

## **Pitch Shifting of Arabic Speech Signal by Source Filter Modelling for Prosodic Transformations**

Ykhlef Fayçal

*Electronic Department, University of Saad Dahleb, Algeria  
ykhlef\_faycal@yahoo.fr*

Ykhlef Farid

*Electronic Department, University of Saad Dahleb, Algeria  
f\_ykhlef@yahoo.fr*

Bensebti Mesaoud

*Electronic Department, University of Saad Dahleb, Algeria  
mbensebti@yahoo.fr*

Guerti Mhania

*Electronic Department, National Polytechnic School of Algiers, Algeria  
mguerti@yahoo.fr*

### ***Abstract***

*This paper proposes a pitch shifting technique based on the use of Source Filter Model (SFM) by Linear Prediction. A complete solution is described, included estimation of model parameters, pitch detection, Unvoiced/Voiced (U/V) classification and speech synthesis at several pitch values. An automatic U/V classification is proposed using two methods, the Autocorrelation by thresholding and the Zero Crossing Rate (ZCR) versus Energy computation. The satisfactory performance of this technique is evaluated by listening tests using sentences in Standard Arabic Language. The limit range of pitch shifting for a clear speech synthesizing is found.*

### **1. Introduction**

Make a synthesis of high quality requires an accurate check of vocal quality parameters which depend mainly on the voicing source. The acoustic parameters which carry prosodic information (pitch, duration, energy) must be adjusted and evaluated comparing the perception capacities [1]. An effective method of speech synthesis consists of selecting a class of basic acoustic units, recording them in natural voice and generating utterances by concatenating appropriately modified segments from the inventory of stored units [2]. This method is called concatenative synthesis. Systems based on concatenative speech synthesis are being used in many speech technology applications [3]. For example, personification of Text-To-Speech synthesis systems and preservation of speaker characteristics. One problem of segments concatenation is that it doesn't generalize well to contexts not included in the training process, partly because prosodic variability is very large.

There are techniques that allow us to modify the prosody of a unit to match the target prosody. These prosodic transformations techniques degrade the quality of the synthetic speech, though the benefits are often greater than the distortion introduced by using them because of the added flexibility [4].

The objective of prosodic transformations is to change the amplitude, duration and pitch of a speech segment. Amplitude modification can be easily accomplished by direct multiplication, but duration and

pitch changes are not so straightforward. One category of prosodic transformations techniques process directly on the time domain of the speech signal, like the TDPSOLA algorithm [5]. The largest problem in concatenative synthesis using these techniques occurs because of spectral discontinuities due to different pitch values at the unit boundaries. Other categories process by analyzing the vocal signal in order to separate the glottal information from the vocal tract one [6]. This representation is called the Source Filter Model (SFM). The use of the SFM allows us to modify the source and filter separately and thus maintain more control over the resulting synthesized signal. The prosodic parameters can be controlled using only the glottal information. This kind of techniques will significantly reduce the problem of spectral discontinuities at unit boundaries [4].

This paper describes a procedure of pitch shifting that leads to clear sounding synthesized speech using the Linear Prediction model. At first, we describe the SFM principal and its use in pitch shifting. Then, we explain the procedure of speech modelling by Linear Prediction. Afterwards, we present the major problems in pitch detection that can limit the precision of a pitch shifter and we present also the techniques used in pitch detection and voice classifications. The experiments and results for the analysis and synthesis of Arabic speech signal at several pitch values are presented in details. We complete this paper by a conclusion to our work.

## 2. Source Filter model

The SFM describes the vocal signal as a time domain glottal signal (with a flat spectral envelope) plus a vocal tract filter containing the information about formants. The source is assumed to carry the glottal information and the filter represents the vocal tract.

### 2.1. Pitch shifting using the model

We assume that the glottal excitation signal is approximately a pulse train in the case of Voiced speech (quasi periodicity) and a white noise in the case of Unvoiced one. So, the spectral envelope of the glottal signal is flat. The shape of the voice spectrum depends only on the vocal tract that acts as a filter which frequency response is the spectral envelope of the output signal [6]. Consequently, if we can design an equivalent filter, we will have a model of the vocal tract. If we keep the vocal tract model as it is and alter only the glottal signal, we will not modify the spectral shape of the voice as long as we preserve the flatness of the glottal spectrum. The pitch shifting operation can be done in this case by changing the periodicity of the source signal in the case of Voiced speech (without altering formants which represent the vocal tract information).

