Wavelet Ridge Track Interpretation in Terms of Formants

Salma Châari, Kaïs Ouni and Noureddine Ellouze.

Unité de Recherche Traitement du Signal, Traitement d'Image et Reconnaissance de Formes
ENIT, BP.37, Le Belvédère, 1002 Tunis, Tunisia
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.

Index Terms: formant tracking, speech analysis, Fourier ridges, wavelet ridges

1. Introduction

Given the potential interest of the first few formants as primary information carriers in human speech, automatic formant trackers have attracted a great deal of interest in many areas of speech processing. In fact, robust formant tracks are utilized to identify vowels [1] and other vocalic sounds [2] [3], to pilot formant synthesizers [4] and in some cases to provide a speech recognition with additional data [5].

Although automatic formant tracking has a wide range of applications, it is still an open problem in speech analysis. Numerous works have been dedicated to develop algorithms for estimating formant frequencies from the acoustic signal. Traditional methods for estimating formants include peak-picking to identify formants from the spectral envelope and linear predictive coding (LPC) analysis which identifies poles corresponding to formant resonances [4]. Nature and complexity of the problem explain the success of dynamic programming algorithms [4][5]. These algorithms utilize dynamic programming in the evaluation of transition costs between two frames. Other algorithms show how active curves could be used to track formants [6][11]. The underlying idea is to deform initial rough estimates of formants under the influence of the spectrogram to get regular tracks close to lines of spectral maxima which are potential formants. Formant tracking algorithms based on auditory models [2] as well as on probabilistic approaches, particularly the Hidden Markov Models (HMM) [1], have been proposed in the same way.

In this paper, two new approaches to interpret Fourier ridge and wavelet ridge tracks in terms of formants are described. Indeed, spectral components of speech signals vary in time as the articulators change position. Such time-frequency (TF) evolution is highlighted by signal decomposition into in elementary functions well concentrated in time and in frequency, termed TF atoms [7]. Windowed Fourier transform and wavelet transform are two important examples of TF decomposition [7]. A spectral line creates high amplitude windowed Fourier at time varying frequencies. The time evolution of such spectral components is therefore analyzed by following the location of large amplitude coefficients, called ridge points [7]. Since formants are spectral peaks that correspond to the resonance frequencies of the vocal tract, we detect ridge points in the TF plane, and we interpret ridge tracks in terms of formants during voiced speech segments. We experimented two approaches for formant tracking. The first consists in a representation of the speech signal using spectrogram. The second consists in a representation of the speech signal using scalogram. Two available techniques based on a LPC analysis have been selected in order to test and compare their performance with that of the formant tracking algorithms proposed in this work.

This paper is organized as follows. Section 2 describes the ridge algorithm, Section 3 presents the proposed formant frequency estimation algorithms, Section 4 describes the experimental results, and Section 5 gives the conclusions and the future work.

2. Formant tracking algorithm

A block diagram showing the main stages of the formant tracking algorithm is presented in Figure 1. Each element of the block diagram is briefly described below.

2.1. Re-sampling and pre-emphasize

The algorithm was implemented for speech signals sampled at Fs = 8 kHz. A first order pre-emphasis filter of the form $1-0.98Z^{-1}$ is applied to partially compensate for the speech source.

2.2. Transformation

Two TF signal decompositions were separately applied to the proposed formant tracker. Then we implement two algorithms. In the first one the processed signal is represented by its spectrogram computed by applying the windowed Fourier transform, while in the second one the processed signal is represented by its scalogram computed by applying the wavelet transform.

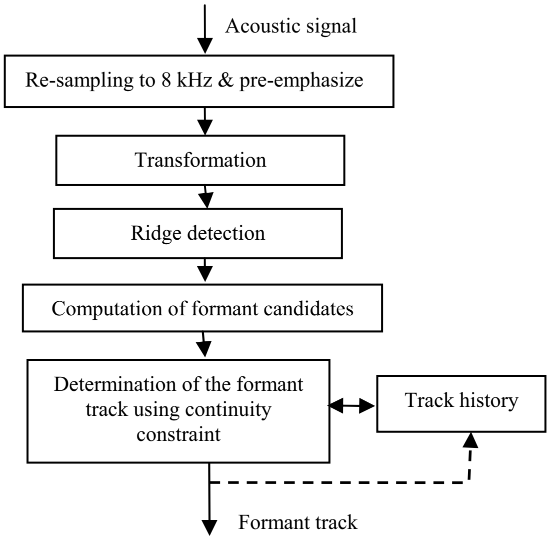
Acoustic signal
Re-sampling to 8 kHz & pre-emphasize
Transformation
Ridge detection
Computation of formant candidates
Determination of the formant track using continuity constraint
Track history
Formant track

Figure 1 *Block diagram of formant tracking algorithm by Fourier/wavelet ridge detection.*

2.3. Ridge detection

We shall begin by defining the windowed Fourier ridges and the wavelet ridges.

