END POINT ANALYSIS FOR DETECTING PITCH BOUNDARIES OF SPOKEN INDIAN DEVNAGARI LANGUAGE NUMERICAL USING AUTOCORRELATION TECHNIQUE

Pramod B. Patil Dr. V. T. Ingole
PRM Institute of Technology and Research, Badnera, Amravati, India

Abstract:

Pitch boundaries detection of the speech is an important task due to the presence of background noise. By detecting pitch for extracting linear predictive codes, allows the generation of high quality speech. Primarily start and end points of speech is detected and thereafter pitch boundaries are recognized. This paper emphasized on accurate end point analysis for detection of pitch boundaries for spoken Indian Devnagari Numerical words by autocorrelation technique.

Keywords: Pitch Boundaries, Autocorrelation

I. Introduction:

Generating speech a source that is passed through a filter with the vocal tract response to produce speech [1][2]. The simplest implementation of this is known as the Linear Predictive Coding (LPC) synthesizer. At every frame, the voiced speech is analyzed to compute the pitch value, filter coefficients, the energy of the excitation and a voicing decision. In order to generate high quality speech with synthesis based on LPC, a method for detecting the pitch boundaries is important, since detecting start and end point is very critical. This paper discusses the correct detection of start and end point of the spoken Indian Devnagari Language Numerical in the presence of background noise and thereby detects the accurate pitch boundaries by autocorrelation technique [3][4].

II. Pitch detection:

The pitch period is estimated by autocorrelation technique by detecting the peak amplitude in the function. The second largest peak relative to the true delay of the correlation is used to determine the pitch period. The peak in the autocorrelation function tends to fall off linearly starting from the first peak [5][6].

III. Methodology:

The spoken data for Indian Devnagari Language Numericals are recorded at 8KHz sampling frequency with 8 bits representing each sample. Figure 1 shows block-processing model for detecting the pitch boundaries.

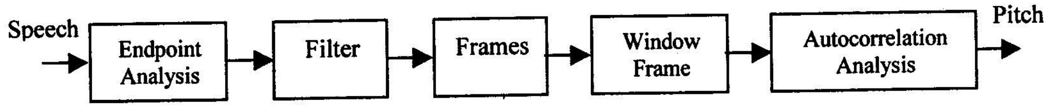

Fig 1. Block Processing Model for Detecting Pitch Boundaries

The data is processed in frame of 200 samples and a vector of pitch is measured. The vector is obtained by pre-emphasizing speech data to flatten it using first order digital filter with coefficient a= 0.97.The signal is processed in the steps of 100 samples to provide smoothing between vectors of pitch. For speech analysis, the hamming window and autocorrelation technique is with different order is used. In the start & End point analysis of speech, each window samples are classified as voiced speech, unvoiced speech or silence, based entirely on measurements made on the signal during the prescribed interval [7]-[11].

IV. Result:

The following results are obtained with the samples of spoken Indian Devnagari Language Numerical "ek". The autocorrelation system order selected is equal to 10, 11 for analyzing the better performance. Fig 2, 3 shows pitch boundaries detected using the start & end point analysis and autocorrelation technique for the spoken Indian Devnagari Language Numerical "ek". Table 1 shows the start and end frames of spoken word "ek".

Spoken Word= ek	Starting Frame	Ending Frame
Order=10	42	80
Order=11	38	78

Table 1. Showing the Start & End frames of spoken word "ek"

V. Conclusions:

Autocorrelation method with start and end point analysis has an advantage of less computation since only a window length is used as signal component. Even though the speech waveform is not always periodic, the pitch boundaries are detected reliably.