### 2.2. The Linear Prediction

Since the term Linear Prediction was first coined by N. Wiener in 1966 [7], the technique has become popularly employed in a wide range of applications. This technique first used for speech analysis and synthesis by Itakura and Saito and Atal and Schroeder in 1968 [7], has produced a very large impact on every aspect of speech research. The importance of Linear Prediction stems from the fact that the speech wave and spectrum characteristics can be efficiently and precisely represented using a very small number of parameters.

Additionally, these parameters are obtained by relatively simple calculations. The Linear Prediction is described as a system identification problem, where the parameters of an AutoRegressive (AR) model are estimated from the signal itself. A stationary signal  $x(n)$  will be known as AR if it obeys to the following model, also known as AR [8]:

$$x(n) + a_1x(n-1) + \dots + a_Mx(n-M) = e(n) \quad (1)$$

Where  $e(n)$  is a centered white Gaussian noise, of variance  $J$ . The coefficients  $a_i$  will be known as predictors and the constant  $M$  as the prediction order. The relation (1) can be interpreted by the  $Z$  transform as follow:

$$X(z) = \frac{1}{A(z)} E(z) \quad (2)$$

With 
$$A(z) = 1 + a_1 z^{-1} + \dots + a_M z^{-M} \quad (3)$$

We describe the signal  $x(n)$  as an output of the filter with Transfer Function (TF)  $1/A(z)$ , i.e. of a recursive filter or all pole system. The input of this system is the white noise  $e(n)$  that is also interpreted as a prediction error. What justifies seeking the optimal coefficients  $a_i$  by minimizing this error or more exactly by minimizing its variance  $J$  [9]:

$$J = E[e(n)^2] = E \left[ \sum_{i=0}^M a_i x(n-i) \sum_{j=0}^M a_j x(n-j) \right] \quad (4)$$

With  $E[.]$  denote the expectation operator. The minimization of this variance (using autocorrelation method) leads us to the normal equations:

$$\begin{bmatrix} R(0) & R(1) & \dots & R(M) \\ R(1) & R(0) & \dots & M \\ \dots & \dots & \dots & R(1) \\ R(M) & M & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_M \end{bmatrix} = \begin{bmatrix} J_{\min} \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (5)$$

Where  $R(i)$  represents the autocorrelation function of the signal  $x_n$  and  $J_{\min}$  the minimized value of the variance error prediction. These equations can be solved using several algorithms. In our case, we have used the Levinson–Durbin algorithm which is suitable for practical implementation of Linear Prediction analysis. The Levinson–Durbin approach finds the solution to the  $m^{\text{th}}$  order predictor from that of the  $(m-1)^{\text{th}}$  order predictor [10]. It is an iterative recursive process where the solution of the zero order predictor is first found, which is then used to find the solution of the first order predictor. This process is repeated one step at a time until the desired order is reached (Table 1). Inputs to the algorithm are the autocorrelation coefficients  $R[l]$ , with the LPCs (Linear Prediction Coding) and RCs (Reflexion Coefficients)  $k_i$  the outputs.

### 2.3. Source filter by Linear Prediction

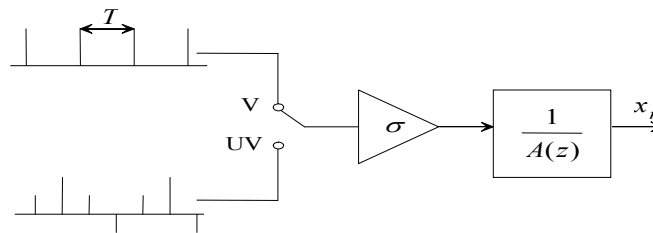
In what precedes, the signal  $e(n)$  was considered as an error prediction which we minimized the variance to calculate the Linear Prediction coefficients  $a_i$ . The first representation shows that we can rebuild the residue  $e(n)$  of the estimation starting from the signal  $x(n)$  using a non-recursive filter represented by the TF  $A(z)$ . Conversely and in the second representation, it will be noted that the residue  $e(n)$  can be considered as an excitation signal that is being used to create the signal  $x(n)$  using a recursive all poles filter  $1/A(z)$ . In the case of a speech, the excitation signal can be periodic (Voiced sounds) or random (Unvoiced sounds) [1]. The model generally adopted to create sounds artificially is crude compared to the complexity of the phonate system but it is completely satisfactory for our need (Figure 1).

