We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Multi-Resolution Spectral Analysis of Vowels in Tunisian Context

Nefissa Annabi-Elkadri, Atef Hamouda and Khaled Bsaies
URPAH Research Unit, Computer Science Department, Faculty of Sciences of Tunis.
Tunis El Manar University, Tunis
Tunisia

1. Introduction

Classic speech spectrogram shows log-magnitude amplitude (dB) versus time and frequency. The sound pressure level in dB is approximately proportional to the volume perceived by the ear. The classic speech sonagram offers a single integration time which is the length of the window. It implements a uniform bandpass filter, the spectral samples are regularly spaced and correspond to equal bandwidths. The choice of the window length determines the time-frequency resolution for all frequencies of sonagram. The more the window is narrower, the better the time resolution and the worse the frequency resolution. This implies that the display resolution of formants, voicing and frictions at low frequencies is less good than the resolution of the bursts in the high frequencies and vice versa. It is so necessary to make the right choice of windows compared to the signal.

(Mallat, 1989, p.674) remarks *"it is difficult to analyze the information content of an image directly from the gray-level intensity of the image pixels... Generally, the structures we want to recognize have very different sizes. Hence, it is not possible to define a priori an optimal resolution for analyzing images."*. To improve the standard spectral output, we can calculate a multi-resolution (MR) spectrum. In the original papers, the MR analysis is based on discrete wavelet transforms (Grossmann & Morlet, 1984; Mallat, 1989; 2000; 2008). Since that it has been applied to several domains: image analysis (Mallat, 1989), time-frequency analysis (Cnockaert, 2008), speech enhancement (Fu & Wan, 2003; Manikandan, 2006), automatic signal segmentation by search of stationary areas from the scalogram (Leman & Marque, 1998).

The MR spectrum, a compromise that provides both a higher frequency and a higher temporal resolution, is not a new method. In phonetic analysis, (Annabi-Elkadri & Hamouda, 2010; 2011) presents a study of two common vowels [a] and [E] in Tunisian dialect and french language. Vowels are pronounced in Tunisian context. The analysis of the obtained results shows that due to the influence of french language on the Tunisian dialect, the vowels [a] and [E] are, in some contexts, similarly pronounced. Annabi-Elkadri & Hamouda (2011 (in press) applies the MRS for an automatic method for Silence/Sonorant/Non-Sonorant detection used the ANOVA method. Results are compared to the classical methods for classifications such as Standard Deviation and Mean with ANOVA who were better. The method for automatic Silence/Sonorant/Non-Sonorant detection based on MRS provides better results when compared to classical spectral analysis. Cheung & Lim (1991) presented a method for

combining the wideband spectrogram and the narrowband spectrogram by evaluating the geometric mean of their corresponding pixel values. The combined spectrogram appears to preserve the visual features associated with high resolution in both frequency and time. Lee & Ching (1999) described an approach of using multi-resolution analysis (MRA) for clean connected speech and noisy phone conversation speech. Experiments show that the use of MRA cepstra results reduces significantly error insertion when compared with MFCCs. For music signals, Cancela et al. (2009) presents two algorithms: efficient constant-Q transform and multi-resolution FFT. They are reviewed and compared with a new proposal based on the IIR filtering of the FFT. The proposed method shows to be a good compromise between design flexibility and reduced computational effort. Additionally, it was used as a part of an effective melody extraction algorithm. In this context, Dressler was interested in the description of spectral analysis to extract melodies based on spectrograms multi-resolution (Dressler, 2006). The approach aims to extract the sinusoidal components of the audio signal. A calculation of the spectra of different resolutions of frequencies is done in order to detect sinusoids stable in different frames of the FFT. The evaluated results showed that the multi-resolution analysis improves the extraction of the sinusoidal.

The aim of this study was an extension of Annabi-Elkadri & Hamouda (2010; 2011) researches. We presented and tested the concept of multi-resolution for the spectral analysis (MRS) of vowels in Tunisian words and in French words under the Tunisian context. Our method was composed of two parts. The first part was applied MRS method to the signal. MRS was calculated by combining several FFT of different lengths (Annabi-Elkadri & Hamouda, 2010; 2011). The second part was the formant detection by applied multi-resolution LPC (Annabi-Elkadri & Hamouda, 2010). We present an improvement of our method of multi-resolution spectral analysis MR FFT. As an application, we used our system VASP for a Tunisian Dialect corpus pronounced by Tunisian speakers.

Standard Arabic is composed by 34 phonemes (Muhammad, 1990). It has three vowels, with long and short forms of [a], [i], and [u]. Arabic phonemes are classified in two classes pharyngeal and emphatic. There are characteristics of semitic languages (Elshafei, 1991; Muhammad, 1990). Arabic has two kinds of syllables: open syllables (CV) and (CVV) and closed syllables (CVC), (CVVC), and (CVCC). Syllables (CVVC) and (CVCC) occur only at the end of the sentence. V is a vowel and C is a consonant (Muhammad, 1990).

In section 2, we presented a brief history of Tunisian Dialect and it's relationship with Arabic and French. In section 3, we presented our calculated method of the multi-resolution FFT. In section 4, we presented the materials and methods composed by our corpus and our system Visual Assistance of Speech Processing Software (VASP). We presented our experimental results in section 5 and we discussed it in section 6. Our conclusion is presented in section 7.

2. History of Tunisian dialect and it's relationship with Arabic and French

The official language in Tunisia is Arabic. But, the popular language is the Tunisian Dialect (TD). It is a mix of Arabic with a lot of other languages: French, Italian, English, Turkich, German, Berber and Spanish. This mixture is related to the history of Tunisia, since it was invaded and colonized by many civilizations like the Romans, Vandals, Byzantains, the Arab Moslems and French.

After French colonization, the French government wanted to spread the French language in the country. The French instituted a bilingual education system with the Franco-Arab schools. Programs of bilingual schools were modeled primarily on the model of French primary education for children of European origin (French, Italian and Maltese), which were added courses in colloquial Arabic. As for the Tunisian children, they received their education in classical Arabic in order to study the Quoran. Only a small Tunisian elite received a truly bilingual education, in order to co-administer the country. Tunisian Muslim mass continued to speak only Arabic or one of its many varieties. The report of the Tunisian Minister of Affairs, Jean-Jules Jusserand, pursuing the logic of Jules Ferry. In a "Note on Education in Tunisia", dated February 1882, Jusserand exposing his ideas: "*We have not at this time we better way to assimilate the Arabs of Tunisia, to the extent that is possible, that they learn our language, it is the opinion of all who know them best: we can not rely on religion to make this comparison, they do not convert to Christianity, but as they learn our language, a host of European ideas will prove to be bound to them, as experience has sufficiently demonstrated. In the reorganization of Tunisia, a large part must be made to education*".

After independence, education of the French language such as Arabic was required for all Tunisian children in primary school. This explains why French has become the second language in Tunisia. It is spoken by the majority of the population.

There are different varieties of TD depending on the region, such as dialect of Tunis, Sahel, Sfax, etc. Its morphology, syntax, pronunciation and vocabulary are quite different from the Arabic (Marcais, 1950). There are several differences in pronunciation between Standard Arabic and TD. Short vowels are frequently omitted, especially where they would occur as the final element of an open syllable. While Standard Arabic can have only one consonant at the beginning of a syllable, after which a vowel must follow, TD commonly has two consonants in the onset. For example Standard Arabic "book" is /*kita?b*/, while in TD, it is /*kta?b*/. The nucleus in TD may contain a short or long vowel, and at the end of the syllable, in the coda, it may have up to three consonants, but in standard Arabic, we cannot have more than two consonants at the end of the syllable. Word-internal syllables are generally heavy in that they either have a long vowel in the nucleus or consonant in the coda. Non-final syllables composed of just a consonant and a short vowel (i.e. light syllables) are very rare in TD, and are generally loaned from standard Arabic: short vowels in this position have generally been lost, resulting in the many initial CC clusters. For example /*?awa?b*/ "reply" is a loan from Standard Arabic, but the same word has the natural development /*?wa?b*/, which is the usual word for "letter" (Gibson, 1998).