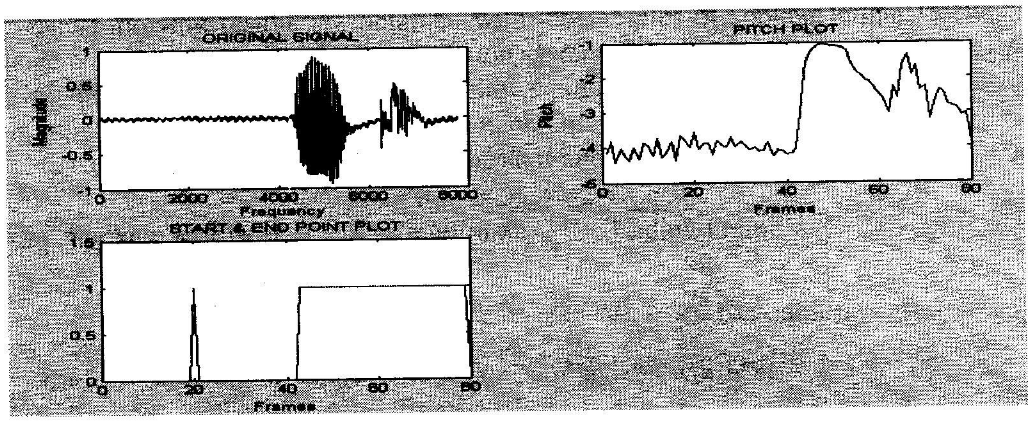

Fig 2. Pitch boundaries for the spoken Indian Devnagari Language Numerical " ek" with system order 10.

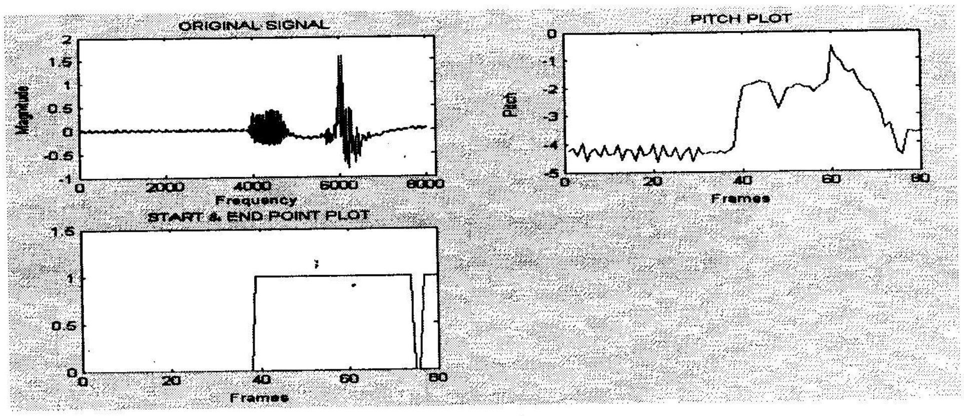

Fig 3. Pitch boundaries for the spoken Indian Devnagari Language Numerical " ek" with system order 11.

References:

1. Samad S, Hussain A, Fah L K, " Pitch Detection of Speech Signals using the Cross Correlation Technique", Proceedings of IEEE on Speech, Audio and Signal Processing, pp 283 - 286,2000.
2. M.J.Ross, H.L.Shaffer, A. Cohen, R. Freudberg, and H.J. Manley, "Average Magnitude Difference Function Pitch Extractor", IEEE Trans. Acoustic, Speech, and Signal Processing, pp. 353-362 Oct. 1974.
3. D. Takin, "A Robust Algorithm for Pitch Tracking (RAPT)", Speech Coding and Synthesis, Netherlands: Elsevier Science. 1995.
4. LA. Atkinson, M. Kondoz and B.G. Evans, "Time Envelop Vocoder, A New LP Based Coding Strategy for Use of Bit-Rate 2.4kb/s and Below", IEEE Journal on Selected Areas on Communications, Vol. 13, No. 2, Feb. 1995.
5. A. M. Kondoz, Digital Speech: Coding for Low Bit Rate Communications Systems, Wiley, England. 1995.
6. B.Gold and L. Rabiner, "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain", Journal of Acoustics. Society, America, Vol. 46, pp. 442-448, Aug. 1969.
7. Rabiner L R, Levinson S E., "Isolated and Connected Word Recognition – Theory and Selected Applications ", IEEE Transaction on Communications, Vol 29, No 5, 1981
8. Rabiner L R, Sambur M R, " Speaker Independent Recognition Of Connected Digits", Bell System Technical Journal Vol 54, pp 202 –205, 1972.
9. Rabiner L R, Schafer R W, Digital Processing of Speech Signals, Englewood Cliffs, NJ Prentice Hall, 1978.
10. Rabiner L R, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of IEEE, Vol 77, No2, pp 257-285, 1989.
11. Rabiner L R, Sambur M R, "An Algorithm for Determining the Endpoints of Isolated Utterances" Bell System Technical Journal, Vol 54, pp 297 –315, 1975.