**TABLE 1: LEVINSON–DURBIN ALGORITHM [10]**

<p>- <b>Initialisation</b> : <math>l = 0</math> , set <math>J_0 = R[0]</math></p>
<p>- <b>Recursion</b> : for <math>l = 1, 2, \dots, M</math></p> <p><u>Step 1.</u> Compute <math>l^{th}</math> RC</p> $k_l = \frac{1}{J_{l-1}} \left( R[l] + \sum_{i=1}^{l-1} a_i^{(l-1)} R[l-i] \right)$ <p><u>Step 2.</u> Calculate LPCs for the <math>l^{th}</math> order predictor:</p> $a_l^{(l)} = -k_l,$ $a_i^{(l)} = a_i^{(l-1)} - k_l a_{l-i}^{(l-1)}; \quad i = 1, 2, \dots, l-1.$ <p>Stop if <math>l = M</math>.</p> <p><u>Step 3.</u> Compute the minimum mean-squared prediction error associated with the <math>l^{th}</math>-order solution</p> $J_l = J_{l-1} (1 - k_l^2).$ <p>Set <math>l \leftarrow l + 1</math> ; return to Step 1.</p>
<p>The final LPCs are :</p> $a_i = a_i^{(M)}; \quad i = 1, 2, \dots, M.$

This model includes:

- a periodic generator of impulses units;
- a random numbers generator with null average value and a unit variance;
- a switch being used to choose the Voiced or Unvoiced sounds;
- a factor  $\sigma$  called gain model, it is selected equal to the minimized value of the variance error prediction;
- an all poles filter.



**Figure 1: Vocal tract model**

**2.4. Pitch shifting using Linear Prediction**

The pitch shifting of the segments stored in LPCs form is a straightforward operation. The period  $T_0$  of the vocal signal being a parameter of the model, it is enough to impose directly the synthesis value  $T$  , what causes to modify the period of the impulses periodic train used in excitation of the Voiced zones [1]. If we want that this operation doesn't affect the power of the signal, it is necessary to modify simultaneously the value of the gain factor  $\sigma$  starting from its initial value  $\sigma_0$  (corresponding to  $T_0$  ).

$$\sigma = \sigma_0 \sqrt{T_0/T} \quad (6)$$

### 3. Pitch detection

The pitch estimation plays a very important role in speech compression, coding, recognition and synthesis, as well as in voice transformation. A good estimation of the pitch period is crucial to improve the performance of speech analysis and synthesis systems [11].

Accurate and reliable measurement of the pitch period of a speech signal from the acoustic pressure waveform alone is often exceedingly difficult for several reasons. One reason is that the glottal excitation waveform is not a perfect train of periodic pulses [12]. Measuring the period of a speech waveform, which varies both in period and in the detailed structure of the waveform within a period, can be quite difficult [13]. A second difficulty in measuring pitch period is the interaction between the vocal tract and the glottal excitation [12]. The formants of the vocal tract can alter significantly the structure of the glottal waveform so that the actual pitch period is difficult to detect [14]. A third problem in reliably measuring pitch is the inherent difficulty in defining the exact beginning and end of each pitch period during voiced speech segments. The choice of exact beginning and end of each pitch period is often quite arbitrary [13]. A fourth difficulty in pitch detection is distinguishing between unvoiced speech and low-level voiced speech. Transitions between unvoiced speech segments and low-level voiced speech segments are very subtle and extremely hard to pinpoint [13]. The fifth problem is that pitch period depends upon sex (female or male), age, emotional and health state of the speaker. It can also be influenced by the accent. In general, the pitch frequency varies from 70 to 250 Hz for men, 150 to 400 Hz for women and from 200 to 600 Hz for children which is a large bandwidth of variation.

The result is that the periodicity of the speech signal is much hard to detect, no one algorithm has been developed so far performs perfectly for all different speakers, applications and environmental conditions [12]. This detection is the main important step in pitch shifting using the SFM model and all remaining result depend on it.

Many algorithms had been reported in the literature for pitch detection which can be classified in tree main categories: Time domain, Frequency domain and hybrid method (Combining both, time and frequency).

- Time domain: such as traditional auto-correlation, Average Magnitude Difference Function (AMDF) [15], and DATA REDUCTION method (DARD) [16].
- Frequency domain methods: such as Cepstral Pitch detector CEP [17].
- Hybrid method: which incorporates features of both Time-domain and Frequency-domain approaches such as the Simplified Inverse Filtering Technique (SIFT) [18] and Spectral equalization LPC method using Newton's transformation [13].

