A new Automatic Formant Tracking approach based on scalogram maxima detection using complex wavelets

Imen Jemaa, Kais Ouni, Yves Laprie, Slim Ouni, Jean-Paul Haton

A new Automatic Formant Tracking approach based on scalogram maxima detection using complex wavelets

Imen Jemaa[1, 2], Kaïs Ouni [1], Yves Laprie[2], Slim Ouni[2] and Jean Paul Haton [2]

[1] *Laboratoire de Signal, Image et Technologies de l'Information*

Ecole Nationale d'Ingénieurs de Tunis, BP.37, Le Belvédère 1002, Tunis, Tunisie

[1] imen_jemaa@yahoo.fr

[1]kais.ouni@enit.rnu.tn

[2]*Equipe Parole, LORIA – BP 239 – 54506 Vandœuvre -lès- nancy, France*

[2]Yves.Laprie@loria.fr

[2]slim.ouni@loria.fr

[2] Jean-Paul.Haton@loria.fr

***Abstract*— In this paper we present a new formant tracking algorithm where the formant frequencies estimation was based on local maxima detection of a time frequency representation. This representation can be shown by a scalogram issued from a complex wavelet transform. The formant frequency candidates are validated as local maxima of scalogram which correspond to wavelet ridges. Then in the proposed algorithm, we have introduced the computation of center of gravity as tracking constraint. We tested our new algorithm by applying it on synthesized and natural voiced speech signals. The formant trajectories obtained by our algorithm were compared to those of manually-edited ones of our Arabic database as reference; those given by Fourier transform method and the LPC analysis used in Praat. The comparison of the results showed globally the adequacy of the first three formant trajectories using complex Morlet wavelet refers to the manually-edited formant tracks.**

***Keywords*— Speech processing, Formant tracking, Scalogram, wavelet ridges, center of gravity, Arabic database.**

I. INTRODUCTION

The problem of formant tracking has received considerable attention in speech recognition research, as formant frequencies are known to be important in determining the phonetic content as well as the articulator information about the speech signal [1]. Indeed, robust formant tracks are utilized to identify vowels [2] and other vocalic sounds [3].

Although automatic formant tracking has a wide range of applications, it is still an open problem in speech analysis. Especially, when anti-formants are present, as in most consonantal sounds, the underlying resonant frequencies i.e. formants are often hidden.

In this paper we describe our new automatic formant tracking algorithm based on the detection of wavelet ridges which are the maximum of the scalogram. This algorithm uses a tracking constraint by calculating the centre of gravity for a set of frequency formant candidates. Then, to report a quantitative evaluation of the proposed algorithm to other automatic formant tracking methods we used manually-edited formant trajectories of our Arabic database as reference prepared specially for this work.

This paper is presented as follows, we present in section 2, the proposed formant tracking algorithm, in section 3, the results obtained by comparing the proposed algorithm with other automatic formant tracking methods. Finally, we give some perspectives in section 4.

II. FORMANT TRACKING ALGORITHM

The block diagram presented in (Fig.1) describes the main steps of the proposed algorithm. Each element of the block diagram is briefly described below.

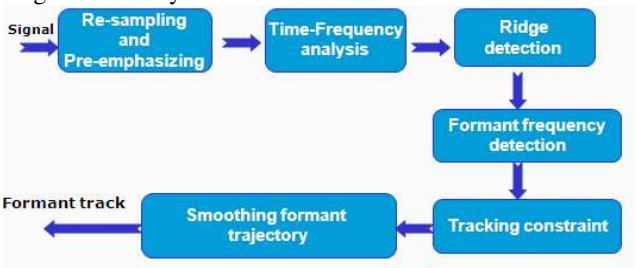

Fig. 1 Block diagram of the proposed formant tracking algorithm

A. *Re-sampling and Pre-emphasizing*

The sampling frequency used for speech signal in the database is16 kHz. Since we are interested in the first three formants, we down-sampled the speech signal at 8 kHz for computational reasons. Then to accentuate the high frequencies, a first order pre-emphasis filter is applied on the processed speech signal.

