

A NEW PITCH DETECTOR BASED ON PRESELECTED INFORMATION
FROM THE LPC ERROR SIGNAL.

S. G. Bakamidis*, G. Carayannis*, N. Skiadas*.

ABSTRACT

Pitch extraction is always a very interesting problem. Many algorithms exist but none of them is always working. The technique studied here is very accurate. The multipulse sequence is proposed as the most efficient vehicle of pitch information. A computer program named MPDE (Multipulse Driven Pitch Extractor) is built which performs extremely well. Its implementation details are discussed next.

I. INTRODUCTION

Many speech interface applications require a reliable pitch extraction algorithm. Speech synthesis, vocoding, speaker identification, etc., are ultimately related to the knowledge of the vocal tract excitation rhythm. Furthermore, theoretical studies on prosody need a high quality pitch detector. Many different pitch extraction algorithms exist and may be categorized in the following classes :

- heuristic methods and vision simulating methods
- cepstrum based methods (ref 1)
- linear prediction methods (ref 1)

The human eye is considered to be as one of the best pitch detectors, so many algorithms have been developed in trying to simulate the eye pitch detection ability. The main disadvantage in these methods is the huge computational complexity and the lack of real knowledge of the eye detection mechanism.

The cepstrum based methods try to extract the pitch from the cepstrum signal which is the Discrete Fourier Transform (DFT) of the logarithm of the DFT of the speech signal computed after adequate windowing. Cepstrum based methods are not very reliable.

In linear prediction error based methods (ref 1) the speech signal is modeled and the inverse filtered signal, that is the error residual is then used for pitch extraction. The principle is that the autocorrelation function or some similar function will be more aware of the pitch information than the signal itself as the vocal tract influence is minimized in the prediction error avoiding thus the masking effects created by the strongly convolved signals. However it must be noticed that none of the existing algorithms which fall in the above mentioned classes is totally accurate and that the degree of accuracy depends largely on the complexity of the algorithm and the associated error correcting logic.

Linear prediction error algorithms (ref 4,5) suffer from a number of drawbacks some of them are :

- speaker dependence
- frequent classification of voiced speech segments as unvoiced.

In the class of linear prediction error based methods, we introduce the principle of pitch extraction from a sequence of selected samples of the error signal $e(n)$. Many possibilities exist for the selection of a new sequence $e'(n)$ as a basis for pitch extraction. One error signal which is studied in more detail here is the multipulse sequence.

* National Technical University of Athens, Division of Computer Science,
Zografou, 15773 Athens, Greece.

There are several algorithms for the generation of the multipulse sequence, the most suitable of them are based on an optimization criterion. The autocorrelation of $e'(n)$ will give the periodicity of the signal if any. The above described principle leads to a number of significant improvements in pitch detection.

- very reliable pitch extraction
- speaker independent performance
- excellent pitch tracking
- it can be combined with dynamic programming and post processing techniques, but generally no postprocessing is needed.

Section II describes the selection principles conducing to the possible sequences $e'(n)$. Section III is referred to implementation issues and discussions and finally in section IV conclusions are drawn.

II. SELECTION PRINCIPLES

A simple method to create $e'(n)$ is to select a number M of samples ($M \ll N$) which greatly contribute to the total energy of $e(n)$ $0 \leq n \leq N-1$ and discard the remaining ones. A more sophisticated way described below is the selection of positions and amplitudes of the nonzero samples of $e'(n)$ via an optimization criterion which is the so-called multipulse excitation of speech. Let us write $e'(n)$ as

$$e'(n) = \sum_{k=1}^M g_k \delta(n-n_k) \quad 0 \leq n \leq N-1 \quad (1)$$

where g_k is the amplitude of the pulse k positioned at n_k and $\delta(n)$ is the unit pulse. The synthetic signal $\hat{s}(n)$ produced when the LPC filter is driven by $e'(n)$ is given by

$$\hat{s}(n) = \sum_{i=1}^p a_i s(n-i) + e'(n) = \sum_{i=1}^p a_i s(n-i) + \sum_{k=1}^M g_k \delta(n-n_k) \quad (2)$$

where a_i $1 \leq i \leq p$ are the LPC coefficients. the minimization of the total error

$$E = \sum_{n=0}^{N-1} |s(n) - \hat{s}(n)|^\lambda \quad (3)$$

with respect to g_k , n_k $1 \leq k \leq M$ leads to a set of nonlinear equations with unknowns the g_k and n_k $1 \leq k \leq M$. λ can be any integer or real number but in practice it is taken to be 1 or 2, so E is the total absolute or the total mean squared error respectively. Minimization of (3) with respect to the pulse locations n_k leads to a set of nonlinear equations which do not have a closed form solution. Many algorithms have been proposed for solving the above problem (ref 3,4) the first one is the maximum crosscorrelation method introduced by Atal and Remde (ref 3) in which pulse amplitudes and locations are obtained in successive stages one pulse at a time.