In TD's non-pharyngealised context, there is a strong fronting and closing of /*a?*/, which, especially among younger speakers in Tunis can reach as far as /*e?*/, and to a lesser extent of /*a*/.

This is an example of Tunisian Arabic sentence (SAMPA and X-SAMPA symbols): '/*ddZ   bA : k   Ukil ?\adailkas mta?\u   milEna bil   lE   sykse*/'. This sentence is a mixture of three languages; '/*ddZ   bA : k*/' in English, '/*Ukil   ?\ada   ilkas mta?\u milEna   bil*/' in Tunisian Arabic, which means '*as usual the show is interesting*' and finally '/*lE   sykse*/' in French, which means '*success*'.

We introduce a study of six common vowels [a], [E], [i], [e], [o] and [u] in TD and French. Vowels are pronounced in Tunisian context. Our study is realized in time-frequency domain.

3. Multi-resolution FFT

It's so difficult to choose the ideal window with the ideal characteristics. The size of the ideal window (Hancq & Leich, 2000) was equal to twice the length of the pitch of the signal. A wider window show the harmonics in the spectrum, a shorter window approximated very roughly the spectral envelope. This amounts to estimate the energy dispersion with the least error. When we calculated the windowed FFT, we supposed that the eneregy was concentrated at the center of the frame (Haton & al., 2006, p.41). We noted the center C_p. So our problem now, is the estimation of C_p.

3.1 The center estimation in the case of the Discrete Fourier Transform (DFT)

We would like to calculate the spectral of the speech signal s. We note L the length of s. The first step is to sample s into frames. The size of each frame was between 10 ms and 20 ms (Calliope, 1989; Ladefoged, 1996) to meet the stationnarity condition. We choosed the Hamming window and we fixed the size to 512 points and the overlap to 50%. Figure 1 shows the principle of the center estimation.
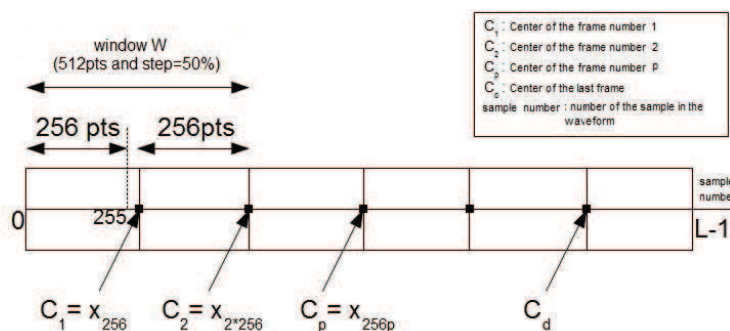

Fig. 1. Signal sampling and windowing for center estimation C_p. The window length $N = 512$ points and overlap = 50%.

For each frame p, the center C_p was estimated:

$$\begin{cases} C_1 = x_{256} & for \quad p = 1 \\ C_2 = x_{2*256} & for \quad p = 2 \\ \quad \vdots \\ C_p = x_{256p} & in \quad general \quad case \end{cases}$$

The center $C_p = x_{256p}$ with $p = 1...[\frac{L-1}{256} - 1]$ and $[\quad]$ the integer part.

Each signal s was sampled into frames. Each frame number p was composed by $N = 512$ points:

$$\begin{cases} s_0(p) = x_{256(p-1)} \\ s_1(p) = x_{256(p-1)+1} \\ \quad \vdots \\ s_{511}(p) = x_{256(p-1)+511} \end{cases}$$

In general case, for the componant number l of s:

$$s_l(p) = x_{256(p-1)+l}$$

The FFT windowing for the frame number p was calculated as:

$$S_k(p) = \sum_{l=0}^{511} s_l(p) e^{-\frac{2j\pi kl}{512}} w(s_l(p) - s_{256}(p)) \tag{1}$$

In general case:

$$S_k(p) = \sum_{l=0}^{N-1} s_l(p) e^{-\frac{2j\pi kl}{N}} w(s_l(p) - s_{\frac{N}{2}}(p)) \tag{2}$$

We noted $C_p = s_{\frac{N}{2}}(p)$ the center of the frame number p with $p = 1...[\frac{2(L-1)}{N} - 1]$, $[\quad]$ the integer part and :

$s_l(p)$: the componant of s number l of the frame p
$S_k(p)$: the componant of S number k of the frame p
L: the length of the signal s
N: the length of the window w

3.2 The center estimation in the case of the MR FFT

To improve the standard spectral, we calculated the MR FFT by combining several FFT of different lengths. The temporal accuracy is higher in the high frequency region and the resolution of high frequency in the low frequencies.

We calculated the FFT windowing of the signal several times NB. The number of steps NB was equal to the number of band frequency fixed a priori. For each step number i ($i \leq NB$), the signal s was sampled into frames $s_i(p_i)$ and windowed with the window w. We noted N_i the length of frames and of w for each step i. C_{i,p_i} was the center of w.

The spectral $S_{i,k}(p_i)$ for each step i was:

$$S_{i,k}(p_i) = \sum_{l=0}^{N_i-1} s_{i,l}(p_i) e^{-\frac{2j\pi kl}{N}} w(s_{i,l}(p_i) - C_{i,p_i}) \tag{3}$$

with: $C_{i,p_i} = s_{i,\frac{N_i}{2}}(p_i)$ the center of the frame p_i when the overlap$= \frac{N_i}{2}$.

In MRS, the overlap $\frac{N_i}{2}$ can not satisfy the principle of continuity of the MRS in different band frequencies. A low overlap causes a discontinuity in the spectrum MRS and thus give us a bad estimation of the energy dispersal. So our problem consisted on the overlap choosing. It was necessary that the frames overlap with a percentage higher to 50% of the frame length. We choosed an overlap equal to 75% (fig. 2).

For the frame $p_i = 1$ of the step number i, we have N_i components:

$$\begin{cases} s_0(1) &= x_0 \\ s_1(1) &= x_1 \\ &\vdots \\ s_l(1) &= x_l \\ &\vdots \\ s_{N_i-1}(1) &= x_{N-1} \end{cases}$$
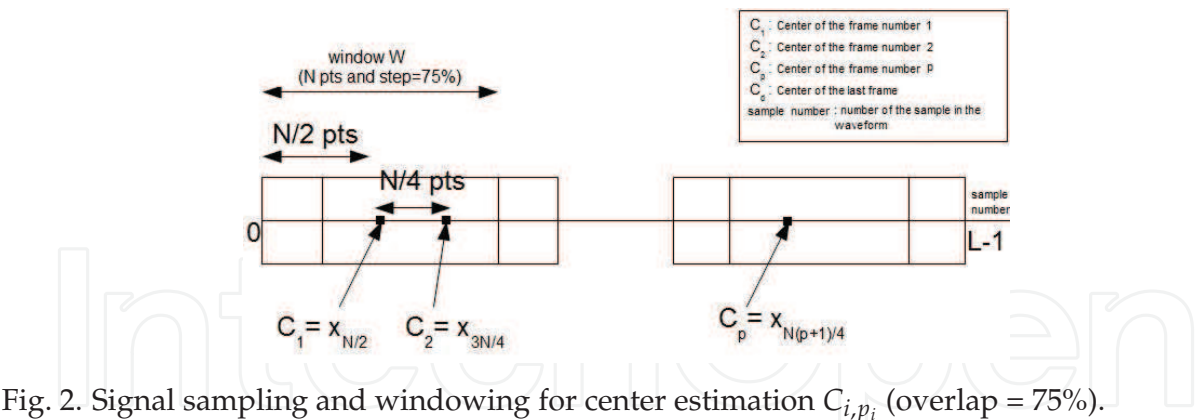
Fig. 2. Signal sampling and windowing for center estimation C_{i,p_i} (overlap = 75%).