B. *Time-Frequency Analysis*

The goal of time-frequency analysis is to find what frequency occurs at what time in a signal. It enables us to analyze the non-stationary signals. The time-frequency analysis used here is the complex wavelet transform and can

be shown using a scalogram. This wavelet transforms uses a complex wavelet which separates both amplitude and phase information [4]**.** To calculate the scalogram representation, we tested three types of complex wavelets which are: Complex Morlet (CMOR), Shanon (SHAN) and Frequency B-Spline (FBSP) which are included in Matlab wavelet toolbox [5].

C. Ridges Detection

The tracking application described here uses the instantaneous frequency to find formants in the speech signal. Indeed, we can show that the instantaneous frequencies are almost as validated spectral maxima in the time-frequency (TF) plane. The accumulation of near instantaneous frequencies can therefore find the formants. We applied this property to detect the instantaneous frequencies (wavelet ridges) in the scalogram and validated as formants using the complex wavelet transform. A complex wavelet is used here to accurately measure the instantaneous frequency varies more rapidly at high frequencies. Thus in this section, we will define the instantaneous frequency of a signal, how to validate them as local maxima (ridges) of a scalogram. For this we consider a complex wavelet defined by [4] [6]:

$$\psi(t) = g(t)\exp(i\eta t) \tag{1}$$

Where η is the frequency center of wavelet ψ, *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 centred 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

The algorithm computes the instantaneous frequencies of the input signal *f* which are considered as local maxima of its scalogram. The voiced speech segments of *f* can be modelled with sums of sinusoidal partials. The instantaneous frequency can be measured from the ridges of the wavelet transform using a normalized scalogram. The normalized scalogram of the processed signal *f* is given in [4] as below:

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

The corrective term $\varepsilon(u,\xi)$ is negligible if a(t), analytic amplitude of *f*, and $\phi'(t)$, the instantaneous frequency which is the derivative of the phase of *f*, vary slowly on the support of $\psi_{u,s}$ and if $\phi'(u) \geq \frac{\Delta\omega}{s}$. Since $\left|\hat{g}(\omega)\right|$ is maximum at $\omega = 0$, the Equation (Eq.2) shows that if the corrective term is neglected, then the scalogram is maximum at $\frac{\eta}{s(u)} = \xi(u) = \phi'(u)$. The corresponding TF points $(u, \xi(u))$ are called wavelet ridges. Therefore, the algorithm detects all local maxima of scaloagram in points $(u, \xi(u))$. Thus, we obtained for each formant the combination frequency candidates.

D. Formant Frequency Detection

In this work, we suppose that the tracking concerns only the three first formant tracks, so in each case of analysis, we have split the set of formant frequency candidates that were detected previously into three wide bands respectively for each formant track. The algorithm proceeds, then, to removing ridges corresponding to small amplitude below a given threshold, because they may be, for instance, instantaneous frequency specific to the fundamental frequency F0 [7] . Then, we calculate the frequencies corresponding to the remaining ridge points.

E. Tracking Constraint

It is considered that in general, formants vary slowly in time, which leads to impose a continuity constraint in the process of selecting formant frequencies from the set of candidates.

In this algorithm we propose the calculation of the centre of gravity for a set of formant frequency candidates, detected by the local maxima detection stage, as a constraint of tracking between signal frames. Since the detection of local maxima gives several candidates close together for one formant, the idea is to calculate the centre of gravity of the set of instantaneous frequency candidates located in the frequency band weighted by the spectral energy. The resulting frequency *f* is given by:

$$\bar{f} = \frac{\sum_{i=1}^{n} p_i f_i}{\sum_{i=1}^{n} p_i} \tag{3}$$

Where f_i is the frequency of the i^{th} candidate and p_i its spectral energy and *n* is the number of instantaneous frequencies considered. Considering centre of gravity instead of isolated candidates allows smooth formant transitions to be recovered.