An examination of $e'(n)$ reveals the periodic structure of the excitation for voiced sounds and this happens because the major pulses are placed at the onset of pitch periods. Thus for voiced sounds the autocorrelation of $e'(n)$ will show a strong peak at the pitch period.

III. IMPLEMENTATION ISSUES AND DISCUSSION

It is well known that the vocal tract excitation for voiced sounds is quasi-periodic, that is, its period is a function of time. Observing voiced speech segments or their residuals after LPC analysis or the corresponding excitation signals we may see that the differences between

consecutive pitch period onsets is not the same in all the segments and this is due to the quasi-periodic nature of speech. The quasi-periodicity is one important reason for failures in pitch detection block algorithms, because the amplitude of the peak in the autocorrelation function or the null in the AMDF function depends on the fundamental frequency shift from one period to the next.

So, it is very common in practice to get from a quasi periodic speech segment, containing several pitch periods, a weak peak (null), which may not exceed a specified threshold resulting in wrong classification. It

is also possible in the above example to get multiple adjacent peaks (nulls). None of them exceed a specified threshold leading to a pitch estimation which does not correspond to the average magnitude of the pitch period in the analysed frame.

The same apply to the autocorrelation function of the multipulse excitation signal $e'(n)$ but in that case $e'(n)$ can be processed properly in order to get a high peak in the average pitch period. This process consists of convolving $e'(n)$ by an even function $f(n)$ with $2L+1$ non zero terms :

$$f(n) = 0.25 \sum_{i=-L}^L (|i|+1) [\delta(n-i) + \delta(n+i)] + \delta(n) \quad i \neq 0 \quad (4)$$

Typical value for L is 2.

The MDPE algorithm described above was implemented in a digital computer and tested extensively with male and female voices from different speakers. The performance evaluation was based on its comparison with both the AMDF algorithm and eye detection. In the simulation the sampling frequency was 8 KHz and the input speech signal bandlimited to 3.2 KHz. The speech signal was analysed with a 10th order linear prediction model and the multipulse excitation sequence of 18 pulses per frame was calculated. Figure 1 shows the pitch contours obtained from AMDF (fig. 1b) and MDPE (fig 1c) algorithms for a voiced sentence. (fig. 1a)

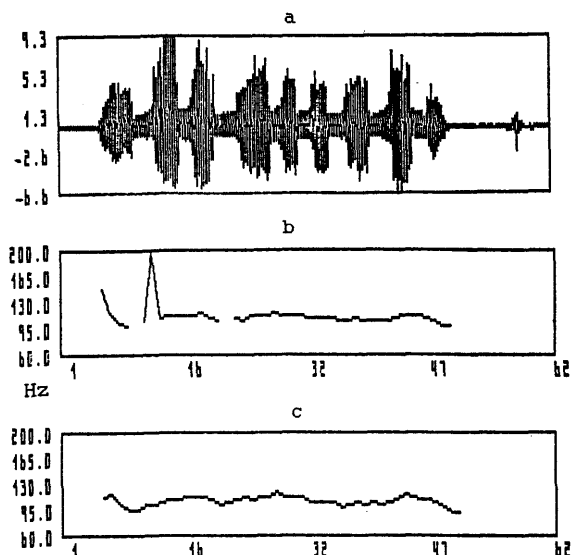


Figure 1 a) Speech waveform
b) Pitch contour from AMDF
c) Pitch contour from MDPE

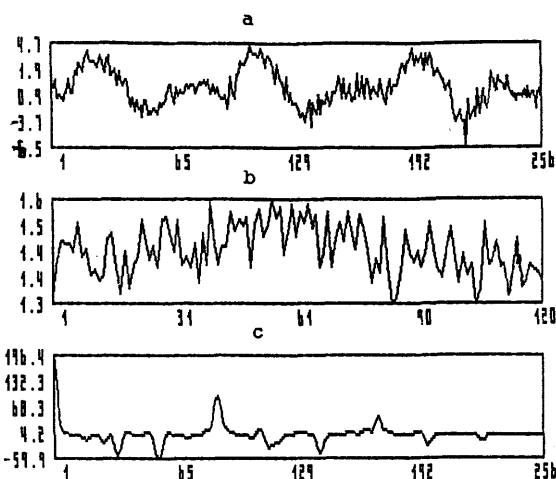


Figure 2 a) Speech segment
b) AMDF of the LPC residual
c) MDPE autocorrelation

Figures 2 through 4 illustrate voiced speech segments (fig 2a-4a) and the corresponding AMDF (fig 2b-4b) and MDPE (fig 2c-4c) signals. In figure 2a a periodic like speech segment is depicted while in figures 2b- c the corresponding AMDF and MDPE signals are shown. It may be observed