For the frame $p_i = 2$ of the step number i, we have N_i components:

$$\begin{cases} s_0(2) &= & x_{N_i} \\ s_1(2) &= & x_{N_i} + 1 \\ & \vdots & \\ s_l(2) &= & x_{N_i} + l \\ & \vdots & \\ s_{N_i-1}(2) &= & x_{N_i} + N_i - 1 \end{cases}$$

In general case, for the frame p_i of the step number i, we have N_i components:

$$\begin{cases} s_0(p_i) &= & x_{(p_i-1)N_i} \\ s_1(p_i) &= & x_{(p_i-1)N_i} + 1 \\ & \vdots & \\ s_l(p_i) &= & x_{(p_i-1)N_i} + l \\ & \vdots & \\ s_{N_i-1}(p_i) &= & x_{(p_i-1)N_i} + N_i - 1 \\ s_{N_i-1}(p_i) &= & x_{p_i N_i} - 1 \end{cases}$$

The center C_{i,p_i} of $p_i = 1$ was:

$$C_{i,1} = \frac{N_i}{2}$$

The center C_{i,p_i} of $p_i = 2$ was:

$$\begin{cases} C_{i,2} = \frac{1}{2}(\frac{1}{4} + \frac{5}{4})N_i \\ = \frac{3}{4} N_i \end{cases}$$

In general case, the center C_{i,p_i} of p_i was:

$$\begin{cases} C_{i,p_i} = & C_{i,p_i-1} + \frac{N_i}{4} \\ = & C_{i,1} + (p_i - 1)\frac{N_i}{4} \\ = & x_{\frac{N_i(p_i+1)}{4}} \end{cases}$$

with : $\frac{N_i(p_i+1)}{4} \leq L$ and $p_i \leq \frac{4L}{N_i} - 1$

The spectral $S_{i,k}(p_i)$ of each step i was :

$$S_{i,k}(p_i) = \sum_{l=0}^{N_i-1} s_{i,l}(p_i)e^{-\frac{2j\pi kl}{N}}w(s_{i,l}(p_i) - C_{i,p_i}) \qquad (4)$$

with: $C_{i,p_i} = x_{\frac{N_i(p_i+1)}{4}}$ the center of the frame p_i and the overlap equal to 75%.

So, the multi-resolution spectral MRS was:

$$S_k(p) = S_{i,k}(p_i)sik_i \leq k \leq k_{i+1} \qquad (5)$$

with: $0 \leq k \leq N_0 + N_1 + \ldots + N_P$ and $\quad 1 \leq p \leq P$.

The size of the ideal window (Hancq & Leich, 2000) is equal to twice the length of the pitch of the signal. A wider window shows the harmonics in the spectrum, a shorter window approximates very roughly the spectral envelope.

To improve the standard spectral, we calculate a multi-resolution spectral (MRS) with two methods; by combining several FFT of different lengths and by combining several windows for each FFT (Annabi-Elkadri & Hamouda, 2010).

Diagrams displayed in figure 3 illustrates the difference between the standard FFT and the MRS. For a standard FFT, the size of the window is equal for each frequency band unlike the MRS windows size. It is dependent on the frequency band.
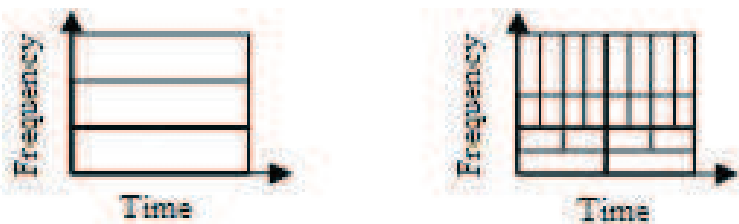

Fig. 3. Standard FFT (on the left) and MR FFT (on the right)

Figure 4 shows the classical sonagram; Hamming window, 11 ms with an overlap equal to 1/3. The sentence pronounced is: "Le soir approchait, le soir du dernier jour de l'anné". Figure 5 shows the multi-resolution sonagram of the same sentence. It offers several time integrations which are combinations of several FFT of different lengths depending on frequency bandwidth.

4. Materials and methods

4.1 Corpus

Our corpus is composed of TD prounounced by Tunisian speakers. The sampling frequency is equal to 44.1 KHz, the wav format was adopted in mono-stereo. We avoided all types
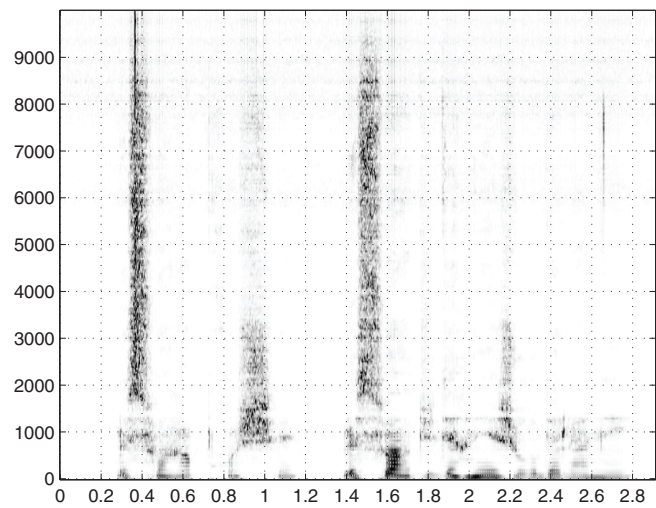
Fig. 4. Classical sonagram (Hamming, 11 ms, overlap 1/3) of this sentence: "Le soir approchait, le soir du dernier jour de l'année"
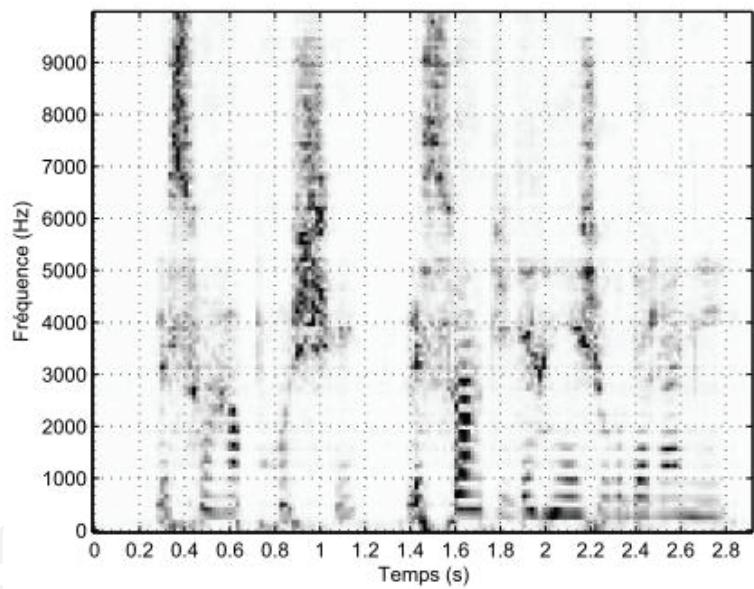

Fig. 5. MR sonagram; Hamming (23, 20, 15, 11 ms), overlap 75%, Band-limits in Hz were [0, 2000, 4000, 7000, 10000] of this sentence: "Le soir approchait, le soir du dernier jour de l'année"

of noise filter that would degrade the quality of the signal and thus, causes information lost. We have recorded a real time spontaneous lyrics and discussions of 4 speakers. We have removed noise and funds sounds like laughing, music, etc. It was difficult to realize a spontaneous corpus because, in real time, it is impossible to have all phonemes and syllables. Another difficulty was the variability of discussion themes and pronounced sounds. For these reasons, we decided to complete our corpus with another one. We prepared a text in Tunisian dialect with all sounds to study. Every phoneme and syllable appeared 15 times. We asked from four speakers: two men and two women, to read the text in a high voice in the same conditions of the first corpus records. All speakers are between 25 and 32 years old.

