Short Time Cepstrum Analysis Method for Pitch Estimation of an Arbitrary Speech Signal Using MATLAB

Harish Chander[1], Balwinder Singh[2], Ravinder Khanna[3]
[1]*Department of Electronics Engineering, I. K. Gujral Punjab Technical University, Jallandhar, India*
[2]*Centre for Development of Advanced Computing, Mohali, India*
[3]*Department of Electronics & Communication Engineering, Maharishi Markandeshwar University, Sadopur, India*
(E-mail: [1]harishrajni@rediffmail.com, [2]balwinder@cdac.in, [3]ravikh2006@gmail.com)

Abstract—In this paper we present the short time cepstrum analysis method for pitch estimation of an arbitrary speech signal. We use MATLAB simulating software for our analysis purpose. The speech signal whose pitch is to be estimated is imported into the MATLAB SPTool box in .wav format. The sampled speech signal is further sliced to reduce its duration for pitch estimation. Pitch estimation of speech signal is useful for speech modeling, speech enhancement, speech synthesis, speech coding and various other application of speech signal processing.

Keywords—*Speech processing; Cepstrum analysis; Speech coding; Speech detection; Speech enhancement*

I. INTRODUCTION

Speech is a mixture of voiced and unvoiced types of sounds. The voiced sounds are quasi-periodic in nature and are produced when we speak vowels and some of the consonants. The unvoiced sounds are non periodic in nature and are produced when we speak other consonants. The voiced sounds have fundamental frequency which depends upon the vibration of the vocal cord of a person during the production of voiced sounds [1, 2].

Many pitch detection algorithms are available in the literature. Mostly these are divided into two groups: block based and event based [3]. In block based pitch estimation, the speech signal is sliced into small segments and it is assumed that pitch remains constant during these small segments of the speech [4]. Event based pitch detection methods are based on pitch marking or epoch detection. Since no assumption of constant pitch over several pitch cycles is made, event based pitch estimators are able to track fast changes in the pitch even during the segments [5]. Block based methods are more robust than event based methods but they neglect the variation in pitch that may exist within the segments. Block based methods do not give sufficiently accurate results required for speech processing applications that require pitch processing synchronization. Event based methods, on the other end, are more prone to errors since they rely on the shape of the speech wave form [6].

In this paper we present the very fundamental method of cepstrum analysis for pitch estimation which was also proposed by Noll [7]. Section II, of this paper describes about the pitch and cepstrum. Section III describes the procedure of importing speech signal into MATLAB workspace and then cepstrum analysis. Results are formulated in section IV and concluded in section V.

II. THEORY OF PITCH AND CEPSTRUM

A. Pitch

The pitch is a subjective attribute of the speech and is related to the fundamental frequency of the voiced speech signal. The relation between frequency and pitch is shown in Fig. 1. For fundamental frequency f Hz, it is measured in mels and is approximated by the relation given in (1) [1].

$$\text{Pitch in mels} = 1127 \text{ loge } (1 + f/700) \qquad (1)$$

Pitch plays an important role in speech processing, especially in voice recognition, speech synthesis, speech modeling, speech coding, and speech enhancement.

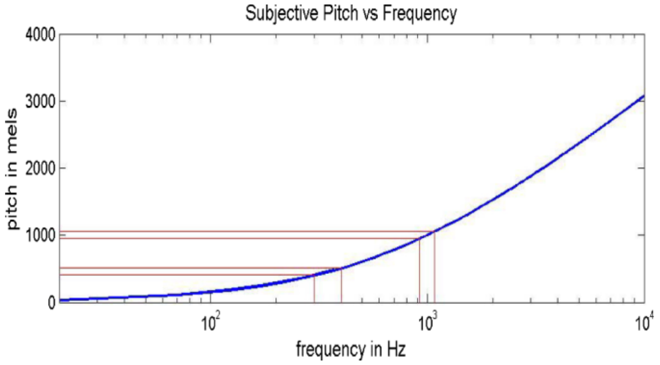

Figure 1 Relation between pitch and frequency

B. Cepstrum

The cepstrum is defined to be the inverse Fourier Transform of the log magnitude spectrum of a signal [8]. The cepstrum of a given signal X($e^{j\omega}$) is given by:

$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} log\left| X(e^{j\omega})\right| e^{j\omega n} d\omega \qquad (2)$$

Where $\log\left|X(e^{j\omega})\right|$ is the logarithm of the magnitude of the DTFT of the given signal. The complex cepstrum of the signal is given by:

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\{\hat{X}(e^{j\omega})\} e^{j\omega n} d\omega \qquad (3)$$

Where $\log\{\hat{X}(e^{j\omega})\}$ is the complex logarithm of $X(e^{j\omega})$ and is defined as:

$$\hat{X}(e^{j\omega}) = \log\{X(e^{j\omega})\} = log\left|X(e^{j\omega})\right| + jarg[X(e^{j\omega})] \qquad (4)$$

Speech is a type of signal whose properties changes very frequently. Due to this the processing of speech signal is done on segments of short duration. This requires replacement of DTFT into STFT (short time Fourier transform) of cepstrum, which is given by:

$$c_{\hat{n}}[m] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|X_{\hat{n}}(e^{j\hat{\omega}})\right| e^{j\hat{\omega} m} d\hat{\omega} \qquad (5)$$

Where$X_{\hat{n}}(e^{j\hat{\omega}})$ is the STFT given by:

$$X_{\hat{n}}\left(e^{j\hat{\omega}}\right) = \sum_{m=-\infty}^{\infty} x[m]\, w\,[\hat{n} - m] e^{-j\hat{\omega} m} \qquad (6)$$

Where $W[\hat{n} - m]$ is the sliding window function. Short time complex cepstrum can similarly be defined by:

$$\hat{x}_{\hat{n}}[m] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\{X_{\hat{n}}(e^{j\hat{\omega}})\} e^{j\hat{\omega} n} d\hat{\omega} \qquad (7)$$

C. Computation of the Cepstrum Using DFT

The cepstrum can be computed using DFT analysis, z-transform analysis or the recursive analysis. Here, we have used DFT analysis which is described below:

DFT is the sampled version of the DTFT of a finite length sequence, $X(k) = X(e^{\frac{j2\pi k}{N}})$ [9]. Using DFT, complex cepstrum can be computed with the help of following equations:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\left(\frac{2\pi k}{N}\right)n} \qquad (8)$$

$$\hat{X}[k] = log\left| X[k] + j \arg\{X[k]\} \right. \qquad (9)$$

$$\tilde{\hat{x}}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}[k] e^{j\left(\frac{2\pi k}{N}\right)n}, \text{ Inverse DFT} \qquad (10)$$

III. CEPSTRUM ANALYSIS PROCEDURE

The procedure to estimate the pitch using cepstrum analysis through MATLAB is explained below [10]:

- Import the .wav format of the speech signal whose pitch is to be estimated into the workspace. Let us name it as mv (for male voice) or fv (for female voice). Set the sampling frequency fs = 8000 Hz. This is saved in the workspace in the form of column vector whose size depends upon the sampling frequency and the duration of the .wav file.
- Slice down the above imported signal to 1 second duration using command mv1s = mv(40001:48000). This will retain the speech file from 6th to 7th second duration. To avoid initial pauses or noises, we have chosen intermediate portion of the speech signal.
- Speech signal has to be analysed for short duration only since its properties changes very frequently. We convert the 1 sec sliced signal into 15 segments of 50 millisecond each by moving the window in steps of 12.5 ms, i.e. 100 samples at a sampling rate of 8000 samples/sec. Let us name these segments mv1s_1 to mv1s_15 for male voice and fv1s_1 to fv1s_15 for female voice. We use command mv1s_1 = mv1s(1:400); mv1s_2 = mv1s(101:500); and so on upto 15th segment as mv1s_15 = mv1s(1901:2300).
- Convert these segments into cepstrum using command cepmv1s_1 = cceps(mv1s_1), for respective segments.
- Import these cepstrum segments into the signal processing tool, SPTool, box of MATLAB. Fig. 2 shows the SPTool box window.

