Formantic Analysis of Speech Signal by Wavelet Transform

L.Falek, A.Amrouche, L.Fergani , H.Teffahi, A.Djeradi

***Abstract*— The goal of this study is to consider the instantaneous frequencies corresponding to the speech signal formant using the wavelet transform. The developed method is based on an analysis of derivative phase of the continuous Morlet wavelet transform coefficients. Using synthesized signals produced by a formant model made it possible to adapt this method to real speech signals and to determinate the instantaneous formantic frequencies for all the types of speech sounds (isolated vowels, syllables and words). Results are satisfactory and join those obtained by other researchers in the field (who were interested especially in isolated vowels).**

***Index Terms*— speech signal, instantaneous frequencies, formants, complex wavelet, phase derivative.**

I. Introduction

Formants are the maximum energy of instantaneous frequencies of speech signal. They are basic components of coding, recognition or synthesis systems of signals. They can also be useful for specialized applications as assistance to medical diagnostic (for larynx pathologies : by speech signal analysis). Considering their growing importance, the formants are the subject of many works. The difficulty in formants calculation is related mainly to none stationnarity of the speech signal. Time-frequency representations (like continuous wavelet transform) made formidable great strides these 30 last years with the very fast evolution of computers calculation capacities. These representations are adapted to signals presenting frequential contents which vary during time (what is the case of the speech signal).

Manuscript received March 23, 2011; revised April 8, 2011. This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here).

L. Falek is with the University of science and technology Houari Boumediene (USTHB), Electronic and Computer Science Faculty, BP 32, El Alia, Algiers, ALGERIA. (e-mail: lfalek@hotmail.fr).

Aissa Amrouche is with the University of science and technology Houari Boumediene (USTHB), Electronic and Computer Science Faculty, BP 32, El Alia, Algiers, ALGERIA. (e-mail: amrouche_a@yahoo.com).

L.Fergani is with the University of science and technology Houari Boumediene (USTHB), Electronic and Computer Science Faculty, BP 32, El Alia, Algiers, ALGERIA. (e-mail: lamifer@msn.com).

H.Teffahi is with the University of science and technology Houari Boumediene (USTHB), Electronic and Computer Science Faculty, BP 32, El Alia, Algiers, ALGERIA. (e-mail: hteffahi@gmail.com).

A.Djeradi is with the University of science and technology Houari Boumediene (USTHB), Electronic and Computer Science Faculty, BP 32, El Alia, Algiers, ALGERIA. (e-mail: adjeradi05@yahoo.com).

They provide a joint representation in time and frequency, contrary to the Fourier transform who represents in only frequential form the contained information in signal temporal , from where the disadvantage of loss of events chronology . Wavelets transform takes up same idea as the Fourier transform by adopting a multi-resolution approach: if we look at a signal with a broad window, we will be able to distinguish from the coarse details. In a similar way, increasingly small details could be observed by shortening the window size. Wavelets analysis objective is thus to carry out a kind of adjustable mathematical microscope. In this study, we developed a determination method of instantaneous variation of a speech signal formantic frequencies based on complex continuous wavelet transform. The method principle is the phase exploitation of coefficients transformation for the instantaneous frequency extraction by using an analytical complex continuous wavelet transform. The application of this method to the speech signal is carried out by taking account of the acoustic characteristics of this signal.

We proceeded in first to the adjustment of method parameters starting from three vowels (/a/, /i/, /u/) obtained using a formants synthesizer, then we widened the application to real speech signals (isolated vowels, syllables and words). The results represented on a spectrogram were compared with those obtained using a traditional method (LPC). They were considered to be satisfactory.