Speakers don't know the text. The sampling frequency is equal to 44,1 KHz, the wav format was adopted in mono-stereo. We avoided the remarkable accents and all types of noise filter that would degrade the quality of the signal and thus, causes information lost. Our corpus was transcribed in sentences, words and phonemes.

4.2 VASP software: Visual assistance of speech processing software

For our study, we have created our first prototype System for Visual Assistance of Speech Processing VASP. It offers many functions for speech visualization and analysis. We developed our system with GUI Matlab. In the following subsection, we will present some of the functionalities offered by our system.

VASP reads sound files in wav format. It represent a wav file in time domain by waveform and in time-frequency domain by spectral representation, classical spectrogram in narrow band and wide band (see figure 6), spectrograms calculated with linear prediction and cepstral coefficients, gammatone, discrete cosine transform (DCT), Wigner-Ville transformation, multi-resolution LPC representation (MR LPC), multi-resolution spectral representation (MR FFT) and multi-resolution spectrogram.
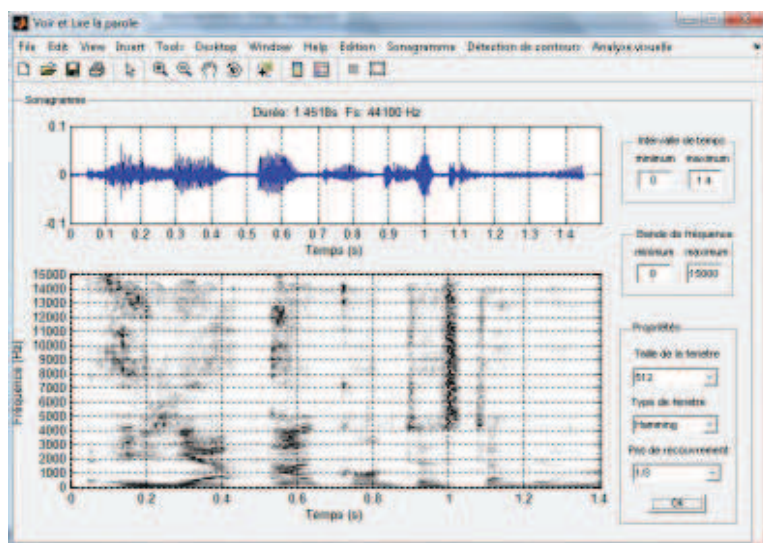

Fig. 6. Classical Spectrogram Interface.

From waveform, we can choose, in real time, the frame for which we want to represent a spectrum (see figure 7). Parameters are manipulated from a menu; we can select the type of windows (Hamming, Hanning, triangular, rectangular, Kaiser, Barlett, gaussian and Blackmann-Harris), window length (64, 128, 256, 512, 1024 and 2048 samples) and LPC factor.

From all visual representations, coordinates of any pixel can be read. For example, we can select a point from a spectrogram and read its coordinates directly (time, frequency and intensity).

VASP offers the possibility to choose a part of a signal to calculate and visualize it in any time-frequency representations.

Our system can automatically detect Silence/Speech from a waveform. From the spectrogram, the system can detect acoustic cues like formants, and classify it automatically to two classes: sonorant or non-sonorant.
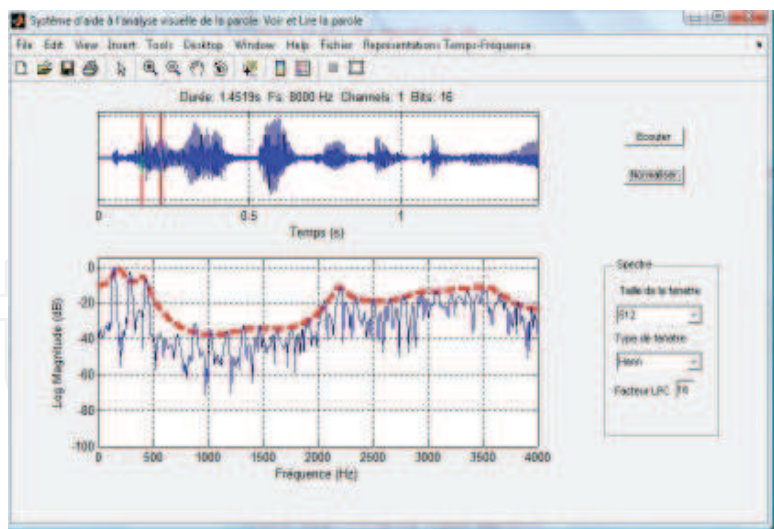
Fig. 7. Waveform and spectrum of the selected signal.

Our system can analyse visual representations with two methods image analysis with edge detection and sound analysis signal. Edge detection is calculated with gradiant method or median filter method (fig.8). The second method is based on detecting energy from a time-frequency representation.
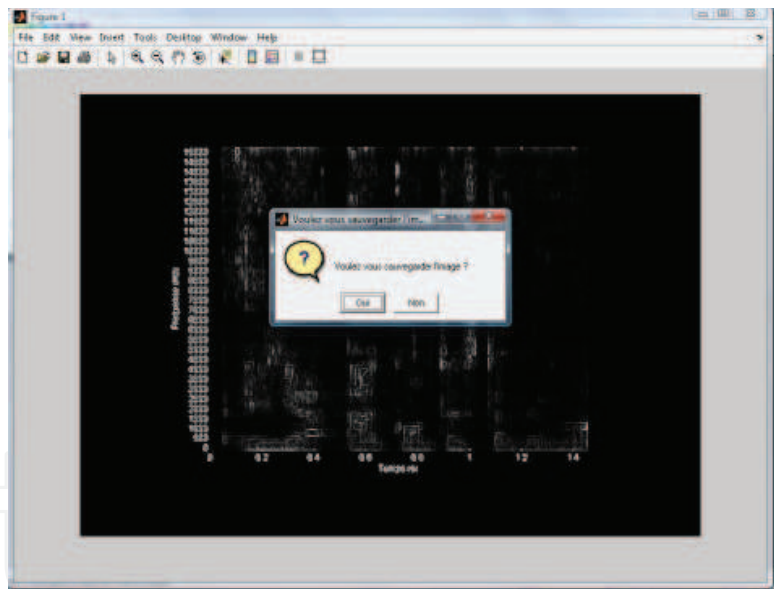

Fig. 8. Edge detection calculated with the gradiant method.

5. Experimental results

Formants frequencies are the properties of the vocal tract system and need to be inferred from the speech signal rather than just measured. The spectral shape of the vocal tract excitation strongly influences the observed spectral envelope. However, not all vocal tract resonances can cause peaks in the observed spectral envelope.

To extract formants frequencies from the signal, we resampled it to 8 KHz. We use a linear prediction method for our analysis. Linear prediction models the signal as if it were

generated by a signal of minimum energy being passed through a purely-recursive IIR filter. Multi-resolution LPC (MR LPC) is calculated by the LPC of the average of the convolution of several windows to the signal.