F. Smoothing Formant Trajectory

To smooth the formant trajectories, first we interpolate the sequence of points resulting from the previous step of calculating the tracking constraint. Then, we calculate the moving average value of each point of the trajectory with the frequencies already previously chosen that is to say, bearing in each time the track history. Following this step, the trajectories corresponding to the first three formants are continuous and very smooth.

III. RESULTS AND DISCUSSION

We tested our proposed formant tracking approach on synthetic vowels (/a/, /i/et /u/) before moving on to real speech signals. To calculate the scalogram, we tested three kinds of complex wavelets: Morlet (CMOR), Frequency B-Spline

(FBSP) and Shanon (SHAN). First, we tested the three wavelets to set the appropriate parameters for each wavelet gives good scalogram resolution and ridge detection. The common parameters for the three wavelets are f_b (bandwidth) and f_c (center frequency of the wavelet) and with the extra parameter of the wavelet FBSP noted m which is the order derivative parameter where $(m \geq 1)$. Finally, we set a choice: CMOR ($f_b = 10$ and $f_c = 1$), FBSP ($m = 10$, $f_b = 1$ and $f_c = 1$) and SHAN ($f_b = 0.1$ and $f_c = 1$). The quantitative evaluation of our method consists in calculating the averaging absolute difference in (Hz) refer to (Eq.4) and the standard deviation normalized with respect to reference values in (%) refer to (Eq.5) for every formant track (F1, F2 and F3). We use our prepared Arabic database as reference [8].

$$Diff = \frac{1}{N} \times \sum_{p=1}^{N} \left| F_r(p) - F_c(p) \right| \quad \text{Hz} \tag{4}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{p=1}^{N} \left(\frac{\left| F_r(p) - F_c(p) \right|}{F_r} \right)^2} \tag{5}$$

Where F_r is the reference formant frequency, F_c the calculated formant frequency corresponding to each formant tracking methods, N the total number of formant frequencies for each formant trajectory and p is the counter of formant frequencies for each formant trajectory: p=1 to N values.

TABLE I
QUANTATIVE RESULTS OBTAINED ON SYNTHETIC VOWELS /a/, /i/ et /u/

		Crêtes d'Ondelette: CMOR			Crêtes d'Ondelette: FBSP			Crêtes d'Ondelette: SHAN		
Suivi de Formants		F1	F2	F3	F1	F2	F3	F1	F2	F3
/a/	Diff	27	26	30	42	14	29	36	37	54
	σ	4	4	4	6	3	4	5	5	8
/u/	Diff	0	18	270	0	8	270	8	47	270
	σ	0	6	90	0	3	90	3	16	90
/i/	Diff	39	15	66	407	472	17	190	681	36
	σ	15	6	27	152	181	9	76	253	15

The color values displayed in the table above are the high difference values (more than or equal to 90Hz) between the values estimated by each method of automatic formant tracking and reference value. The red values correspond to the calculation of the mean absolute difference (Diff in Hz) and the blue ones correspond to the calculation of the corresponding standard deviations (in %).

When comparing the results shown in the table1 for the proposed algorithm by testing the three complex wavelets we notice that the results using the wavelet CMOR are largely better than the other wavelets especially for the synthetic vowel /i/. However, we notice globally the performance of the proposed automatic formant tracking algorithm based on wavelets detection concerning three synthetic vowels.

Then, we examined results obtained for the short vowel /a/ within the syllable CV preceded by one consonant from every phonetic class for each formant track (F1, F2 and F3). The CV occurrences were taken from the four following sentences:" عَرَفَ وَالِيًا وَقَائِدً" ("**ɛarafawa:liyanwaqa:3idan**" which means "He knew a governor and a commander"),"لَقَدْ كَانَ مُسَالِمًا وَقُتِلَ." ("**laqadka:namusa:limanwaqutila**" which means "He was a pacifist and was killed"), "هِيَ هُنَا لَقَدْ آبَتْ"("**hiyahuna: laqad3a:bat**" which means "She is here and she was pious") and "قَادَ الجَيْشَ."("**qa:daljayŝa**" which means "He commanded the army").