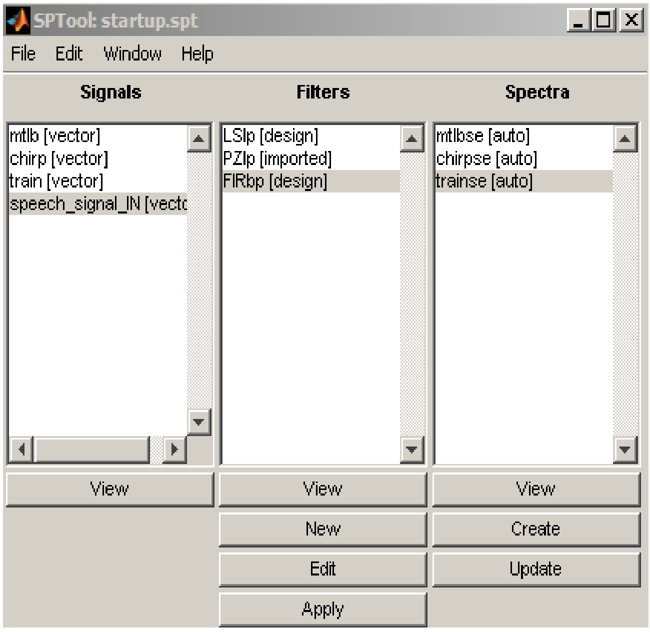

Figure 2 SPTool box browser window

- For each segment of the cepstrum, obtain time versus amplitude graph of short time cepstra from the signal window by selecting respective segment and clicking view. Graph for segment 1 is shown in Fig. 3.
- Create spectrum for each segment using spectra window of the SPTool. Obtain STFT for each segment.

Figure 4 shows FFT spectrum for segment 1. We have used number of samples N = 1024 for spectrum view.

- For pitch estimation set all the graphs of cepstrum spectrum one above the other in sequence. Likewise set all the graphs of cepstrum one above the other in sequence as shown in Fig. 5.

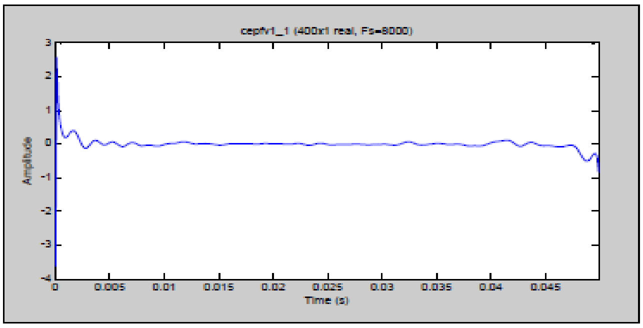

Figure 3 Short time cepstra of segment 1

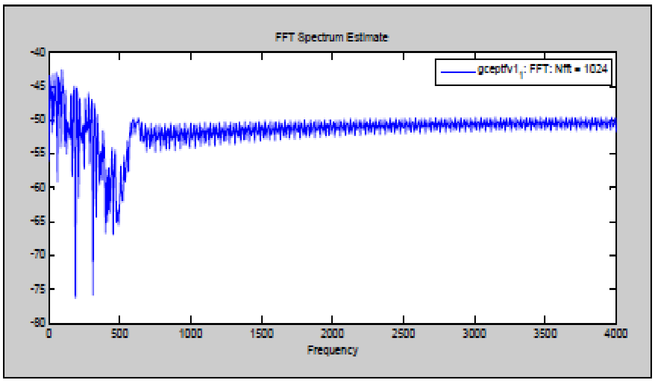

Figure 4 Spectrum of segment 1

IV. OUTCOMES AND RESULTS

From Fig. 5 we analyse that spectrum for segments 1 through 5 are non periodic and thus it is unvoiced portion of the speech sample. Spectrum for segment 6 and 7 seems to be partly voiced and partly unvoiced. Spectrum of segments 8 through 15 are of voiced portion of the sample speech. F1, F2 and F3 are marked as three formant frequencies of the sample speech.

On the right side of Fig. 5 we can see peaks between 11 – 12 ms. These peak points give accurate estimate of the pitch of the arbitrary speech sample.

V. CONCLUSION

In this paper we have exploited the MATLAB SPTool for accurate estimation of the arbitrary sample speech signal. We have used the fundamental cepstrum analysis method as described in [7] by Noll. The pitch estimation is useful for various speech processing such as speech synthesis, speech coding, speech enhancement and speech compression.