Each algorithm present advantages and inconveniences which are due to several problems quoted above. The basic assumption that is made in all these cases is that if a quasiperiodic signal has been suitably processed to minimize the effects of the formant structure, then simple time-domain measurements will provide good estimates of the pitch period [13].

In our work, since we use the Autocorrelation function to compute LPCs parameters, we have preferred to estimate the pitch from the residual  $e(n)$  of the estimation using the traditional autocorrelation technique. This approach enables us to avoid the vocal tract interaction where the measurement (the residual excitation) contains only information on the excitation source (pitch). In order to separate the components of the speech signal, we need to classify each speech segment into two categories: Voiced or Unvoiced for both pitch detection and the choice of the LPC model generators used in excitation. In our work, the U/V decision is made using two methods. The first one evaluates the measurement of the Zero Crossing Rate (ZCR) versus Energy of all stationary speech

segments. The second one uses peaks decision of the Autocorrelation function for each stationary speech segment. The decision rule for ZCR versus Energy classification is based on the following criteria:

- If the Energy (Eq.7) is smaller than an Energy threshold then the segment is Unvoiced otherwise we compute the ZCR (Eq. 8).

$$E = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2 \quad (7)$$

$$ZCR_i = ZCR_{i-1} + \frac{1}{2} |(\text{sign}(x(i)) - \text{sign}(x(i-1)))| \quad (8)$$

Where x denotes N samples of a speech signal.

The function sign(x) returns 1 if the element is greater than zero, 0 if it equals zero and -1 if it is less than zero. The initial value of ZCR is null.

- If the ZCR is larger than a ZCR threshold then the segment is Unvoiced, otherwise the segment is decided to be Voiced.

The decision rule for the second method is based on the comparison between the First Peak Value (FPV) of the autocorrelation function calculated for each stationary segment in the range of the pitch period existence (depending on the sex of the speaker) and the initial one calculated at time zero IPV (Initial Peak Value) of the same function [19]:

- If the FPV is greater or equal than a Peak Value Threshold (PVT) than the segment is decided to be Voiced, otherwise the segment is Unvoiced.

The PVT is generally taken equal to 50 % of the IPV, but in our case using representative phonetically equilibrated sentences in Standard Arabic Language another PVT is given in the next section.

#### 4. Experiments and results

In this section, we will establish a speech signal model of analysis and synthesis for pitch shifting based on the following steps:

- acquisition of the signal;
- emphasis of the speech signal using a transmittance filter  $(1 - 0.95z^{-1})$ ;
- blocks constitution of N samples with shifts of L samples;
- U/V classification and pitch detection of Voiced segments;
- estimation of the model parameters for each segment;
- synthesis by Add and Overlap using Hamming window and application of a de-emphasis filtering.

Generally, we estimate that the TF of the model must comprise a pair of poles by kHz of bandwidth; the glottal excitation and the radiation of the lips require together 3 or 4 poles.

Then, if the sampling frequency  $F_e$  is equal to 10 kHz, we choose  $M$  equal to 13 or 14 [9]. In our case,  $F_e$  is chosen equal to 16 kHz and the order  $M$  to 19 (16 for the TF and 3 for the glottal excitation and radiations with the lips). The duration of the signal segments depend on the selected method and conditions under which it is applied. The practice shows that the window must encroach over several fundamental periods for the Voiced sounds.

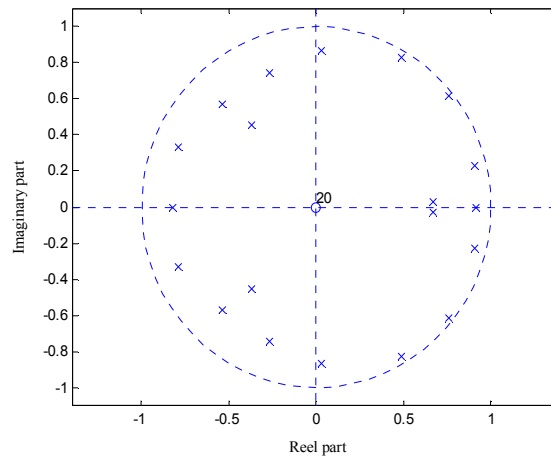
We usually use windows of 30 ms shifted by 20 ms hence at this duration the speech signal is considered as a stationary one. For  $F_e$  equal to 16 kHz, the analysis segments are setup to 480 samples shifted by 320. The same choice is remained for Unvoiced segments.