To our knowledge, there is no normative studies in Standard Arabic vowels like those of Peterson and Barney (Peterson & Barney, 1952) for American english, and those of Fant and al. (Fant, 1969) for Swedish.

We applied VASP on TD and French language. We measured the two first formants F1 and F2 of vowels in Tunisian and French words. We compared our experimental results with those for Calliope (Calliope, 1989). Their corpus is constituted of vowels repetitions in [p_R] context and pronounced by 10 men and 9 women.

The analysis of the obtained results shows that due to the influence of French language on the TD, the vowels [a], [E], [i], [e], [o] and [u] are, in some contexts, similarly pronounced.

5.1 Tunisian dialect and French spectral analysis of [a] in Tunisian context

We measured the two first formants F1 and F2 of vowels [a] in Tunisian and french words in Tunisian context. Figures 9(a) and 9(b) show scatters of formants respectively for [a] in Tunisian words and in french words. There was a high matching of the two scatters for F1 (400-700 Hz) and for F2 (500-3000 Hz).
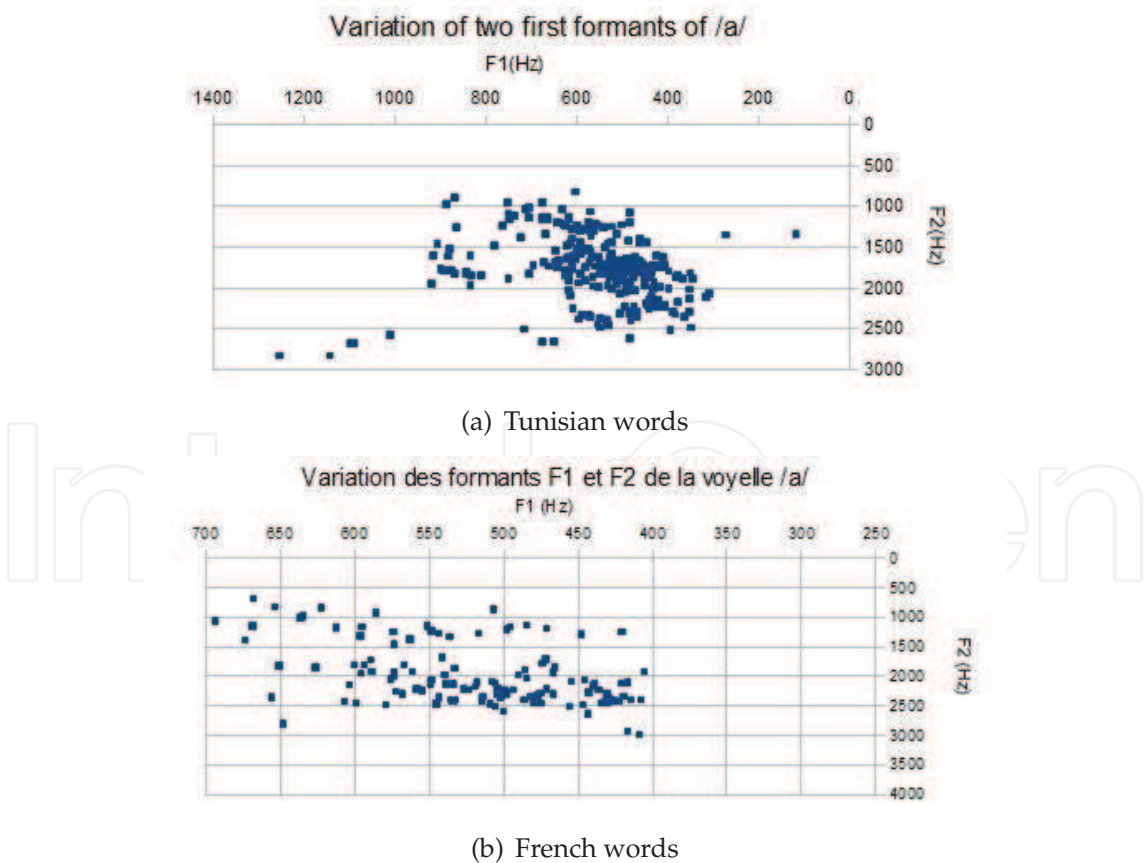
(a) Tunisian words

(b) French words

Fig. 9. Variation of two first formants of [a].

5.2 Tunisian dialect and French spectral analysis of [E] in Tunisian context

We measured the two first formants F1 and F2 of vowels [E] in Tunisian and french words in Tunisian context. Figures 10(a) and 10(b) show scatters of formants respectively for [E] in Tunisian words and in french words. There is a high matching of the two scatters for F1 (300-550 Hz) and for F2 (1500-3000 Hz).
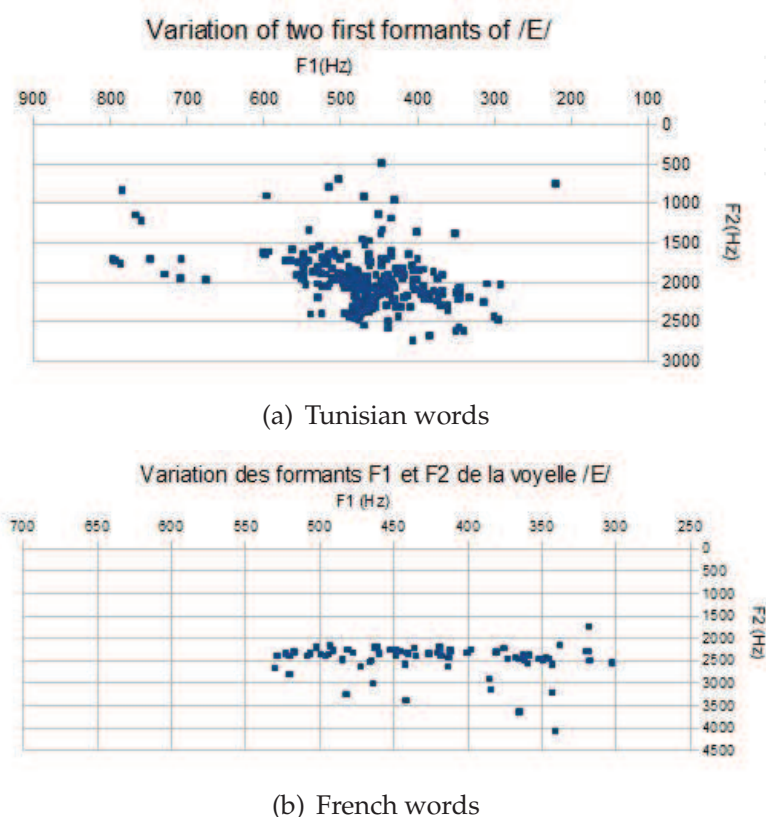

(a) Tunisian words

(b) French words

Fig. 10. Variation of two first formants of [E].

5.3 Tunisian dialect and French spectral analysis of [i] in Tunisian context

We measured the two first formants F1 and F2 of vowels [i] in Tunisian and french words in Tunisian context. Figures 11(a) and 11(b) show scatters of formants respectively for [i] in Tunisian words and in french words. There is a high matching of the two scatters for F1 (250-400 Hz) and for F2 (1800-2500 Hz).

5.4 Tunisian dialect spectral analysis of [o] and [e] in Tunisian context

We measured the two first formants F1 and F2 of vowels [o] and [e] in Tunisian words. Figure 12(a) and 12(b) show scatters of formants for [o] and [e] in Tunisian words.