that the AMDF null is not very clear while the MDPE autocorrelation shows a clear peak located at the average pitch value. Figure 3a depicts a transition from /m/ to /a/. Again, the MDPE method shows clear peak in the autocorrelation while the AMDF shows no distinct null. Finally in

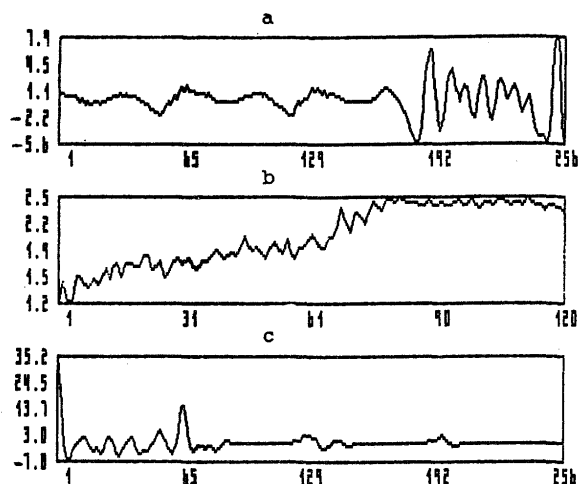


Figure 3 a) Speech segment
b) AMDF of the LPC residual
c) MDPE autocorrelation

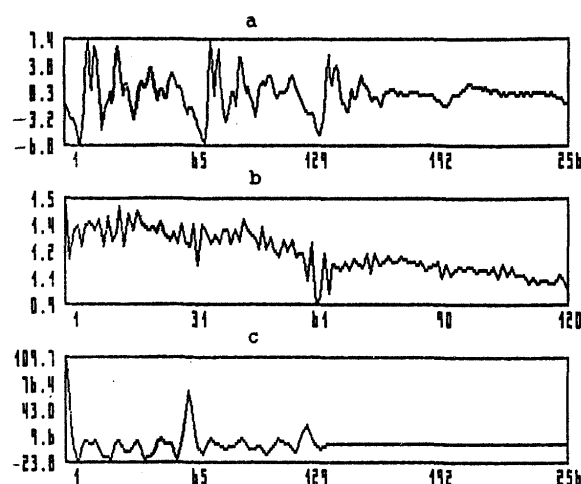


Figure 4 a) Speech segment
b) AMDF of the LPC residual
c) MDPE autocorrelation

figure 4 the new method indicates the signal periodicity clearer than that of the AMDF.

IV. CONCLUSION

The pitch extraction system discussed in this paper is almost error free. Excellent performance is achieved without using complicated correction logic. The system introduced here demonstrates that pulse positions are strongly related to the pitch information in multipulse coding techniques. The MDPE algorithm can be a valuable tool for pitch extraction either alone or in a multiexpert pitch measurement environment. Further improvements are presented in (ref 6).

REFERENCES

1. Rabiner L. R. and Schafer R. W., Digital processing of speech signals (Prentice Hall, Englewood Cliffs, New Jersey 1978)
2. B. S. Atal and J. R. Remde A new model of LPC excitation for producing natural sounding speech at low bit rates. IEEE Proc. Int. Conf. on ASSP Paris, France 1982 pp. 614-617
3. A. Parker, S. T. Alexander and H. J. Trucel, Low bit rate speech enhancement using a new method of multipulse impulse excitation, IEEE Pproc. Int. Conf. on ASSP San Diego CA, USA, 1984 pp. 1.5.1-1.5.4
4. C. K. Un and S. C. Yang, A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF, IEEE Trans. Acoust. Speech and Signal Processing vol. ASSP-25 PP 565-571 December 1977.
5. J. D. Markel, The SIFT algorithm for fundamental frequency estimation IEEE Trans. Electroacoustics vol. AU-20 PP 367-377 Dec. 1972
6. S. G. Bakamidis and G. Carayannis, A new multipulse driven pitch extraction system. Submitted for publication.