II. ANALYTICAL SIGNAL AND INSTANTANEOUS FREQUENCY

The analytical signal concept was posed by Ville in 1948 [13] with an aim of defining the instantaneous frequency. The analytical signal is a complex signal associated with a real signal. It has interesting properties, in particular with regard to its Fourier transform, which is null for the negative frequencies.
An analytical signal z(t) can be calculated starting from the Hilbert transform [11] of a real signal $x(t)$ such as:

$$H(x(t)) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(s)}{t-s} ds \qquad (1)$$

$zx(t) = x(t) + iH(x(t))$ with H(x(t)) the Hilbert transform of x(t) : It is possible to define [6], starting from the analytical signal an instantaneous frequency $fx(t)$ [4], where $zx(t)$ is an analytical signal

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$f_x(t) = \frac{1}{2\pi}\frac{d\ arg(z_x)}{dt}(t) \quad (2)$$

III. MORLET CONTINUOUS WAVELET TRANSFORM

A. The continuous Morlet wavelet

The complex continuous Morlet wavelet transform also makes it possible to define the concept of the instantaneous frequency, when one employs an analytical wavelet [7]. This is why we choose the complex continuous Morlet wavelet like mother wavelet. The latter is an analyzing wavelet for small oscillations (a centre frequency FC: around 1Hz). Moreover, it is very quite localized in time (between -4 and 4s) and in frequencies (a peak around 0.8 Hz). What makes of it very a good candidate for the analysis of the speech signal. The Morlet wavelet is inspired by the elementary Gabor signal. It is obtained by modulation of a Gaussian. It is given by the following relation [3]

$$\psi(t)=\frac{1}{C}\exp j\omega_c t\left[\exp\left(-\frac{t^2}{2\sigma_t^2}\right)-\sqrt{2}\exp-\frac{\omega_c^2\sigma_t^2}{4}\exp\left(-\frac{t^2}{\sigma_t^2}\right)\right] \quad (3)$$

Where: C: factor which makes it possible to standardize energy.

$\omega_c=2\pi F_c$ (F_c : is the wavelet centre frequency)

$\sigma_t=(1/(2\pi\sigma_f)$ (σ_t σt is the standard deviation of Gaussian and $4\sigma_t$ is the effective wavelet duration and $4\sigma_f$ its band-width).

The product $\omega_c\sigma_t$ fixes the bond between the width of the wavelet Gaussian envelope and its oscillation frequency F_C [8]. To have a wavelet family, the product must be constant. For the Morlet wavelet, this last must take rather large values ($\omega_c\sigma_t \geq 5$ in practice).

For low oscillations: $\omega_c = 2\pi Fc=5.486$ rad/s were *Fc*≈0.8Hz. One usually uses for ωc values: $5\leq \omega_c \leq 6$. Then for Fc : $(0.8\leq Fc \leq 1)$ Hz [3]

B. The wavelet transform

Compared to the Fourier transform, the basic idea of the wavelet transform is to break up a signal x(t) according to another base that of the sinusoids, each wavelet basis having particular properties which guide its use for the type of problem arising. Signal x(t) thus will be broken up on a functions family relocated and dilated starting of a single function ψ (t) called wavelet mother. The family puts itself in the form [3 9 12]

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (4)$$

and is called wavelet analyzing, with "a" parameter of dilation or scale parameter, defining the width of the analysis window .

The variable “a” cheek the role of the opposite frequency: more "a" is weak less, the wavelet analyzing is wide temporally, and therefore more the centre frequency of its spectrum is high. The parameter “b” is the translation parameter locating the wavelet analyzing in the temporal field. The Modification of "a" and "b" allow to have wavelet at the desired frequency and the desired moment. By noting ψ * (t) complex combined of ψ (t), the wavelet Transform of signal x (t) is defined by [3 9 12]

$$(W_\psi x)(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\ \psi^*\left(\frac{t-b}{a}\right) dt \quad (5)$$

This analysis makes it possible to locally describe the contents of x(t) in the vicinity of (a, b) in the time-scale plan. It indicates to us relative importance of the frequency 1/a around the point b (or at the moment b) for a signal x(t). Thus, if x(t) vibrates at a frequency definitely less raised or, on the contrary, much higher than 1/a, the module of the wavelet transform coefficient will be very small and almost negligible. It becomes consequent only if the signal contains a component of this frequency at the point considered. The coefficients of the wavelet transform are thus a way of locating with precision the appearance of a frequency given to one moment given in a signal. This decomposition is function of two variables "a" and "b" and evaluates the relevance of the use of the wavelet in the description of x(t)