The following phonetically representative Arabic sentences, stored on a disk, were used as samples to produce synthesized speech with different pitch values. These sentences cover Occlusive (Voiced/Unvoiced), Fricative (Voiced/Unvoiced), Nasal, Liquid, Vibrant, African phoneme, Semi Vowel and Vowel of Standard Arabic sounds:

1. "The seventh course"
2. "The Arabic press"
3. "And the sound shushed again"
4. "The geographic location"
5. "The apparition of Islam and its propagation"
6. "The paradise Family"

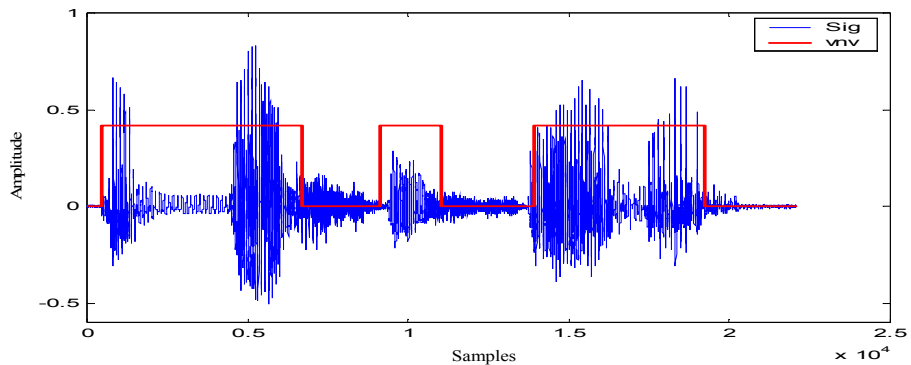
الدرس السابع  
الصحافة العربية  
وسكت الصوت مرة أخرى  
الموقع الجغرافي  
ظهور الإسلام و انتشاره  
أهل الجنة

These sentences were spoken by adult male speaker, in Arabic language and sampled at 16 kHz. The LPCs Coefficients represent a  $z^{-1}$  polynomial which is not other than the TF denominator of the vocal tract. As this last is essentially stable, the poles must be inside the unit circle (Figure 2).



**Figure 2: TF Poles of the phoneme [a] over 20 ms with AR order M=20 from the context (course "الدرس") of the first sentence**

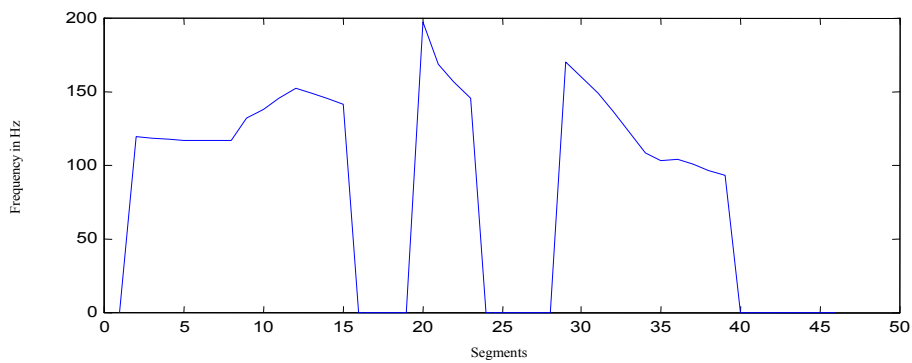
As we have explained in section 3, our pitch detection method is based on the use of the autocorrelation function from the residual  $e(n)$  of the estimation. The noise effects problems can be attenuated by using a low pass filter. We have used an 8<sup>th</sup> order Butterworth low pass filter with a cut-off frequency of approximately 800 Hz. The Energy and the ZCR thresholds used in the U/V classification are given for each input signal starting from practical tests which give better results. In order to automate the system, we have proposed a technique to determine these thresholds. We start first by establishing a row consisted of various calculated Energies of all input segments by an ascending order. Then, among these various successions of Energy, we determine the one which represents the 25<sup>th</sup> percent of the row indices (positions of the Energy row), which indicates finally the desired threshold Energy. The ZCR threshold is calculated in the same way by determining in this time the rate which represents the 75<sup>th</sup> percent of the row indices (positions of the ZCR row) (Figure 3).



