

# Suivi de formants par détection de crêtes d'ondelettes

Salma Châari<sup>1</sup>, Kaïs Ouni<sup>1,2</sup>, Nouredine Ellouze<sup>1</sup>

<sup>1</sup>Unité de Recherche Traitement du Signal, Traitement d'Images et Reconnaissance de Formes  
Ecole Nationale d'Ingénieurs de Tunis (ENIT)  
BP.37, Le Belvédère, 1002, Tunis, Tunisie

<sup>2</sup>Institut Supérieur des Technologies Médicales de Tunis (ISTMT)  
Avenue Docteur Zouheir Essafi, 1006, Tunis, Tunisie  
kais.ouni@enit.rnu.tn - n.ellouze@enit.rnu.tn

## ABSTRACT

This paper proposes two new approaches for formant tracking using Fourier and wavelet ridges. The speech signal is decomposed into Time-Frequency representations issued from windowed Fourier transform and wavelet transform. Formant tracking is achieved by exploring ridges from time-frequency representation and imposing continuity constraints on formant trajectories. These approaches are validated by their application on synthesized vowels. The formant tracking shows a reliable estimation in comparison with the values given in synthesis. In the same way, the two analyses are applied on natural voiced speech. All the results are compared to two traditional methods using some type of LPC analysis. Fourier ridge detection analysis is shown to be a useful tool for formant tracking compared to traditional methods.

## 1. INTRODUCTION

Etant donné l'intérêt potentiel des formants en tant que porteurs fondamentaux de l'information dans la parole, le suivi de formants peut jouer un rôle déterminant dans l'identification des voyelles et autres sons vocaliques, le pilotage des synthétiseurs à formants [1], et dans quelques cas, l'alimentation d'un système de reconnaissance automatique [2]. De nombreux efforts ont été consacrés au développement de techniques de suivi de formants. Ces techniques reposent communément sur la détection des pics spectraux obtenus par les coefficients de prédiction linéaire (LPC) [3]. La nature et la complexité du problème expliquent l'adoption de la programmation dynamique par d'autres algorithmes [4]. D'autres méthodes utilisent les maxima d'énergie sur un spectrogramme [5]. Des techniques issues d'études sur la perception [2] ainsi que des approches probabilistes, particulièrement les modèles de Markov cachés [1], ont été de même proposées.

L'objet de ce papier est le développement et l'étude de deux techniques de suivi de formants utilisant les crêtes de Fourier à fenêtre et les crêtes d'ondelettes. En effet, la parole est un signal dont la composition fréquentielle varie avec le temps. De telles évolutions temps-fréquence (TF) sont mises en évidence en décomposant le signal en fonctions élémentaires bien concentrées en temps et en fréquence, appelées atomes TF [6]. La transformée de Fourier à fenêtre et la transformée en ondelettes sont deux

exemples importants de décomposition TF [6]. Une trajectoire spectrale produit après transformation des coefficients de grande amplitude à des fréquences qui dépendent du temps. On analyse alors l'évolution de ces trajectoires en suivant la localisation des coefficients de grande amplitude, appelés points de crête [6]. Comme les formants sont des crêtes spectrales qui correspondent aux fréquences de résonance du conduit vocal, il s'agit alors de détecter les points de crête dans le plan TF, et d'étudier leur validation comme étant des formants en voisement. Les deux approches proposées utilisent une représentation du signal de parole de type spectrogramme, dans le cas de l'analyse de Fourier, et de type scalogramme, dans le cas de l'analyse en ondelettes. Ces deux techniques ont été validées sur des exemples de voyelles synthétiques et naturelles, puis évaluées par comparaison des résultats à deux méthodes classiques basées sur une analyse LPC.

Ce papier est organisé de la manière suivante : dans la partie 2, on présente l'algorithme de crêtes. Puis, dans la partie 3, on explicite les deux algorithmes mis en oeuvre. Quelques résultats sont donnés dans la partie 4 avant de conclure et de présenter des perspectives dans la partie 5.