5.5 Tunisian dialect and French spectral analysis of [u] in Tunisian context

We measured the two first formants F1 and F2 of vowels [u] in Tunisian and french words in Tunisian context. Figures 13(a) and 13(b) show scatters of formants respectively for [u] in Tunisian words and in french words. There is a high matching of the two scatters for F1 (300-440 Hz) and for F2 (1500-3000 Hz).
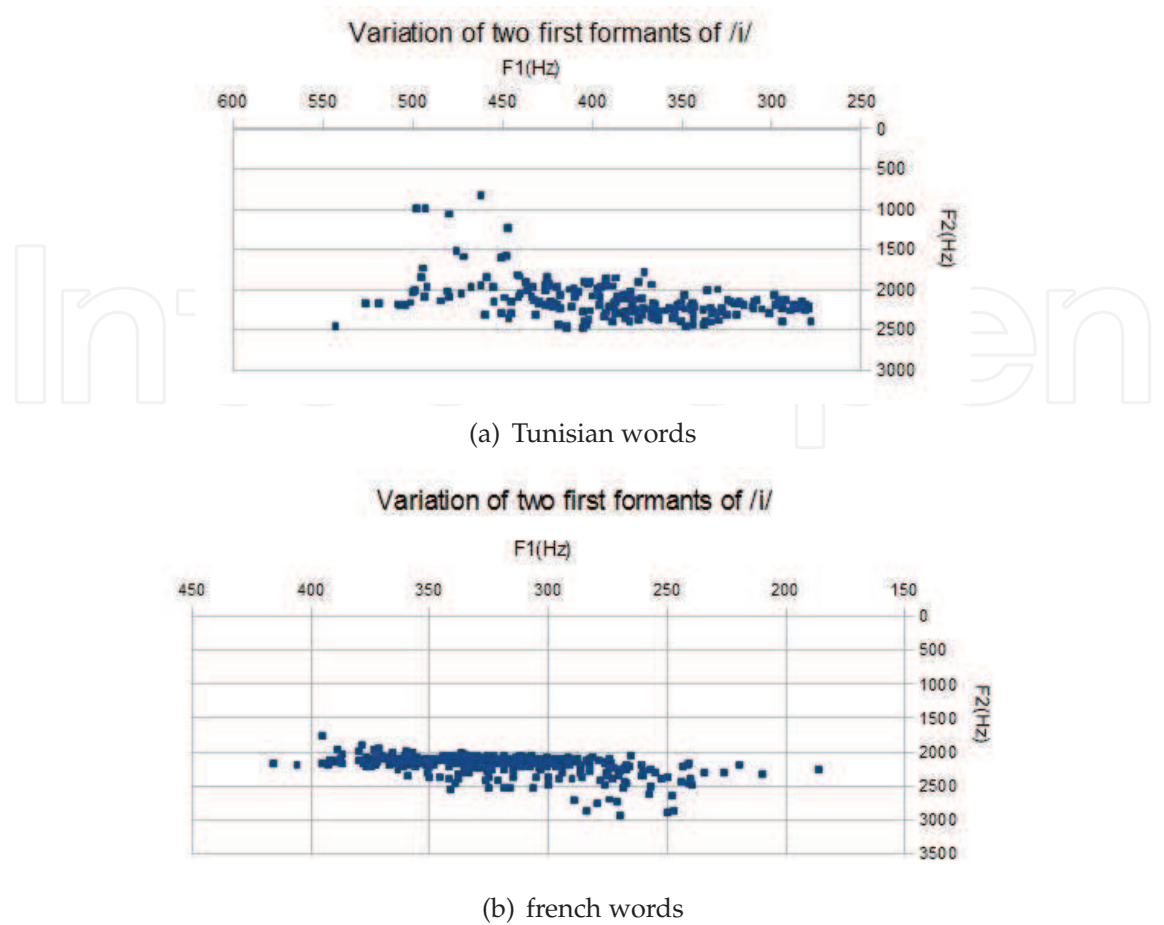
(a) Tunisian words

(b) french words

Fig. 11. Variation of two first formants of [i].

5.6 Results of Tunisian dialect formants variation

Figure 14 show scatters of formants mean for vowels [a], [e], [E], [o], [i] and [u] in Tunisian words.

6. Discussion

We compared our experimental results with those for Calliope (1989) in tables 1, 2, 3, 4 and 5. Their corpus was constituted of two repetitions of [a], [E], [i], [o], [e] and [u] in [p_R] context and pronounced by 10 men and 9 women.

	Calliope		French		Tunisian Dialect	
	F1	F2	F1	F2	F1	F2
Med	684	1256	512	2137	515	1758
SD	47	32	74	600	158	383

Table 1. Values of median (Med) and Standard Deviation (SD) of two first formants (F1, F2) of [a] for Tunisian Dialect and French language

In French, [i] was a front vowel. The difference between the two first formants was great. F1 was around 200 Hz and F2 was greater than 2000 Hz. For the Tunisian dialect, we note that
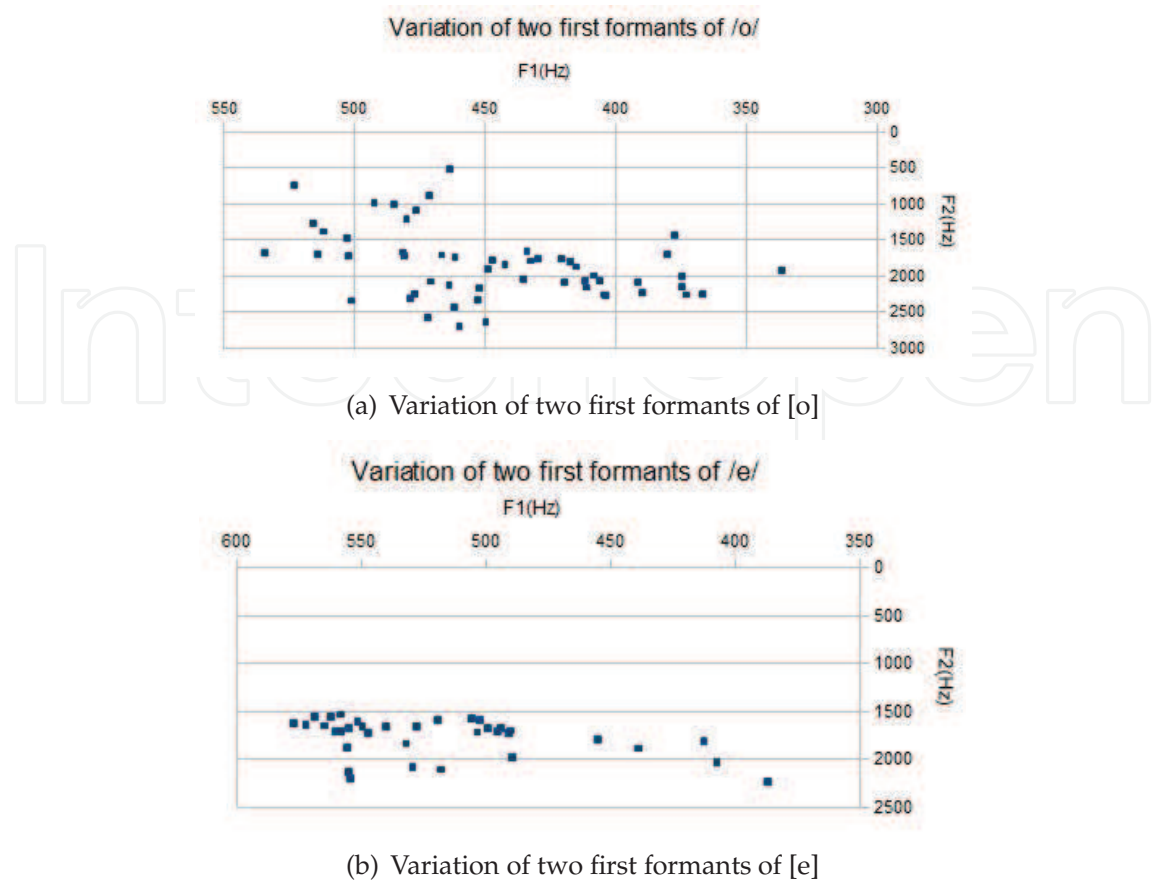
(a) Variation of two first formants of [o]

(b) Variation of two first formants of [e]

Fig. 12. Variation of two first formants of [o] and [e] for Tunisian words.