**Figure 3: U/V classification of the sentence -the seventh course- “الدرس السابع”**

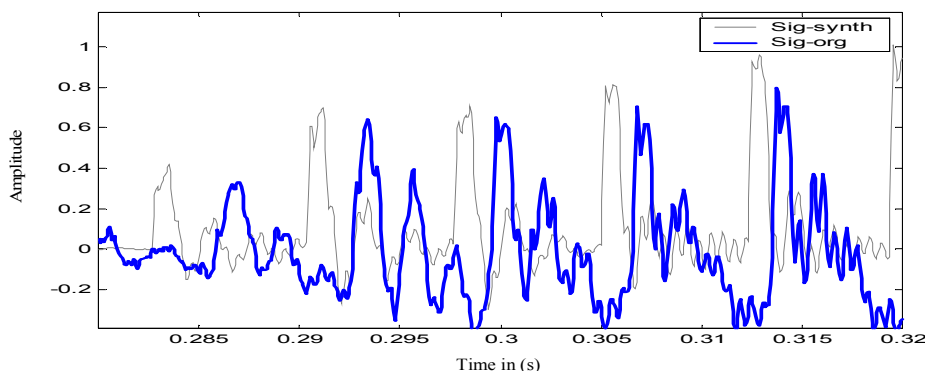
The decision rule for U/V thresholds is based on the principle that about 75% of Standard Arabic speech sounds are Voiced and remain 25% Unvoiced sounds [1].

The same result can also be obtained using the peaks decision of the Autocorrelation function by a choice of a PVT equals to 34% of a given IPV for each frame and for all sentences. The pitch evaluation of the input signal (the first sentence) according to the analysis blocks is represented in the Figure 4.



**Figure 4: Pitch evaluation of the sentence -the seventh course- “الدرس السابع”**

The synthesized phoneme [a] and the original one taken from the context – course - “الدرس” at a duration of 10 ms are represented in the Figure 5, where sig-synth denote the synthesized signal and sig-org the original one. The delay observed is due to the filtering operations. For obtaining a synthetic signal of different pitch values, we have to change simultaneously the periodic generator of impulses units used as a source excitation of Voiced sounds and the gain factor given by Eq. 6.



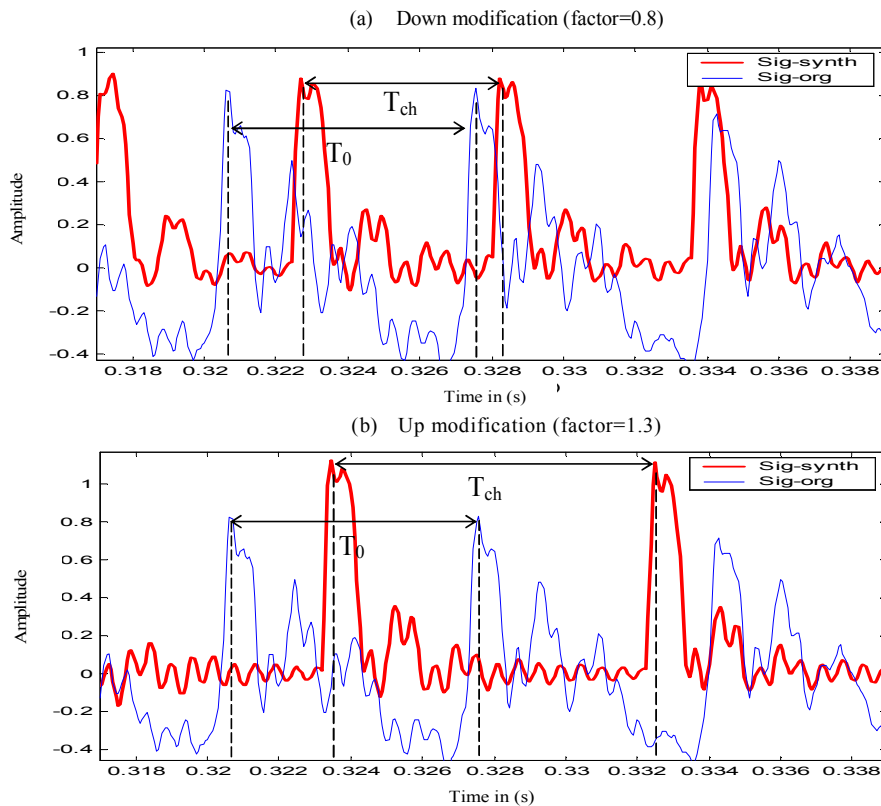
**Figure 5: Temporal representation of the original and synthesized phoneme [a]**

The Figure 6 shows the original phoneme [a] taken from the context – course - “الدرس” and the shifted version synthesized using modifications factors of 0.8 and 1.3. The legends sig-org and sig-synth denote respectively the original signal and shifted one. Also, the legends  $T_0$  and  $T_{ch}$  denote respectively the original and the shifted pitch periods using previously modifications factors. The practical tests show that using the SFM by Linear Prediction, the synthesized signal suffer from amplitude distortion even though we have used the gain corrector. Furthermore, the natural of the synthesized signal is not well reproduced essentially for nasals and transistorises sounds. These problems are basically due to the AR model and sources errors and also the accuracy of the pitch detection and classification algorithms.