## 2. ALGORITHME DE CRETES

### 2.1. Fréquence instantanée

Un signal  $f$  appartenant à  $L^2(\mathfrak{R})$  est analytique si sa transformée de Fourier  $\hat{f}$  est nulle sur les fréquences négatives. La partie analytique  $f_a$  d'un signal réel  $f$  est donnée par sa transformée de Fourier  $\hat{f}_a$  [6] :

$$\hat{f}_a(\omega) = \begin{cases} 2\hat{f}(\omega) & \text{pour } \omega \geq 0 \\ 0 & \text{pour } \omega < 0 \end{cases} \quad (1)$$

et  $f_a$  peut se décomposer en module et phase complexe [6]:

$$f_a(t) = a(t)e^{i\phi(t)} \quad (2)$$

Comme  $f = \text{Re}[f_a]$ , on en déduit :  $f(t) = a(t)\cos(\phi(t))$  (3)

La fréquence instantanée est définie comme la dérivée de la phase [6] :

$$\omega(t) = \phi'(t) \quad (4)$$

## 2.2. Détection des crêtes

### Crêtes de Fourier à fenêtre

Les atomes de Fourier à fenêtre [6] sont construits en translatant de  $u$  et en modulant à la fréquence  $\xi$  une fenêtre réelle et symétrique  $g(t)$ . Il a été démontré [6] que la fréquence instantanée du signal traité  $f$  est reliée à la transformée de Fourier à fenêtre  $Sf(u, \xi)$  si  $\xi \geq 0$  :

$$Sf(u, \xi) = \frac{\sqrt{s}}{2} a(u) e^{i[\phi(u) - \xi u]} \left( \hat{g} \left[ s \left[ \xi - \phi'(u) \right] \right] + \varepsilon(u, \xi) \right) \quad (5)$$

où  $s$  est la dilatation initiale appliquée à la fenêtre  $g$ ,  $\hat{g}$  est la transformée de Fourier de  $g$ , et  $\varepsilon(u, \xi)$  un terme de correction. Comme  $|\hat{g}(\omega)|$  est maximum en  $\omega=0$ , l'équation (5) montre que pour  $u$  fixé et  $\xi$  variable, le spectrogramme  $|Sf(u, \xi)|^2$  est maximum en  $\xi(u) = \phi'(u)$ . Les maxima du spectrogramme aux points  $(u, \xi(u))$  forment les crêtes de Fourier à fenêtre. Si l'amplitude et la fréquence instantanée varient peu à l'intérieur de la fenêtre de Fourier et si la fréquence instantanée est supérieure à la bande passante de la fenêtre, de manière à pouvoir négliger le terme correctif  $\varepsilon(u, \xi)$  dans (5), alors les fréquences aux points de crêtes sont assimilables aux fréquences instantanées  $\xi(u) = \phi'(u)$  [6].

Les crêtes de Fourier d'un signal analytique contenant plusieurs composantes spectrales de fréquences suffisamment éloignées permettent de les séparer, et les lignes de crête décrivent l'évolution temporelle de chaque composante spectrale [6]. Ainsi, les crêtes de Fourier à fenêtre constituent un outil de détection de fréquences instantanées lorsque celles-ci ne sont pas trop proches.

### Crêtes d'ondelettes

La même approche est adoptée dans le cas des ondelettes analytiques comme les ondelettes de Gabor [6], afin de séparer la phase et l'amplitude. Soit une ondelette complexe définie par :

$$\psi(t) = e^{i\eta t} g(t) \quad (6)$$

avec  $\eta$  est le centre fréquentiel de sa transformée de Fourier  $\hat{\psi}$ , et  $g(t)$  la fenêtre utilisée, réelle et symétrique est translatée de  $u$  et modulée par la fréquence  $\xi$  [5]. Une famille d'atomes TF  $\psi_{u,s}$  s'obtient en dilatant l'ondelette  $\psi$  par un facteur  $s$ , et en la translatant par  $u$ . Pour  $f(t) = a(t) \cos(\phi(t))$ , le scalogramme normalisé  $P_w f$  est égal à [6] :