To quantitatively evaluate the proposed algorithm, we compare it with our automatic method based on Fourier ridges detection [13] and the LPC method integrated in Praat [14] using our prepared database as a reference [8].

TABLE II
RESULTS OBTAINED ON THE VOWEL/a/ PRONOUNCED BY A MALE SPEAKER M1.

Loc.M1			LPC			Fourier			Complex wavelet CMOR			Complex wavelet FBSP			Complex wavelet SHAN		
			F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Plosive voiced:	/d/	Diff	9	30	10	31	64	50	37	132	25	88	43	52	7	90	83
		σ	3	6	2	7	14	11	8	27	6	17	9	13	2	18	20
Plosive voiceless:	/q/	Diff	50	59	73	44	89	13	46	73	20	16	44	91	20	61	82
		σ	10	10	15	10	17	2	8	16	4	4	8	20	4	14	15
Fricative voiced:	/H/	Diff	44	46	28	24	49	33	67	42	34	162	35	37	78	57	106
		σ	7	7	5	4	8	6	9	7	5	24	6	6	13	8	15
Fricative voiceless:	/f/	Diff	28	27	56	13	22	36	17	40	41	10	15	28	42	27	158
		σ	6	5	12	3	4	8	4	8	8	2	3	6	8	5	29
Nasal:	/m/	Diff	100	17	94	71	23	41	53	39	116	72	101	280	70	8	77
		σ	20	4	22	14	5	8	10	8	23	13	20	50	13	2	14
Lateral:	/l/	Diff	49	25	42	17	54	37	74	49	32	50	39	79	26	118	94
		σ	10	5	10	3	9	7	12	8	5	8	7	13	5	19	15
Tap:	/r/	Diff	64	50	149	19	38	128	48	83	162	55	41	156	24	59	191
		σ	13	9	30	4	9	24	9	16	32	10	9	31	5	11	35
Semi-vowel:	/w/	Diff	72	41	88	81	32	47	49	25	43	124	39	47	45	25	117
		σ	25	9	29	14	6	9	10	5	9	23	7	9	8	5	20

When comparing the results shown in the table above for the proposed algorithm by testing the three complex wavelets we observe as well as the results using the wavelet CMOR are

slightly better than the other wavelets so there are an accurate formant tracking since values error of the mean absolute difference and standard deviation (less than 10 Hz) are very small compared to the values of other wavelets and thus its formant tracking is near or close to the manual formant trajectories reference.

The results show that the formant tracking is generally good for all three formants of the short vowel / a / whatever the nature of the consonant which preceded it. This is true for all three methods Praat, Fourier and wavelet. By cons, the error values (the red values) are localized in F3 in the case where the preceding consonant is / r /.

Moreover we can see also that the proposed algorithm using CMOR wavelet transform and Fourier method are also very close to the reference against LPC method via Praat.

Then, we compared our new method based on wavelet (CMOR) detection with the same automatic formant tracking Fourier and LPC methods with other male speakers M2, M3, M4 and M5 for the same vowel / a /. The syllables CV are always taken from the same four sentences cited above. In this paper we presented only the Table III with two male speakers M2 and M3.

From the results studied on five speakers (M1, M2, M3, M4 and M5), we notice that the number of errors (red values) is generally less for three formant tracking by the proposed method using CMOR than other methods Fourier and LPC. This method therefore provides an accurate formant tracking (F1, F2 and F3) very close to the reference. The results of Fourier and wavelet methods are very similar in some cases since both have fewer errors than the LPC method.

The results show that the formant tracking is generally good for all three formants of the short vowel / a / whatever the nature of the consonant which preceded it. This is true for all three methods Praat, Fourier and wavelet. By cons, the error values (the red values) are localized in F3 in the case where the preceding consonant is / r /. It can be explained by the fact that the vowel / a / is preceded by a consonant tap / r / (very short) and followed by the fricative / f / and this is due to the effect of coarticulation. These results are verified for all speakers, since in this case the same sentence is pronounced.