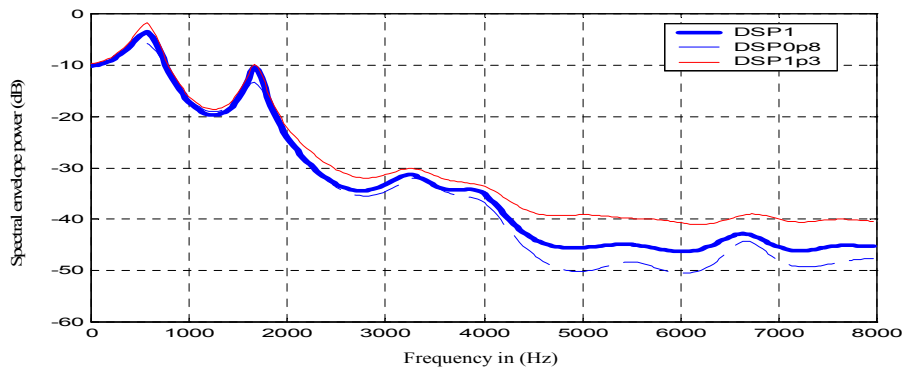
In fact, there are glottal excitations models more sophisticated such as those of Rosenberg or Liljencrants-Fant that approximate more the shape of the excitation signal [17]. The U/V classification used in this work does not consider mixed sounds (like Voiced Fricatives). The use of a mixed excitation returns to the synthesized signal a certain roundness which has been missed in the U/V binary model classification. It is very difficult to improve the quality or the natural of synthesized signal by increasing the sampling frequency or the model order. The advantage of this technique is to change the pitch of the speech signal without altering the spectral envelope which preserves the voice timbre as well as possible (Figure 7). The legends DSP0p8 and DSP1p3 denote respectively the spectral envelope power (in dB) of the synthesized signal using 0.8 and 1.3 factors modifications and DSP1 the spectral envelope power of the synthesized signal without changing the pitch value. Furthermore, the spectral discontinuities due to different pitch values at the unit boundaries are well diminished by the SFM.

The thresholds of the modifications factors reached by any pitch shifting technique are theoretically included between two ends, one maximal and other minimal. To find these limits, we have used the sentences given at the top, spoken by adult male speaker, in Standard Arabic language to evaluate the capability of this technique.

These data base sounds cover all the categories of Standard Arabic sounds After listening tests, we have noticed that ones we are out of a minimum value of 0.4 and a maximum one of 2 (modifications factors) the vocal signal obtained by this technique is becoming increasingly robot-like from where a loss of the naturalness is appearing.



**Figure 6: Pitch modification of the phoneme [a]**



**Figure 7: Spectral envelopes of the interpolated and the synthetic phoneme [a] on a 30 ms interval**

## 5. Conclusion

The performance of the Source Filter Modelling by Linear Prediction designed for pitch shifting of speech signal was investigated by synthesizing sentences in Standard Arabic Language. We have proposed in this work a complete solution included the algorithms used for pitch detection, U/V classification, AR parameters estimations and synthesizing using Add and Overlap method. The pitch shifting operation using this technique is well done by preserving the spectral envelope and the timbre of the speech signal. The synthesized sentences suffer from synthesizing errors due essentially to AR

modelling, pitch detection, U/V classifications and the excitations sources. The optimization of all criterions is a very difficult operation. This technique can be explored in real time applications from the fact that using the Autocorrelation function for voice classification, pitch detection and voice modelling can be done frame by frame without explicit information of the last frames. We can conclude that pitch shifting using SFM by Linear Prediction technique is an elementary operation. The complexity of this technique is located in the analysis approach. The use of combined pitch shifting method between Source Filter Model and time domain technique will improve the natural speech synthesizing. Our future works will focus on increasing the naturalness of the modified Arabic sounds and to explore it for the modification of the other prosodic parameters.

