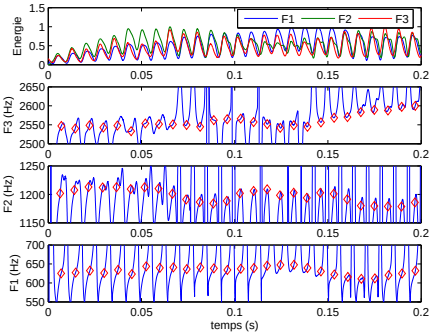

FIG. 6: Evolution de l'énergie des formants et des fréquences instantanées des trois premiers formants. Les losanges rouges marquent les fréquences estimées des formants.

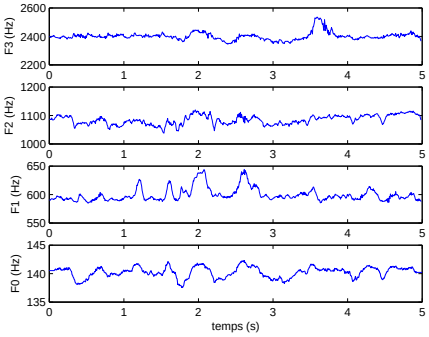

FIG. 7: Evolution de F_0 et des fréquences des trois premiers formants pour un signal réel.

constate cependant une influence des variations de la fréquence fondamentale et de la variation des autres formants proches en fréquence.

RÉFÉRENCES

[1] B. Boashash. Estimation and interpreting the instantaneous frequency of a signal - part i : Fundamentals. *Proceedings of the IEEE*, 80(4) :520 – 539, 1992.

[2] L. Cnockaert, F. Grenez, and J. Schoentgen. Fundamental frequency estimation and vocal tremor analysis by means of morlet wavelet transforms. *Proc. ICASSP, Philadelphia (USA)*, pages 393–396, 2005.

[3] T. Le-Tien. Some issues of wavelet functions for instantaneous frequency extraction in speech signals. *Proc. IEEE Tencon 1997*, pages 31–34, 1997.

[4] St. Mallat. *A Wavelet Tour of Signal Processing*. San Diego : Academic Press, 2nd edition, 1999.

[5] D.B. Percival and A.T. Walden. *Wavelet methods for time series analysis*. Cambridge University Press, 2000.

[6] Pr. Rao and A. D. Barman. Speech formant frequency estimation : evaluating a nonstationary method. *Signal Processing*, 80(8) :1655–1667, 2000.

[7] B. Yegnanarayana and R.N.J. Veldhuis. Extraction of vocal tract system characteristics from speech signals. *IEEE trans. on speech and audio processing*, 6(4) :313–327, july 1998.