$$\frac{\xi}{\eta} P_w f(u, \xi) = \frac{1}{4} a^2(u) \left| \hat{g} \left( \eta \left[ 1 - \frac{\phi'(u)}{\xi} \right] \right) + \varepsilon(u, \xi) \right|^2 \quad (7)$$

où  $\varepsilon(u, \xi)$  est un terme correctif. Comme  $\hat{g}$  est maximum en  $\omega=0$ , l'expression (7) montre que, si l'amplitude et la fréquence instantanée varient peu sur le support de  $\psi_{u,s}$  et si la fréquence instantanée est supérieure à la bande passante de  $\hat{g}$  de manière à pouvoir négliger le terme

correctif  $\varepsilon(u, \xi)$ , alors le scalogramme atteint son maximum lorsque  $\frac{\eta}{s(u)} = \xi(u) = \phi'(u)$ . Les maxima

correspondants aux points  $(u, \xi(u))$  sont appelés des crêtes d'ondelettes [6]. Les fréquences où le scalogramme normalisé est maximal sont assimilables aux fréquences instantanées dans la limite de la résolution de la transformée. Les crêtes d'ondelettes d'une somme de signaux analytiques permettent de distinguer les fréquences instantanées respectives si les différences relatives des fréquences sont supérieures à la largeur de bande relative de l'ondelette. Dans ces conditions, les crêtes d'ondelettes constituent un outil de détection de fréquences instantanées lorsque celles-ci ne sont pas trop proches [6].

## 3. ALGORITHME DE SUIVI DE FORMANTS

### 2.1. Cas de l'analyse de Fourier

L'algorithme de suivi de formants utilise les crêtes détectées dans le plan TF. Le signal de parole est tout d'abord pré-accentué pour rehausser les hautes fréquences. Ensuite, une analyse du signal est opérée par spectrogramme. Puis, l'algorithme de crêtes calcule les fréquences instantanées à partir des maxima locaux du spectrogramme. La phase suivante est la détection de tous les maxima locaux du spectrogramme, suivie par une procédure d'interpolation parabolique de ces maxima et un seuillage pour mieux localiser les crêtes de Fourier. Les fréquences instantanées ainsi calculées à partir des points de crêtes forment les fréquences candidates aux formants.

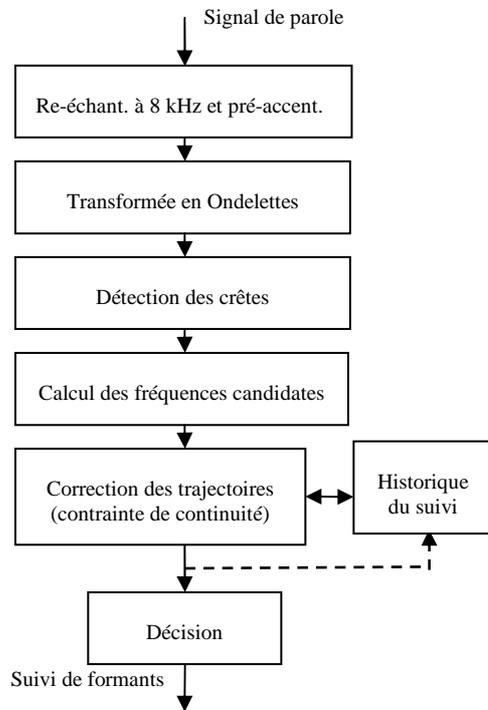
L'observation du fait que les formants varient en général lentement en fonction du temps a mené à imposer des contraintes de continuité au processus de sélection des formants [1] [4]. La contrainte choisie dans ce travail est basée sur le calcul de la distance euclidienne entre chaque combinaison candidate et la combinaison de formants précédemment sélectionnée. Après avoir récupéré les fréquences candidates, chaque combinaison possible de celles-ci va être considérée. Puis, la distance entre chaque combinaison candidate et la combinaison précédente choisie sera calculée, et celle correspondant à la distance minimale sera sélectionnée comme étant la combinaison des formants suivante.