2.3.1. Windowed Fourier ridges

Windowed Fourier transform family of atoms [7] is generated by time translations and frequency modulations of a real and symmetric window $g(t)$. This atom has a frequency center ξ and is symmetric with respect to *u*.

Let $f(t) = a(t)\cos(\phi(t))$. It has been demonstrated [7] that the instantaneous frequency of the processed signal *f* is related to the window Fourier transform $Sf(u,\xi)$ if $\xi \geq 0$

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

where *s* is the scaling which has been applied to the Fourier window *g*, $\hat{g}$ is the Fourier transform of *g* and $\varepsilon(u,\xi)$ is the corrective term. Since $|\hat{g}(\omega)|$ is maximum at $\omega=0$, (5) shows that for each *u* the spectrogram $|Sf(u,\xi)|^2$ is maximum at $\xi(u)=\phi'(u)$. The windowed Fourier ridges are the maxima of the spectrogram at the TF points $(u,\xi(u))$.

2.3.2. Wavelet ridges

Approximatively analytic wavelets, like Gabor wavelets [7], are used here to separate the phase and amplitude information of signals [7]

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

where η is the frequency center of its Fourier transform $\hat{\psi}$, $g(t)$ is a real and symmetric window used here. The family of TF atoms $\psi_{u,s}$ is obtained by scaling the base atom ψ by *s* and translating it by *u*. The previous function $\psi_{u,s}$ is centered on *u*, like the windowed Fourier atom. If η denotes the frequency center of the base wavelet ψ, then the frequency center of a dilated wavelet is η */s*.

Let $f(t) = a(t)\cos(\phi(t))$. The normalized scalogram $P_W f(u,\xi)$ of the processed signal *f* is given by [7]

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

where $\varepsilon(u,\xi)$ is the corrective term. Since $|\hat{g}(\omega)|$ is maximum at $\omega=0$, (7) shows that if the amplitude and frequency have slow variations over the support of $\psi_{u,s}$, and if the instantaneous frequency is higher than the window's passing band, in order to be able to neglect the corrective term $\varepsilon(u,\xi)$, then the scalogram is maximum at $\frac{\eta}{s(u)}=\xi(u)=\phi'(u)$. The corresponding points $(u,\xi(u))$ are called wavelet ridges [7].

2.3.3. Ridge detection

In each case of analysis, the ridge algorithm detects thus all local maxima of the TF representation. These points define curves in the (u,ξ) plane that are the ridges of the used transformation.

2.4. Computation of formant candidates

The ridge algorithm proceeds to a parabolic interpolation and a threshold to remove ridges corresponding to small amplitude caused by noise variations or shadows of other instantaneous frequencies created by the side-lobes of the window. Then we calculate the frequencies corresponding to the remaining ridge points. These frequencies are the formant candidates that might be chosen from to form the formant tracks. We have to indicate however, that, in the present work, we do not take into account problems of modulation and whether or not the formant tracks are coupled [9][10].

2.5. Determination of the formant track using continuity constraint

Here we suppose that the track concerns only the three first formants. After the threshold elimination of the undesirable ridge points, we might end up with no formants, only one formant (either first, second or third), two, three or more, at a given instant. This might cause a discontinuity in the tracks.

The observation that formants are in general slowly varying functions of time has led to attempts to force continuity constraints on the formant selection process using heuristics [1] [4]. In this case, a continuity criterion is used to choose the closest combination or set of candidates to the previous one. The continuity rule is based on the Euclidean Distance (2-norm distance) between each set and the previous formant set chosen. After the candidates are chosen every possible combination of them will be considered. Then the distance of each candidate set from the previous chosen set will be calculated and the one with the minimum distance will be chosen as the next formant set.

Since we are interested in obtaining the first three formants and F3 is known to be lower than 4 kHz [1], it is advantageous to downsample the signal to 8 kHz to avoid obtaining formant candidates above 4 kHz, and to let us use a lower order analysis which offers less computing time [1][8].