TABLE III
RESULTS OBTAINED ON THE SHORT VOWEL /a/ IN SYLLABLE CV WITH DIFFERENT TYPES OF CONSONANT FOR TWO DIFFERENT MALE SPEAKERS M2 AND M3

		Loc.M2								Loc.M3							
		Praat			Crêtes de Fourier			Crêtes d'Ondelette: CMOR		Praat			Crêtes de Fourier			Crêtes d'Ondelette: CMOR	
Suivi de Formants		F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Occlusive voisée : /d/	Diff	14	21	24	36	37	18	17	100	28	9	39	35	21	53	65	53	186	62
	σ	3	4	5	8	8	4	4	19	6	2	8	7	4	12	14	9	35	11
Occlusive non voisée: /q/	Diff	96	89	60	32	103	57	26	25	22	19	45	95	45	49	91	17	36	38
	σ	22	22	12	7	21	10	5	5	4	3	9	16	9	8	15	3	7	7
Fricative voisée: /ɛ/	Diff	56	157	77	37	71	60	82	82	20	37	64	87	40	125	82	33	46	60
	σ	10	37	16	5	11	13	13	11	4	6	10	16	6	18	18	5	7	17
Fricative non voisée: /f/	Diff	26	34	114	9	12	72	38	17	44	25	68	138	54	58	35	23	81	26
	σ	6	7	25	2	3	17	8	3	9	7	14	33	11	12	7	5	15	6
Masale: /m/	Diff	56	44	129	57	79	43	67	29	37	38	34	256	30	134	93	21	24	71
	σ	16	13	29	13	18	10	13	7	8	11	9	67	7	25	19	4	5	13
Latérale: /l/	Diff	61	73	52	38	74	35	68	110	44	44	37	84	44	116	64	21	70	12
	σ	10	18	9	6	12	6	10	17	7	8	8	21	10	19	17	4	11	2
Tap: /r/	Diff	29	35	151	11	46	101	45	94	101	73	149	259	37	73	341	41	75	340
	σ	5	7	28	2	11	19	8	16	21	17	81	64	7	13	68	8	14	62
Semi-voyelle: /w/	Diff	53	49	98	78	37	99	54	31	39	53	36	87	47	33	85	67	39	31
	σ	10	9	24	15	7	22	10	6	9	12	6	16	12	7	23	11	8	5

TABLE IV
RESULTS OBTAINED ON THE SHORT VOWEL /a/ IN SYLLABLE CV WITH DIFFERENT TYPES OF CONSONANT FOR TWO DIFFERENT FEMALE SPEAKERS W3 AND W4

		Loc.W3								Loc.W4							
		Praat			Crêtes de Fourier			Crêtes d'Ondelette: CMOR		Praat			Crêtes de Fourier			Crêtes d'Ondelette: CMOR	
Suivi de Formants		F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Occlusive voisée :	Diff	41	95	121	26	95	79	25	235	77	20	19	317	41	35	396	22	71	444
/d/	σ	8	18	24	5	19	16	5	38	15	5	4	61	9	7	78	5	13	86
Occlusive non voisée:	Diff	35	27	30	52	65	43	40	74	38	8	26	15	45	59	42	81	26	28
/q/	σ	6	4	5	10	9	7	6	11	6	2	5	5	8	10	8	14	4	5
Fricative voisée:	Diff	20	54	31	36	44	75	80	92	35	18	39	76	48	38	43	70	72	162
/H/	σ	2	9	4	4	5	9	8	10	4	2	5	11	6	5	7	9	9	21
Fricative non voisée:	Diff	20	122	134	14	65	54	35	154	63	24	55	34	47	20	39	32	52	10
/f/	σ	4	24	30	3	13	10	7	29	12	5	10	6	9	3	7	5	9	2
Nasale:	Diff	12	56	50	16	40	73	78	64	51	39	36	30	68	70	35	20	49	52
/m/	σ	2	7	7	2	6	10	10	9	7	7	7	6	11	12	7	4	8	10
Latérale:	Diff	24	31	64	38	38	187	20	106	64	16	15	23	26	32	42	33	40	37
/l/	σ	3	4	8	5	5	27	3	14	8	3	2	3	6	5	6	4	6	7
Tap:	Diff	30	54	187	44	81	89	97	90	66	36	161	356	43	219	120	41	128	149
/r/	σ	4	10	24	7	14	12	16	14	10	6	34	55	7	43	21	7	22	23
Semi-voyelle:	Diff	14	62	11	64	63	103	16	33	79	82	128	21	94	145	35	40	77	33
/w/	σ	2	8	2	9	9	17	2	6	12	18	21	4	15	26	5	6	14	5