IV. METHODOLOGY

The implementation of the method was made in Matlab language, by taking account of the mother Morlet wavelet parameters quoted in part III.A. We started by validating the method for a theoretical signal then we passed to the application on a speech signal

A. Application on a theoretical signal

We chose a signal x(t) composed of 3 sinusoids such as:

$x(t)= 5\sin(2\pi*300t) + 10\sin(2\pi*2000t) + 15\sin(2\pi*3500t)$

The scales were selected so that they cover the required frequencies: 2 wavebands: [200, 1500]; [1400, 5000].

Product values $\omega c\sigma_t \geq 5$ like known in the literature.

We calculated the module of the wavelet transform coefficients (amplitude of the module according to the scales).

The result obtained is illustrated in fig.1. The instantaneous frequencies are estimated by the maximum of the module of the coefficients of the wavelet transform as specified of Fig.1. We locate then the scales which correspond to maximum of the module of the coefficients.

We calculate the frequencies by deriving coefficients phase of the wavelet transform which corresponds to these scales compared to time: (figure2). These last correspond well to the frequencies of signal x(t) (relative to each peaks of the module: max1, max3, max3) : F1=500Hz, F2=2000Hz et F3=3500Hz. Those can thus be used for a first estimate of instantaneous frequencies signal. Fig.3 represents the energy of one of the frequencies of Fig.2. In fig.4, we have maximum energies of the frequencies obtained of Fig.3 (represented on the spectrogram of signal x(t)), correspondents well at the frequencies of signal x(t).

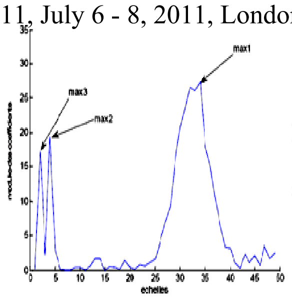


Figure 1: modulate coefficients **of** the wavelet transform according to the scales

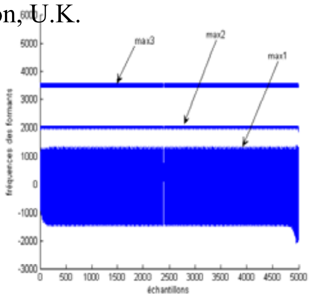


Figure 2. Phase derivative of wavelet transform corresponding for each maximum

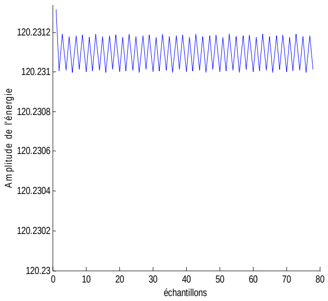


Figure 3. Coefficients energy of the wavelet transform, for frequencies one frequency

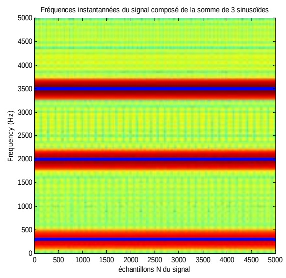


Figure 4: Maximum energy for the 3 instantaneous

These results show that the method is applicable on a theoretical signal.

B. Application to a speech signal

The speech signal is a non stationary signal, being able to contain a big number of very variable instantaneous frequencies. What makes its analysis delicate by this method. For that, we started by analyzing a signal of word of which we know all frequencies preliminary in order to try to fix the parameters of the method.