Comme nous nous sommes intéressés à obtenir les trois premiers formants et comme F3 est connu d'être inférieur à 4kHz [1], il est avantageux de re-échantillonner le signal à 8kHz afin d'éviter l'obtention de formants candidats au-dessus de 4kHz, et nous permettre d'opérer une analyse d'ordre inférieur qui offre moins de temps de calcul [1].

### 2.2. Cas de l'analyse en ondelettes

Une analyse du signal de parole est opérée par scalogramme. Les différentes étapes de cet algorithme sont résumées dans le schéma donné par la figure (1).

L'algorithme de crêtes calcule les fréquences instantanées à partir des maxima locaux du scalogramme. L'algorithme de suivi de formants trouve l'ensemble des fréquences candidates aux formants. La même approche adoptée dans le cas de l'analyse de Fourier pour le processus de sélection des formants est aussi utilisée dans ce deuxième algorithme basé sur l'analyse en ondelettes.



**Figure 1 :** Algorithme de suivi de formants par détection de crêtes d'ondelettes.

#### 4. RÉSULTATS

Nous avons appliqué les deux algorithmes proposés sur des voyelles synthétiques. Dans le cas de l'analyse de Fourier, la représentation des signaux s'opère par spectrogramme en étudiant l'apport de cinq fenêtres analysantes : la fenêtre rectangulaire, Hamming, gaussienne, Hanning et Blackman. La largeur de la fenêtre considérée est de 10 ms et le pitch est pris égal à 100 Hz. De même, nous avons envisagé d'utiliser différentes ondelettes mères en analyse multirésolution, à savoir l'ondelette de Gabor, Morlet et Chapeau Mexicain, afin d'étudier l'apport de chacune d'entre elles. Les résultats de l'estimation des trajectoires des formants de la voyelle synthétique /a/, dans le cas des fenêtres analysantes de type rectangulaire et Hamming, sont respectivement donnés par les figures (2) et (3). La figure (4) illustre le scalogramme et les trajectoires formantiques estimées, associées à cette voyelle, en utilisant une ondelette mère de type Gabor. La comparaison des performances de cinq fenêtres d'analyse dans le cas de Fourier, a montré qu'il est particulièrement important de s'assurer que la transformée de Fourier de la fenêtre d'analyse utilisée ait des lobes latéraux de faible amplitude. Les fenêtres de Hamming, Hanning, Blackman et gaussienne ont fourni une bonne qualité de suivi. D'autre part, dans le cas de l'analyse

multirésolution, parmi les trois ondelettes mères choisies, l'ondelette de Gabor fournit la meilleure qualité de suivi.

Afin d'évaluer les performances des deux méthodes proposées sur des entités de parole naturelle extraites de la base TIMIT, une étude comparative avec deux autres méthodes de suivi de formants a été menée. La première méthode est basée sur le spectre LPC (WinSnoori) [3][5], la deuxième combine les bancs de filtres avec les coefficients LPC [7]. La comparaison de la qualité de l'estimation est vérifiée au moyen des représentations spectrographiques. Les figures (5), (6), (7) et (8) donnent un exemple de suivi. Cet exemple illustre respectivement le suivi des trois premiers formants du mot "carry" par détection de crêtes de Fourier, par détection de crêtes d'ondelettes, par spectre LPC, et finalement par la méthode LPC combinée au banc de filtres. Cette étude comparative a montré que le suivi par spectre LPC (WinSnoori), par bancs de filtres combinés aux coefficients LPC et par détection de crêtes de Fourier fournit des estimations qui se rapprochent. Cependant, la méthode de suivi par détection de crêtes d'ondelettes est plus ou moins capable de fournir une estimation fidèle des trajectoires formantiques, à cause des oscillations des courbes de suivi.