3. Results and discussion

The two proposed algorithms were applied on synthesized vowels. In the case of Fourier analysis, the spectrogram of the signal is calculated using the windowed Fourier transform, while studying the contribution of five analyzing windows: Rectangular, Hamming, Gaussian, Hanning and Blackman window. We use a 10 ms duration window and we consider the pitch equal to 100 Hz. In the same way, we considered different base wavelets in the case of wavelet analysis, like Gabor, Morlet and Mexican Hat wavelet, for comparison purposes. Figures 2 and 3 shows formant tracks for the synthetic vowel /a/ respectively with a rectangular and Hamming window. Figure 4 displays the scalogram and formant tracks for this vowel using Gabor wavelet as analyzing wavelet. The comparison of the five used windows performance in the Fourier analysis case showed that it is particularly important to ensure that the Fourier transform of the analysis window used has negligible side-lobes. Hamming, Hanning, Blackman and Gaussian windows have provided reliable formant tracks. In the wavelet analysis case, among the three analysing wavelets considered, the Gabor wavelet provides the best quality of formant tracks.

In order to evaluate the two proposed techniques are compared to two available formant trackers: that based on LPC spectra (WinSnoori) [4] and that using a time-varying adaptive filterbank [8]. The four formant trackers were run on real speech entities from the TIMIT database. The test results with real speech are demonstrated by means of spectrographic representations, due to unavailability of reference data for formant frequencies of real speech chunks. Figures 5, 6, 7 and 8 show an utterance from a male speaker with the formant tracks done respectively by the formant tracking algorithm using respectively Fourier ridge detection, wavelet ridge detection, LPC spectra and time-varying adaptive filterbank. Though most part of three kinds of trajectories tracked by Fourier ridge detection, LPC spectra and time-varying adaptive filterbank methods are closer to each other, we find high errors in tracking F3 by LPC spectra and time-varying adaptive filterbank methods. However, the wavelet ridge detection method is shown to be not enough capable to provide legible formant trajectories, because of the curve track oscillations.

4. Conclusions and future work

We have presented in this paper two techniques of formant tracking of speech signals. The first technique is based on the Fourier ridges detected from the spectrogram determined by the windowed Fourier transform. The second technique is based on the wavelet ridges detected from the scalogram computed by the wavelet transform. The first method is shown to be a useful tool for formant tracking compared to traditional methods. However, the second method is not enough capable to provide legible formant trajectories. In the future we propose to introduce new parameters in order to eliminate the maximum undesirable ridge points and to take into account the problems of modulation and whether or not the formant tracks are coupled. Also, a correction of the formant trajectories by smoothing will be investigated.

5. References

[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. Speech and Audio Processing*, 2002.

[3] A. Bonneau et Y. Laprie, Elitist identification of stops from formant transitions, In 15th International Congress of Phonetic Sciences - ICPhS'2003, Barcelona, Spain. 2003.

[4] S. McCandless. An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra. *IEEE ASSP-22*, volume 2, pages 135-141, April 1974.

[5] K. Xia et C. Espy-Wilson, A new strategy of formant tracking based on dynamic programming", vol. 3, 55-58, ICSLP 2000, Pékin, Chine, 2000.

[6] Y. Laprie, "A concurrent curve strategy for formant tracking", In *Proceedings of the International Conf. on Spoken Language Processing*, Jegu, Corée du sud, 2004.

[7] S. Mallat, "A wavelet tour of signal processing", second edition, Academic press, 1998.

[8] 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.

[9] A. Potamianos et P. Maragos, Speech Formant Frequency and Bandwidth Tracking Using Multiband Energy Demodulation, Journal of Acoustical Society of America, vol.99 (6), pp. 3795--3806, 1996.

[10] L. Deng, A. Acero, et I. Bazzi., Tracking vocal tract resonances using a quantized nonlinear function embedded in a temporal constraint, IEEE Transactions on Speech and Audio Processing, 2006.

[11] L. Welling et H. Ney, Formant estimation for speech recognition, IEEE Transactions on Speech and Audio Processing, Volume 6, Issue 1, pp.36 - 48, 1998.

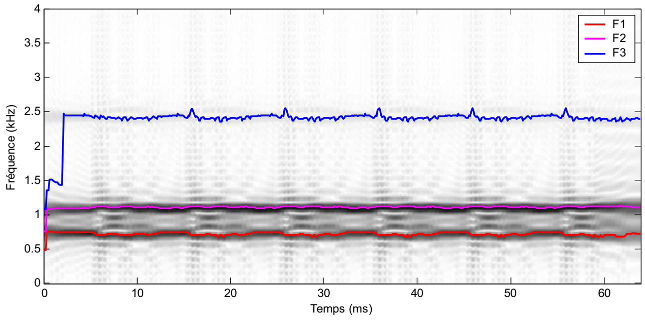

Figure 2 *Formant tracks by Fourier ridge detection for the synthetic vowel /a/superimposed on its wideband spectrogram computed using a rectangular window.*