	Calliope		French		Tunisian Dialect	
	F1	F2	F1	F2	F1	F2
Med	530	1718	436	2374	464	2017
SD	49	132	63	354	90	360

Table 2. Values of median (Med) and Standard Deviation (SD) of two first formants (F1, F2) of [E] for Tunisian Dialect and French language

	Calliope		French		Tunisian Dialect	
	F1	F2	F1	F2	F1	F2
Med	308	2064	307.69	2175.12	384.3	2181.32
SD	34	134	415.37	241.61	61.57	249.6

Table 3. Values of median (Med) and Standard Deviation (SD) of two first formants (F1, F2) of [i] for Tunisian Dialect and French language

the [i] retains the characteristic of front vowel. The average value of F1 was 443 Hz, more higher than that of the French language. It can be considered as a lax variant of the vowel [i]. It was like close-mid front vowel. The tongue was positioned halfway between a close vowel and a mid vowel but it was less constricted.
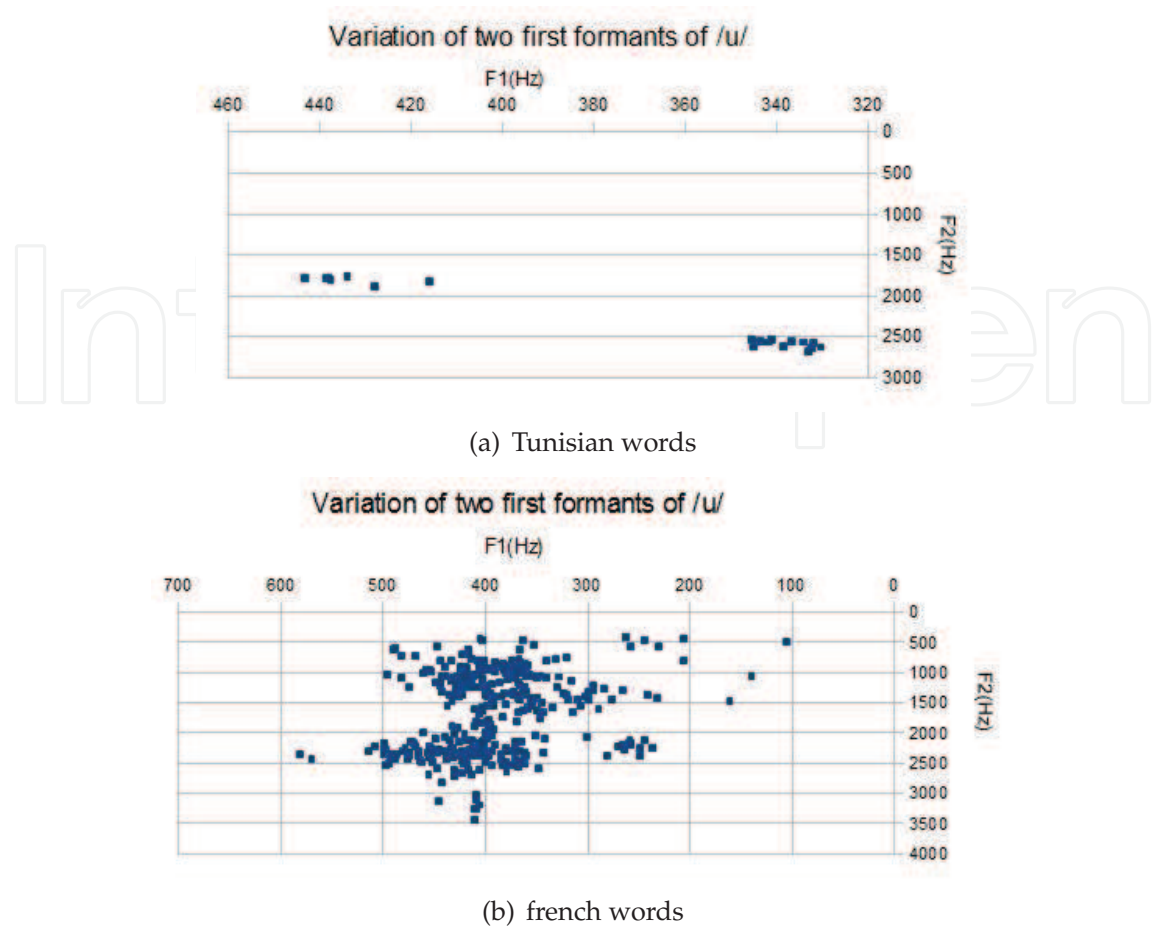
(a) Tunisian words

(b) french words

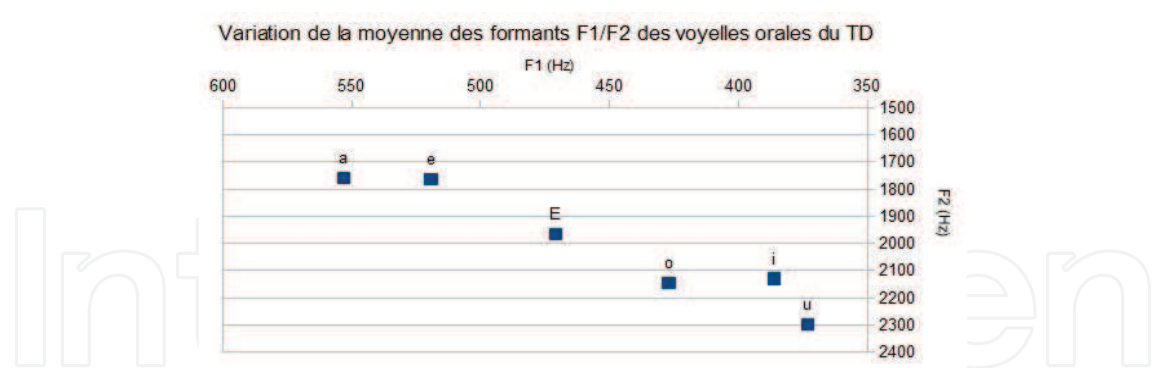Fig. 13. Variation of two first formants of [u].

Fig. 14. Variation of two first formants of Tunisian Dialect Vowels.

	[o] for Calliope		[e] for Calliope		[o] for Tunisian Dialect		[e] for Tunisian Dialect	
	F1	F2	F1	F2	F1	F2	F1	F2
Med	383	793	381	1417	381.4	2184.32	380.23	2188.11
SD	22	63	44	106	37.88	258.24	39.85	153.44

Table 4. Values of median (Med) and Standard Deviation (SD) of two first formants (F1, F2) of [e] and [o] for Tunisian Dialect and French language

	Calliope		French		Tunisian Dialect	
	F1	F2	F1	F2	F1	F2
Med	315	764	575.52	2355.61	381.58	2183.57
SD	43	59	199.38	172.25	48.05	392.42

Table 5. Values of median (Med) and Standard Deviation (SD) of two first formants (F1, F2) of [u] for Tunisian Dialect and French language

We can say the same thing about [u] for Tunisian dialect. The average value of F1, equal to 406 Hz, was greater than that of the French language (200 Hz) and the average value of F2, equal to 2195 Hz, was higher than that of the French language (1200 Hz).