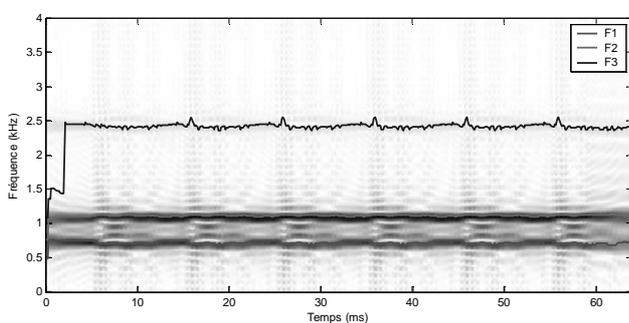
#### 5. CONCLUSION ET PERSPECTIVES

Nous avons présenté dans ce papier deux techniques de suivi des trois premiers formants des signaux de parole. La première technique se base sur la détection des crêtes de Fourier à fenêtre à partir des spectrogrammes déterminés par la transformée de Fourier à fenêtre. La deuxième technique se base sur la détection des crêtes d'ondelettes à partir des scalogrammes issus de la transformée en ondelettes analytiques. La première méthode montre l'intérêt de la détection des crêtes de Fourier à fenêtre dans le suivi des trajectoires formantiques. Cependant, la deuxième méthode s'est montrée plus ou moins capable de fournir une estimation fidèle des trajectoires formantiques. En perspectives, nous proposons d'introduire de nouveaux paramètres en vue d'éliminer au maximum les points de crête indésirables. Une correction des trajectoires par lissage s'avère aussi intéressante.

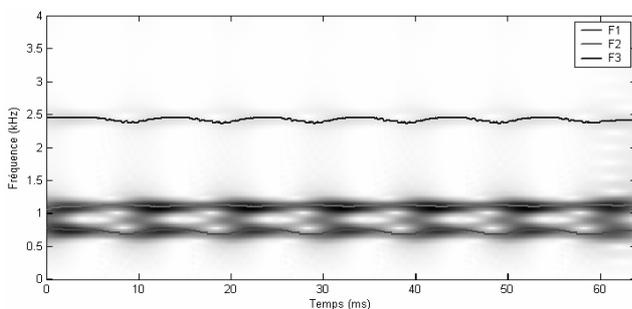
#### BIBLIOGRAPHIE

- [1] A. Acero. Formant analysis and synthesis using Hidden Markov Models. In *Proc. of the Eurospeech Conference*, Budapest, 1999.
- [2] J A. M. A. Ali, J. V. der Spiegel and P. Mueller. Robust auditory-based speech processing using the average localized synchrony detection. *IEEE Trans. on Speech and Audio Processing*, 2002.
- [3] S. McCandless. An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra. *IEEE ASSP-22*, volume 2, pages 135-141, April 1974.

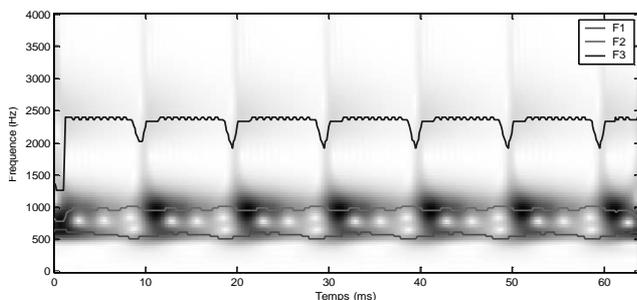
- [4] K. Xia and C. Espy-Wilson. A new strategy of formant tracking based on dynamic programming. In *Proc. Intl. Conf. on Spoken Language Processing*, Beijing, Chine, October 2000.
- [5] Y. Laprie. A concurrent curve strategy for formant tracking. In *Proceedings of ICSLP*, Jegu, Corée du Sud, 2004.
- [6] S. Mallat. Une exploration des signaux en ondelettes. *Les Editions de l'Ecole Polytechnique*, Novembre 2000.
- [7] K. Mustapha and I.C. Bruce. Robust formant tracking for continuous speech with speaker variability. In *IEEE Transactions on Speech and Audio Processing*, January 2005.