When doing the same tests with female speakers, we found that overall formant tracking is correct for all three methods and three female speakers W3, W4 and W5. (In this paper we presented only the Table IV with two female speakers W3 and W4)

The results in table IV show that the formant tracking is generally good for all three formants of the short vowel / a / whatever the nature of the consonant which preceded it. This is true for all three methods Praat, Fourier and wavelet except in cases where the vowel is preceded by the consonants / r / and / d /, we notice that we always have a good F1 for the three methods even if there are some mistakes sometimes located especially in F3 or F2. This observation is verified for all three female speakers.

The case of the vowel / a / preceded by / d / can be explained by the fact that vowels are proceeded by dental consonants there is slight changes in F2 and F3, but this phenomenon is not specific to Arabic.

Then we tested our approach on all different long and short vowels. These tests were made to four speakers M1, M4, M2 and M3. In this paper we presented only the Table V with two male speakers M2 and M3.

We notice that the tracks F1, F2 and F3 are correct for vowels / a / and A /. This is true for all three methods and four speakers because there is practically no errors (red values) for these two cases. We notice from the results that errors occur most frequently in F3 for vowels / i / and / I / and specifically for vowels / u / and / V /. We also found that the results of wavelet and Fourier methods were often close in some cases but the proposed method is sometimes better.

However, the LPC method presented poor tracking formants concerning vowels / u / and / U /.

We did the same tests on the three female speakers W3, W4 and W5.

To sum up, we conclude that in the case of tests on long and short vowels with male speakers, we notice from the results that errors occur most frequently in F3 for vowels / i / and / I / and specifically for vowel / u / and / V /.

In the case of female speakers, the results showed that the formant tracking of F1, F2 and F3 is correct only for the vowels / a / and A /. The results obtained in the case of female speakers confirm the trend observed over the male speakers.

TABLE V
RESULTS OBTAINED ON THE TERMS OF SHORT AND LONG VOWELS EXTRACTED FROM SENTENCES ISSUED OF OUR DATABASE UTTERED BY TWO MALE SPEAKERS M2 AND M3.

		Loc.M2						Loc.M3					
		Praat			Crêtes de Fourier			Praat			Crêtes de Fourier		
Suivi de Formants		F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
a	Diff	26	34	114	9	12	72	25	68	138	54	58	35
	σ	6	7	25	2	3	17	7	14	33	11	12	7
A	Diff	54	99	60	22	44	23	46	61	49	26	37	69
	σ	10	17	13	4	7	4	8	11	9	4	6	10
i	Diff	28	53	161	18	160	208	31	54	84	21	29	206
	σ	12	20	73	7	57	71	10	20	28	7	8	69
I	Diff	64	47	83	29	148	182	61	89	91	29	151	197
	σ	20	14	25	9	42	49	20	29	31	11	50	67
u	Diff	26	55	90	9	43	92	61	67	358	40	93	295
	σ	10	20	31	3	14	26	24	17	114	13	22	69
U	Diff	70	224	169	30	39	85	60	138	196	111	95	112
	σ	25	140	82	9	11	48	19	89	67	34	30	30