1) Speech Signals Analysis obtained using formants synthesizer (Klatt)

We synthesized 3 vowels (/a/, /i/ and /u/) whose Fi formants are constant (see table 1). On the basis of the assumption that the product $\omega_c\sigma_t$ fixes the bond between the width of the Gaussian envelope of the wavelet and its oscillation frequency FC and that to have a family of wavelets, the product must be constant (for the Morlet wavelet: $0.8\leq$ FC ≤ 1et $\omega_c\sigma t\geq 5$), it then acts to find the couples of parameters (FC, $\omega c\sigma_t$) which make it possible to detect formants values introduced into the synthesizer with precision. These couples of values will be then used and adjusted for real speech signals. The method used is the following one:

- Choice of 5 wavebands allowing covering all the formantic frequencies (5 formants) for the three sounds. These last will be used to fix the scales for the wavelet transform: [200,900], [900,1500], [1500,3000], [3000,4500], [4500, fe/2] in Hz;
- We then varied centre frequency F_C of the wavelet mother by step of 0.5Hz with $0.8\leq F_C\leq 1$
- For each value of FC we varied the product $\omega_c\sigma_t$ by step of 0.5 while starting of 5 like city in the literature.

Couples of parameters (F_C, $\omega_c\sigma_t$) obtained for the 3 sounds are illustrated by table 2. The corresponding results are illustrated in fig 5. These results join those obtained by L.Cnockaert [8] while reasoning on the centre frequency (F0, F1, F2, F3) of the wavelet signal.

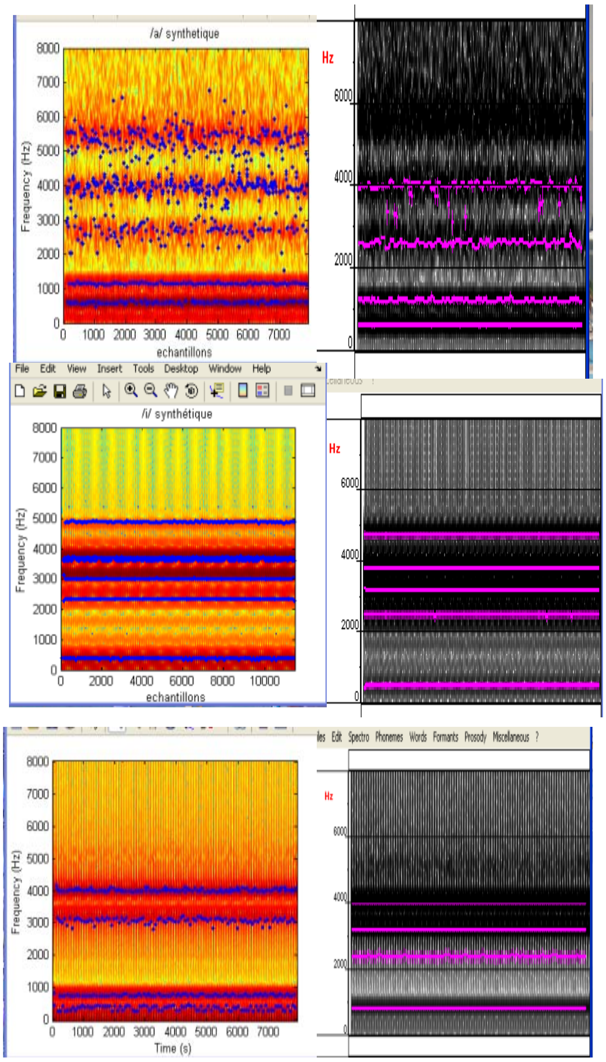
Table 1: Values fixed for the formants for the synthesis of the 3 vowels

	F0	F1	F2	F3	F4	F5
/a/	200	700	1200	2700	3900	5700
/i/	200	270	2400	3370	3800	4900
/u/	300	350	900	3300	4000	5200

Table 2. Couples of parameters (F_C, $\omega_c\sigma_t$) valid for the 3 synthetics vowels

Frequencies bands	Fc	$\omega_c\sigma_t$
[200,900]	0.8	**4.5**
[900,1500]	1	**8**
[1500,3000]	0.9	**8**
[3000,4500]	1	**10**
[[4500,fe/2]	1	**10**

The instantaneous formantic frequencies found on fig.5 correspond well to the frequencies fixed in table 1. These results will be applied for the calculation of the formants in the case of a real speech signal.