## 6. References

- [1] F. Ykhlef, Pitch modification for Text To Speech synthesis of Arabic language (French Lang.), Master These, Saad Dahleb University, Blida, Algeria, Sep. 2005.
- [2] R. Ansari, D. Kahn, M. Macchi, "Pitch modification of speech using a low-sensitivity inverse filter approach", IEEE Signal Processing Letters, vol. 5, no. 3, Mar. 1998, pp. 60-62.
- [3] A. Syrdal, R. Bennett, and S. Greenspan, Applied speech technology, Boca Raton, FL: CRC, 1995.
- [4] Xuedong Huang, Alex Acero and Hsiao-Wuen Hon, Spoken language processing: A Guide to Theory, Algorithm and System Development, Prentice-Hall Inc., 2001.
- [5] X. Sun, "Voice quality conversion in TD-PSOLA speech synthesis", IEEE Proc. of the Acoustics, Speech, and Signal Processing, vol. 2, 2000, pp. 953-956.
- [6] Patrick Bastien, Pitch shifting & voice transformation. VoicePro Whitepapers, tc- helicon vocal technologies, reachable from [www.tc-helicon.com/](http://www.tc-helicon.com/).
- [7] Sadaoki Furui, Digital speech processing, synthesis and recognition, Second Edition (Revised and Expanded), Marcel Dekker, New York, 2001.
- [8] Calliope, Speech and it automatic processing (French Lang.), Ed. J.P. Tubach, Masson, Paris, 1989.
- [9] R. Boite, H. Boulard, T. Dutoit, J. Hancq and H. Leich, Speech processing (French Lang.), Polytechnic Press and Romandes University, Electrical Collection, Lausanne, 2000.
- [10] W. C. Chu, Speech coding algorithms: Foundation and evolution of standardized coders, Hoboken, John Wiley & sons inc., New Jersey, 2003.
- [11] Li Hui, Bei-Qian Dai, Lu Wie, "Pitch detection algorithm based on AMDF and ACF", IEEE Proc. of the Acoustics, Speech, and Signal Processing, vol. 1, 2006, pp. I-I.
- [12] Hui Ding; Bo Qian; Yanping Li; Zhenmin Tang, "A method combining LPC-based cepstrum and harmonic product spectrum for pitch detection", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Dec. 2006, California.
- [13] Rabiner, L.Cheng, M, Rosenberg, A, McGonegal, C, "A comparative performance study of several pitch detection algorithms", IEEE Trans. Acoustics, Speech, Signal Processing, vol. 24, issue 5, Oct. 1976, pp.399-418.
- [14] Xin Xu, Miyanaga, Y, "A robust pitch detection in noisy speech with band-pass filtering on modulation spectra", IEEE International Symposium on Communications and Information Technology, Oct. 2005, China.
- [15] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor", IEEE Trans. Acoustics, Speech, Signal Processing, vol. ASSP-22, Oct. 1974, pp. 353-362.
- [16] N. J. Miller, "Pitch detection by data reduction", IEEE Trans. Acoustics, Speech, Signal Processing (Special Issue on IEEE Symposium on Speech Recognition), vol. ASSP-23, Feb. 1975, pp. 72-79.
- [17] A. M. Noll, "Cepstrum pitch determination", J. Acoust. SOC. Amer., vol. 41, Feb. 1967, pp. 293-309.
- [18] J. D. Markel, "The SIFT algorithm for fundamental frequency estimation," IEEE Trans. Audio Electroacoust., vol. AU-20, Dec. 1972, pp. 367-377.
- [19] J. Fernandez, Acquisition, compression and restitution algorithms of variable speed speech signal, study and implementations (French Lang.). Doctorate these. National High School of Electronic and its applications, France, April 1995.

## AUTHORS



Mr. Ykhlef Fayçal received the B. Eng. degree in electronic engineering from Saad Dahleb university of Algeria in 2002 and his M. Sc. degree in Image and speech from the same university in 2005. In 2006 he worked as an assistant teacher at Saad Dahled university where he still continuing his research at electronic departement of Saad Dahled university. From 2007 till now he is an researcher at the Center of Advanced Technologies Development in Algiers, Algeria. His research interests are in image processing, speech analysis, synthesis, enhancement and signal processing.



Mr. Farid Ykhlef received the B.S. and M.S. degrees in electronic engineering from University of Blida, Algeria, in 1998 and 2002 respectively. In 2003, he worked as a researcher at Scientific and technical research center in welding and control, Algeria. He joined the University of Blida, Algeria, as an Assistant Professor in the department of Electronic, since 2004. His current research interests include digital filter design, speech processing, adaptive filtering, acoustic echo cancellation, systems identification and fast algorithms.