		Loc.M2						Loc.M3					
		Crêtes d'Ondelette: CMOR						Crêtes d'Ondelette: CMOR					
Suivi de Formants		F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
a	Diff	38	17	44	17	184	213	61	89	91	33	150	253
	σ	8	3	9	3	36	21	8	11	9	5	6	28
A	Diff	17	36	21	14	67	263	31	54	84	58	18	346
	σ	3	7	4	3	27	88	10	20	28	20	7	90
i	Diff	14	67	263	5	53	57	20	29	31	11	50	71
	σ	5	27	88	5	30	24	7	8	59	7	8	59
I	Diff	17	184	213	47	90	86	40	93	358	55	55	441
	σ	5	53	57	14	9	24	13	22	69	19	18	103
u	Diff	47	30	86	23	90	93	111	95	112	64	109	302
	σ	14	9	24	8	30	43	34	30	30	19	32	84
U	Diff	23	90	93									
	σ	8	30	43									

IV. Conclusions

In this paper we present a new formant tracking algorithm where the formant frequencies estimation was based on the detection of instantaneous frequencies similar to local maxima of a time frequency representation issued from a complex wavelet transform. The formant frequencies candidates are validated as local maxima of time frequency representation which correspond to wavelet ridges. We explored our Arabic database labeled in terms of formants as reference to evaluate the proposed algorithm. Following tests on synthetic and real signals, we found that our proposed algorithm using CMOR wavelet transform present an accurate formant tracking compared to LPC analysis and refer to the reference.

Future work will target the improvement of this method especially for high frequency formants and its adaption on speaker's gender variability.

Acknowledgment

This work is supported by the CMCU (Comité Mixte Franco-Tunisien de Coopération Universitaire). Research Project CMCU, 11G1411.

References

[1] I.Bazzi, A.Acero, L.Deng., "An expectation maximization approach for formant tracking using a parameter-free non-linear predictor", in Proc. ICASSP. Vol.1, 464-467, 2003.

[2] A. Acero., "Formant Analysis and Synthesis using Hidden Markov Models". in Proc. of the Eurospeech Conference. Budapest, 1999.

[3] J A. M. A. Ali, J. V. D. Spiegel, P. Mueller., "Robust Auditory-based Processing using the Average Localized Synchrony Detection", in IEEE Transaction Speech and Audio Processing, 2002.

[4] S. Mallat., "A Wavelet Tour of Signal Processing". Academic Press, 1999.

[5] http://www.mathworks.com/help/toolbox/wavelet/ug/f8-24282.html#f8-40149

[6] F. Chaplais., "A wavelet tour of signal processing by Stéphane Mallat", A short presentation, 2001. http://cas.ensmp.fr/~chaplais/Wavtour_présentation/

[7] S. Châari, K. Ouni and N. Ellouze., « Wavelet Ridge Track Interpretation in Terms of Formants". in Proc. of INTERSPEECH-ICSLP, 1017-1020. Pittsburgh, Pennsylvania, USA, 2006

[8] I.Jemaa, O.Rekhis, K.Ouni et Y.Laprie, "An Evaluation of Formant Tracking Methods on an Arabic Database", in Proc. INTERSPEECH (International Speech), 1667-1670, Brighton, UK , 2009.

[9] M. Boudraa, B. Boudraa and B. Guerin., « Twenty Lists of Ten Arabic Sentences for Assessment". in Act of Communication ACUSTICA, 86:870-882, 2000.

[10] http://www.loria.fr/~laprie/WinSnoori/.

[11] S. Ghazeli., "Back consonants and backing coarticulation in Arabic", PhD, University of Texas, Austin, 1977.

[12] A. Brahim, "An Acoustic study of temporal organization in Arabic specific to Tunisian speakers", PhD, University of Manouba, Tunis, 1997.

[13] I.Jemaa, K.Ouni et Y.Laprie, "Evaluation of Automatic Formant tracking Method using Fourier Ridges" In Journal Cognitive Computation, 170-179, 2010.

[14] http://www.praat.org/