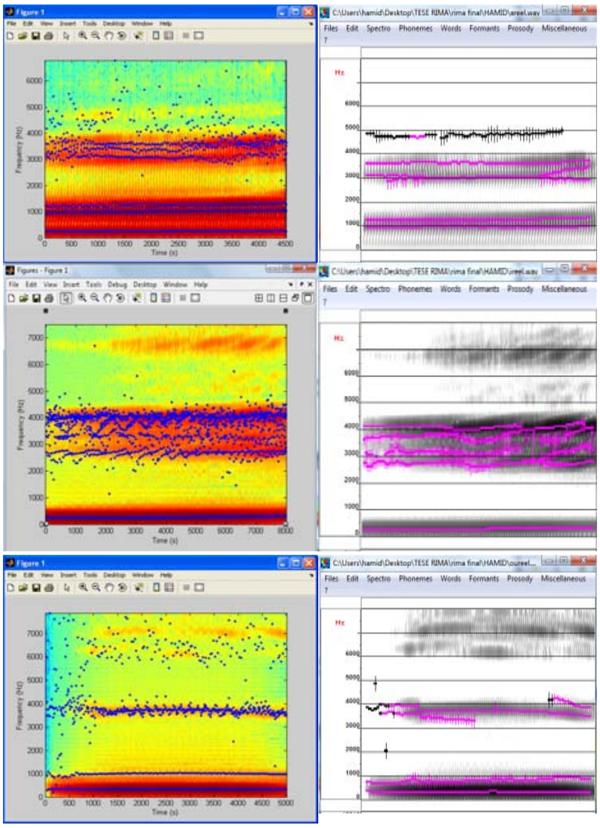


Figures 5 : Results obtained for the vowels [a], [I], [u] (on the left with wavelet transform and on the right, with the software Winsnoori using the LPC)

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

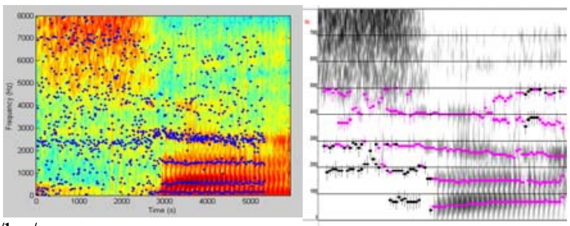
2) Application method for real speech signals

On the basis of couple of parameters obtained for synthetic speech signals, we applied the method to real speech signals (vowels, syllables and words) and we represented the results on the spectrogram relative to each sound (Fig 6, 7, 8)

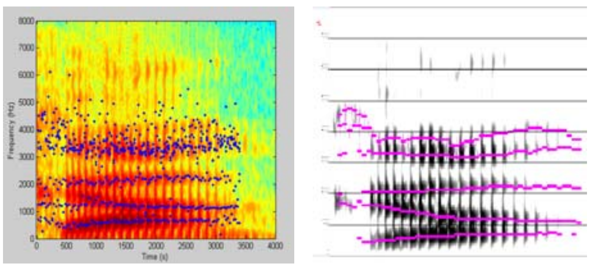


Figures 6. Case of the real isolated vowels: /a/, /i/, /u/. On the left, with wavelet transform and on the right, with the LPC (software Winsnoori)