The vowel [a] was open front. The average value of F1 was equal to 443 Hz and the average value of F2 was equal to 2090 Hz for Tunisian dialect. F1 was less than the average of first French formant (700 Hz). F2 takes a greater value than the average second French formant (1500 Hz). Therefore, we can say that the position of the tongue was narrower and it was positioned as far forward as possible in the mouth for the Tunisian dialect.

In opposition to all vowels studied, the [E] of the Tunisian dialect was same to the [E] of the French language. We note an average of 436 Hz for F1 and 2120 Hz for F2.

For the vowels [a] and [E] the F1 median was nearer to Calliope for TD than for French language. For TD and French language, the F1 medians are far from Calliope for F2 median but near each other. For the vowel [E], the F1 and F2 Standard Deviation was high and far from Calliope for TD and French language. This may be explained by the facts that the position of the tongue was narrower and it was positioned as far forward as possible in the mouth for the Tunisian dialect. High Standard Deviation was related to the small size of our corpus.

We remarked that Tunisian speakers pronounce vowels in the same way for both French language and TD.

7. Conclusion

The analysis of the obtained results shows that due to the influence of French language on the Tunisian dialect, the vowels are, in some contexts, similarly pronounced. It will be interesting to extend the study to other vowels, on a large corpus and to compare it with the study of other languages corpus like Standard Arabic, Berber, Italian, English and Spanish.

8. References

Annabi-Elkadri, N. & Hamouda, A. (2010). Spectral analysis of vowels /a/ and /E/ in tunisian context, *2010 International Conference on Audio, Language and Image Processing*, number CFP1050D-ART in *978-1-4244-5858-5*, IEEE/IET indexed in both EI and ISTP. (in Press).

Annabi-Elkadri, N. & Hamouda, A. (2011). Analyse spectrale des voyelles /a/ et /E/ dans le contexte tunisien, *Actes des IXe Rencontres des Jeunes Chercheurs en Parole RJCP*, Université Stendhal, Grenoble, pp. 1–4.

Annabi-Elkadri, N. & Hamouda, A. (2011 (in press)). Automatic Silence/ Sonorant/ Non-Sonorant Detection based on Multi-resolution Spectral Analysis and ANOVA

Method, *International Workshop on Future Communication and Networking*, IEEE, Hong Kong.

Calliope (1989). *La parole et son traitement automatique*, collection technique et scientifique des télécommunications, MASSON et CENT-ENST, Paris, ISBN :2-225-81516-X, ISSN : 0221-2579.

Cancela, P., Rocamora, M. & Lopez, E. (2009). An efficient multi-resolution spectral transform for music analysis, *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 309–314.

Cheung, S. & Lim, J. (1991). Combined multi-resolution (wideband/narrowband) spectrogram, *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-91*, IEEE, pp. 457–460.

Cnockaert, L. (2008). *Analysis of vocal tremor and application to parkinsonian speakers / Analyse du tremblement vocal et application à des locuteurs parkinsoniens*, PhD thesis, F512 - Faculté des sciences appliquées - Electronique.

Dressler, K. (2006). Sinusoidal extraction using an efficient implementation of a multi-resolution FFT, *Proceeding of the 9th International Conference on Digital Audio Effects (DAFx-06)*, pp. 247–252.

Elshafei, M. (1991). Toward an arabic text-to-speech system, *The Arabian J. Science and Engineering*, Vol. 4B, pp. 565–583.

Fant, G. (1969). Stops in cv syllables, *STL-QPSR*, Vol. 4, pp. 1–25.

Fu, Q. & Wan, E. A. (2003). A novel speech enhancement system based on wavelet denoising, *Center of Spoken Language Understanding, OGI School of Science and Engineering at OHSU .:* .

Gibson, M. (1998). *Dialect Contact in Tunisian Arabic: sociolinguistic and structural aspects*, PhD thesis, University of Reading.

Grossmann, A. & Morlet, J. (1984). Decomposition of hardy functions into square integrable wavelets of consonant shape, *SIAM Journal on Mathematical Analysis* 15(4): 723–736.

Hancq, R. B. H. B. T. D. J. & Leich, H. (2000). *Traitement de la parole*, Presses Polytechniques et Universitaires Romandes. ISBN 2-88074-388-5.

Haton, J. & al. (2006). *Reconnaissance automatique de la parole*, DUNOD.

Ladefoged, P. (1996). *Elements of Acoustic Phonetics*, The University of Chicago Press.

Lee, C. C. Y. W. T. & Ching, P. (1999). Two-dimensional multi-resolution analysis of speech signals and its application to speech recognition, *International Conference on Acoustics, Speech, and Signal Processing, ICASSP99*, Vol. 1, IEEE, pp. 405–408.

Leman, H. & Marque, C. (1998). Un algorithme rapide d'extraction d'arêtes dans le scalogramme et son utilisation dans la recherche de zones stationnaires / a fast ridge extraction algorithm from the scalogram, applied to search of stationary areas, *Traitement du Signal* 15(6): 577–581.

Mallat, S. (1989). A theory for multiresolution signal decomposition : the wavelet representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 11: 674–693.

Mallat, S. (2000). *Une Exploration des Signaux en Ondelettes*, Editions de l'Ecole Polytechnique, Ellipses diffusion.

Mallat, S. (2008). *A wavelet Tour of Signal Processing*, 3rd edition edn, Academic Press.

Manikandan, S. (2006). Speech enhancement based on wavelet denoising, *Academic Open Internet Journal* 17(1311–4360): .

Marcais, W. (1950). *Les parlers arabes, Initiation à la Tunisie*, d. Adrien Maisonneuve, Paris.

Muhammad, A. (1990). *Alaswaat Alaghawaiyah, (in Arabic)*, Daar Alfalah, Jordan.
Peterson, G. & Barney, H. (1952). Control methods used in a study of the words, *Journal of Acoustical Society of America, JASA* 24: 175–184.
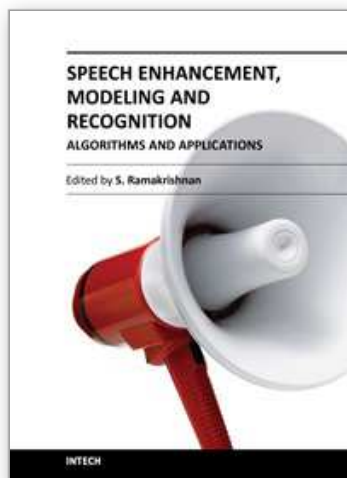
This book on Speech Processing consists of seven chapters written by eminent researchers from Italy, Canada, India, Tunisia, Finland and The Netherlands. The chapters covers important fields in speech processing such as speech enhancement, noise cancellation, multi resolution spectral analysis, voice conversion, speech recognition and emotion recognition from speech. The chapters contain both survey and original research materials in addition to applications. This book will be useful to graduate students, researchers and practicing engineers working in speech processing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Nefissa Annabi-Elkadri, Atef Hamouda and Khaled Bsaies (2012). Multi-Resolution Spectral Analysis of Vowels in Tunisian Context, Speech Enhancement, Modeling and Recognition- Algorithms and Applications, Dr. S Ramakrishnan (Ed.), ISBN: 978-953-51-0291-5, InTech, Available from: http://www.intechopen.com/books/speech-enhancement-modeling-and-recognition-algorithms-and-applications/multiresolution-spectral-analysis-of-vowels-in-tunisian-context-

INTECH
open science | open minds