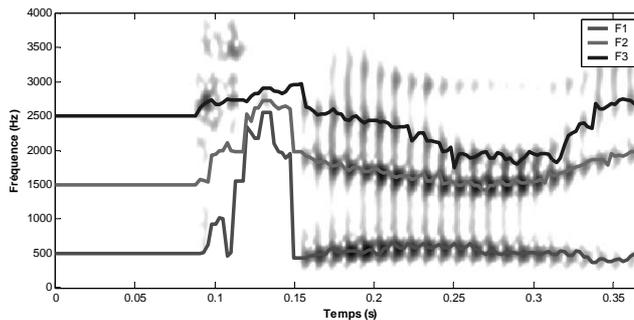
**Figure 2 :** Trajectoires formantiques estimées, superposées au spectrogramme à bande large de la voyelle /a/ avec une fenêtre rectangulaire.



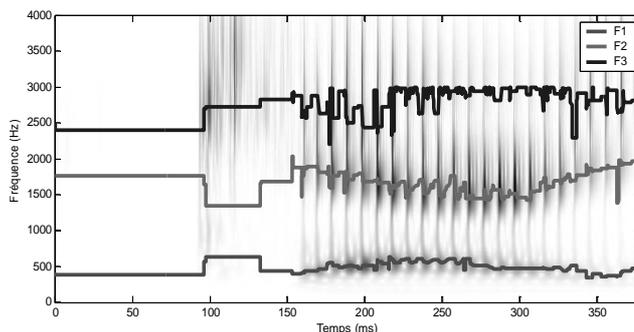
**Figure 3 :** Trajectoires formantiques estimées, superposées au spectrogramme à bande large de la voyelle /a/ avec une fenêtre de Hamming.



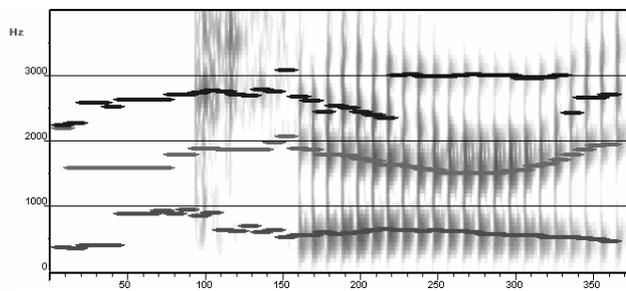
**Figure 4 :** Trajectoires formantiques estimées, superposées au scalogramme représentant la voyelle /a/ avec une ondelette analysante de Gabor.



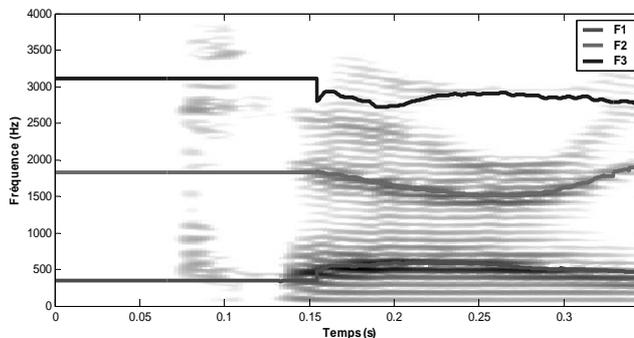
**Figure 5 :** Estimation des trajectoires formantiques du mot "carry" par détection de crêtes de Fourier (fenêtre de Hamming).



**Figure 6 :** Estimation des trajectoires formantiques du mot "carry" par détection de crêtes d'ondelettes (ondelette de Gabor).



**Figure 7 :** Estimation des trajectoires formantiques du mot "carry" par la méthode LPC (WinSnoori).



**Figure 8 :** Estimation des trajectoires formantiques du mot "carry" par la méthode LPC combinée au banc de filtres.