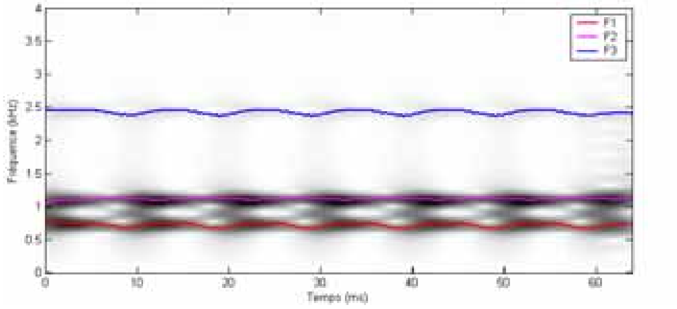

Figure 3 *Formant tracks by Fourier ridge detection for the synthetic vowel /a/superimposed on its wideband spectrogram computed using a Hamming window.*

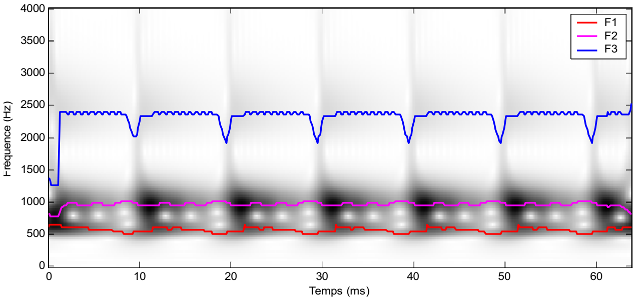

Figure 4 *Formant tracks by wavelet ridge detection for the synthetic vowel /a/superimposed on its scalogram computed using a Gabor wavelet.*

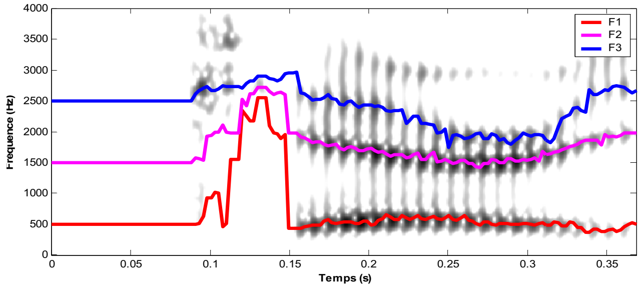

Figure 5 *Formant tracks by Fourier ridge detection for the word "carry" superimposed on its wideband spectrogram computed using a Hamming window.*

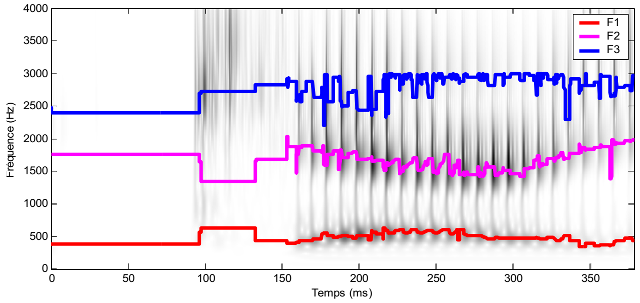

Figure 6 *Formant tracks by wavelet ridge detection for the word "carry" superimposed on its scalogram computed using a Gabor wavelet.*

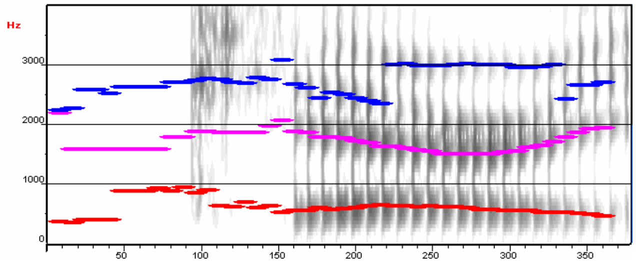

Figure 7 Formant tracks by LPC spectra method for the word "carry" superimposed on its spectrogram.

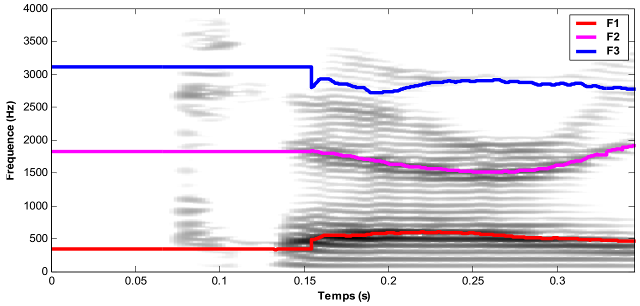

Figure 8 *Formant tracks by the time-varying adaptive filterbank method for the word "carry" superimposed on its spectrogram.*