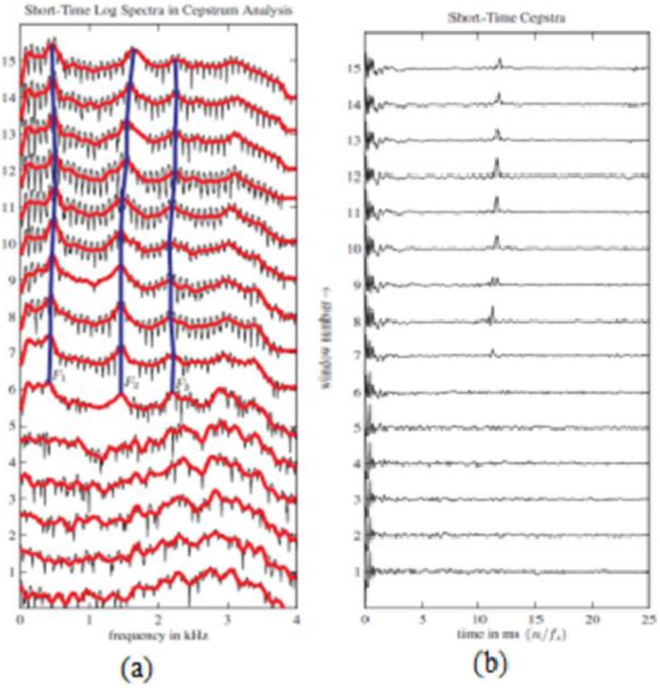

Figure 5 Display of (a) short time log spectra of 15 segments and (b) the corresponding short time cepstra

REFERENCES

[1] Lawrence R. Rabiner and Ronald W. Schafer on Introduction to Digital Speech Processing, Signal processing, vol.1, Nos. 1-2, NOW, 2007.

[2] Harish Chander Mahendru, "Quick review of human speech production mechanism," IJERD, vol. 9, issue 10 , pp. 48-54, 2014.

[3] P. Veprek and M. S. Scordilis, "Analysis, enhancement and evaluation of five pitch determination techniques," Speech Communication., vol. 37, pp. 249–270, 2002.

[4] T. F. Quatieri on Speech Signal Processing, Upper Saddle River, NJ: Prentice-Hall, 2002.

[5] T. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction of voiced speech," IEEE Trans. Acoust. Speech, Signal Process., vol. ASSP-23, no. 6, pp. 562–570, Dec. 1975.

[6] Barbara Resch, Mattias Nilsson, Anders Ekman, and W. Bastiaan Kleijn, "Estimation of the instantaneous pitch of speech," IEEE, transactions on audio, speech, and language processing, vol. 15, no. 3, 2007.

[7] A. M. Noll, "Cepstrum pitch determination," Journal of the Acoustical Society of America, vol. 41, no. 2, pp. 293–309, February 1967.

[8] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency alanysis of times series for echos: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking," Proceedings of the Symposium on Time Series Analysis, (M. Rosenblatt, ed.), New York: John Wiley and Sons, Inc., 1963.

[9] Proakis, J.G., and D.G. Manolakis on Digital Signal Processing: Principles, Algorithms, and Applications (Englewood Cliffs, NJ: Prentice Hall, 1996)

[10] Signal Processing Tool Box, User's guide, version 4.2, Math Works Incorporation.

Harish Chander, born in 1968, has graduated in Electronics and Communication Engineering from Institution of Engineers, India, in 1994. He has completed his Masters in Digital Systems from MNNIT, Allahabad, India, in 2002. He has served the Indian Air Force as electronics & telecommunication engineer for 20 years. For the last 11 years he has been serving various academic organizations at different positions and presently he is working as Controller of Examinations with Institution of Electronics & Telecommunication Engineers, India. He is pursuing his PhD from I K Gujral Punjab Technical University, Jallandhar, India. His areas of research are signal processing, speech processing and wireless communication.

Dr. Balwinder Singh has graduated in Electronics and Communication Engineering from National Institute of Technology, Jallandhar, India, in 2002. He has completed his Masters in Microelectronics from Punjab University, Chandigarh, India, in 2004 and PhD from Guru Nanak Dev University, Amritsar, India, in 2014. He is presently working as Senior Engineer & Coordinator ACS Division, Centre for Development of Advanced Computing (C-DAC), Mohali, India. His Research Interest are Low power VLSI Design and Testing, Digital IP cores and analog modules, Sensor and MEMS design and modeling, FPGA based embedded systems and Image Processing for Embedded systems.

Dr. Ravinder Khanna has graduated in Electrical Engineering from Indian Institute of Technology (IIT), Dehli in 1970 and has completed his Masters and Ph.D in Electronics and Communication Engineering from the same Institute in 1981 and 1990 respectively. He worked as an Electronics Engineer in Indian Defense Forces for 24 Years where he was involved in teaching, research and project management of some of the high tech weapon systems. Since 1996 he has full time Switched to academics. He has worked in many premiere technical institutes in India and abroad. Currently he is Professor and Dean Research with Maharishi Markandeshwar University, Sadopur, Haryana, India. He is active in the general area of Computer Networks, Image Processing and Natural Language Processing.