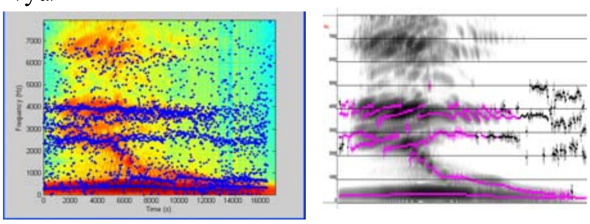
/sa/



/ka/

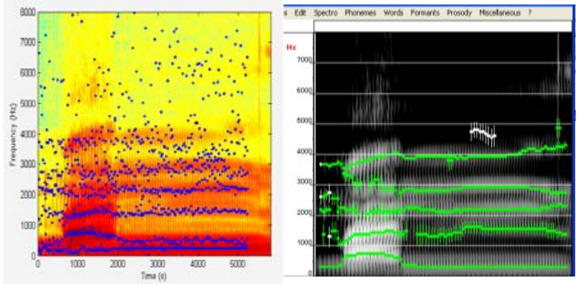


/yu/

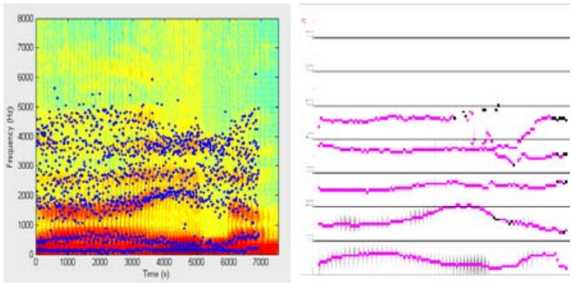


Figures 7. Case of real syllables On the left, with wavelet transform and on the right, with the LPC (software Winsnoori)

/men/



/aʕila/



Figures 8. Case of Arabic words. On the left, with wavelet transform and on the right, with the LPC (software Winsnoori)

V. Conclusion

This study showed that the continuous complex wavelet transform of a speech signal can make it possible to estimate the frequencies of the formants, when the parameters of the transform are quite selected. Indeed, the values of the product $\omega_c\sigma_t$ obtained for the analyzed sounds (isolated vowels, consonants, syllables and words) made it possible to obtain satisfactory results and join those obtained by another researcher for isolated vowels. The difficulty of this method lies in the choice of these parameters like all the other methods of formants detection; however, it has the advantage be applied directly to non-stationary signals. In addition, we noticed the appearance of additional points to those of the existing formants on the various figures of the results presented for the case of real signals. We think that this problem can be related to the presence of other frequencies in the speech signal (harmonic noises and frequencies, aspiration,…) which energy is important and detected by our method.

References

[1] Bilezikian JP, Raisz LG, R. G. (2002). Principles of Bone Biology- Second Edition. Aca-demic Press. 95, 96

[2] Castellengo, M. and Dubois, M. (2005). Timbre ou timbres? propriété du signal, de l'instrument ou construction cognitive. 88

[3] C. Chui, An introduction to wavelets, Academic Press,1995.

[4] Cohen, L. and Lee, C. (1989). Instantaneous frequency and time-frequency distributions. IEEE International Symposium on Circuits and Systems, 1989., pages 1231–1234. 11.

[5] Emiya, V. (2004). Spectrogramme d'amplitude et de fréquences instantanées (safi). Master's thesis, Aix-Marseille II. 52

[6] Flandrin, P. (1993). Temps-fréquence. Hermès. 11, 27, 29, 3.

[7] J. Lardiès, M.N. Ta M.N. , M. Berthillier, Modal parameter estimation from output-only data using the avelet transform, *Archive of Applied Mechanics,* Vol. 73, pp. 718-733; 2004

[8] Laurence Cnockaert. « Analyse du tremblement vocal et application à des locuteurs parkinsoniens ». Thèse de doctorat en sciences de l'ingénieur. Bruxelles, Décembre 2007. Université libre de Bruxelles ULB.

[9] S. Mallat, Une exploration des signaux en ondelettes, Editionsdel'école Polytechnique, 2000.

[10] Navarro, L. (2007a). Analyse temps-fréquence de signaux vibratoires issus d'un réacteur de culture osseuse. In Journée de la recherche de l'EDSE. 143.

[11] Rihaczek, A. and Bedrosian, E. (1966). Hilbert transforms and the complex representation of real signals. *Proceedings of the IEEE*, 54(3) :434–435. 11.

[12] B. Torrésani, Analyse continue par ondelettes, Editions du CNRS, Paris, 1995

[13] Ville, J. (1948). Théorie et applications de la notion de signal analytique. *Câbles et transmissions*, 1 :61–74. 11, 13.

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)