International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 10
ISSN: 2321-8169
5852 - 5854

A Model for Pitch Estimation Using Wavelet Packet Transform Based CEPSTRUM Method

Muhaseena T K
Electronics and Communication Department
Ilahia College of Engineering and Technology
Muvattupuzha, India
e-mail: muhaseenatk2009@gmail.com

Lekshmi M.S
Electronics and Communication Department
Ilahia College of Engineering and Technology
Muvattupuzha, India
e-mail: *lekshmims@gmail.com*

Abstract—A computationally efficient model for pitch estimation of mixed audio signals is presented. Pitch estimation plays a significant role in music audition like music information retrieval, automatic music transcription, melody extraction etc. The proposed system consists of channel separation and periodicity detection. The input signal is created by mixing two sound signals. First removes the short time correlations of the mixed signal. The model divides the signal into number of channels using wavelet packet transform. Computes the cepstrum of each channels and sums the cepstrum functions. The summary cepstrum function is further processed to extract the pitch frequency of two input signal separately. The model performance is demonstrated to be comparable to those of recent multichannel models. The proposed system can be verified by simulating the system in MATLAB.
Keywords*-Multipitch analysis, periodicity detection, cepstrum, auditory modeling, wavelet packet transform.*

*****

I. INTRODUCTION

Pitch estimation is very important for the processing of sound signal. Pitch is the basic part of the sound signal. Basically pitch means the highest or lowest tone of a particular sound that is perceived by our ear. It is depend on the number of vibrations that is produced by our vocal chords per second.

Many methods have been already proposed for the estimation of pitch frequency of human sound. For the voice of a single speaker we can easily compute the pitch frequency. But the complexity of the computation of pitch frequency increases by increasing the the complexity of the input signal. That is when the input signal is mixed with the voices of two speakers then the pitch determination is too difficult.

This paper is organized as follows. Section II presents theliterature survey and the existing methods. Section III presents the proposed system which includes the block diagram and expansion of each block. Section IV presents the experimental results, followed by conclusion in section V.

II. REVIEW

Recent years many models are proposed for the human pitch perception and periodicity detection. Most of the models are used filterbank approach to implement the function of cochlea. The unitary pitch analysis model of Meddis and O'Mard and its predeccessors by Meddis and Hewitt are among the best known recent models of time domain pitch analysis. Here the filter bank divide the single channel signal into number of channels.

The envelop of the each channel is computed by using the half wave rectification and lowpass filtering. After computing the envelop of each channel determine the periodicity by the autocorrelation method. In autocorrelation method compared the signal with its past and future values. Finally sums the autocorrelation of each channel to get the summary autocorrelation function. This summary autocorrelation functions are the good indicators of the pitch.

Fig.1 shows the above discussed pitch determination model using filterbank.There are different approaches for the computation of autocorrelation function.

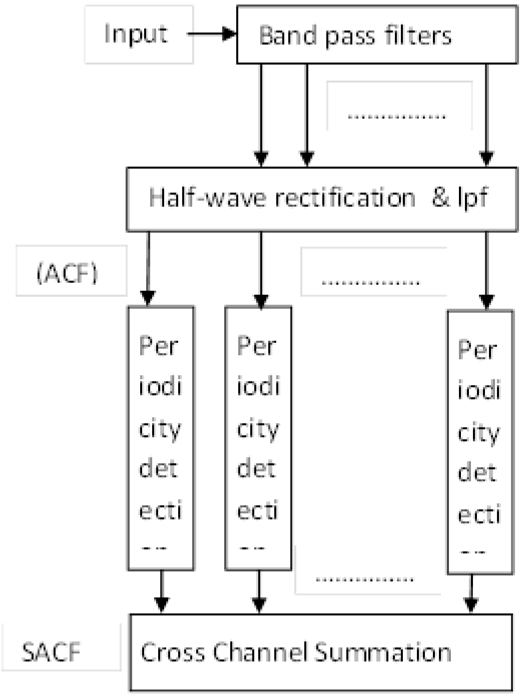

Figure.1 Pitch estimation using filter bank

By using the spectral smoothness principle a model is proposed by Anssi P Klapuri for the determination of pitch. Here assumes the spectral envelopes of real sounds are continuous. The smoothing operation corrected approximately half of the pitch errors that occurred in a system. By using spectral smoothness principle the spectral components of a source can be separated from the mixture. Now the multipitch estimation is reduces to single pitch estimation.

Another method of pitch detection is by using the AMDF. The fundamental frequency f0 is the main part of the pitch. In average magnitude difference function instead of correlating the input speech at various delays a difference signal is formed between the delayed speech and the original. Then the absolute value of the magnitude is taken for each delay. The nonlinear processing is usually used for the pitch tracking to reduce the effect of formant structure.

Fig.2 shows the another method to find out the multipitch of the complex signals. Here first the short time correlation of the input signals are removed by using the pre-whitening filter . The linear prediction by pre-whitening implements the critical band auditory resolution of spectral modeling and reduce the filter order for following steps .

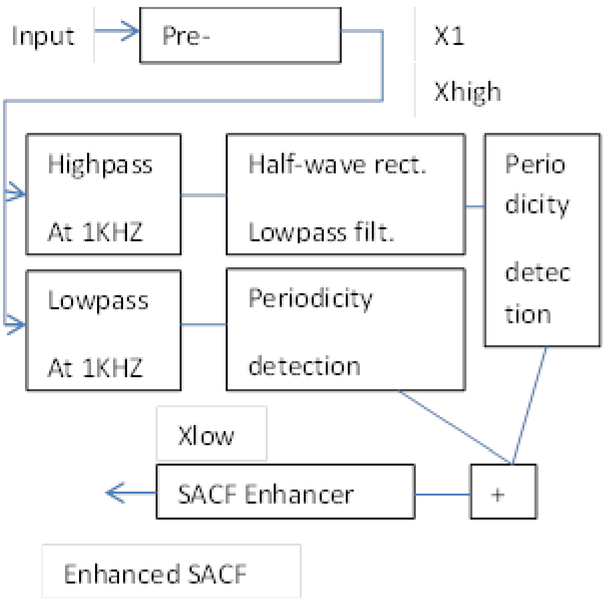

Figure.2:Pitch estimation using autocorrelation method

Then the single signal channel is divided into two channel that is below and above 1000Hz. The highpass channel is again halfwave rectified and lowpass filtered in order to separate the low channel. Compute the periodicity of the each channel using autocorrelation. Autocorrelation uses the discrete fourier transform and inverse fourier transform. Autocorrelation is computed by using the following equation

$$x2 = IDFT(\,|DFT(\text{xlow})|^{K}) + IDFT(\,|DFT(\text{xhigh})|^{K}) \qquad (1)$$

From the SACF signal it is too difficult to estimate the true pitch. For that enhance the SACF signal.

III. PROPOSED METHOD

The proposed system consists of mainly four parts. The are pre-whitening section, channel separation, periodicity detection and enhancing the periodicity. The block diagram of proposed model is shown in fig.3.

A. Pre-whitening

Initially the voice signal of two speakers mixed together to get a mixture. Mixture signal is the input to the proposed system. Then by using the wraped linear prediction compute the linear prediction coefficients. The wraped linear prediction is similar to that of ordinary linear prediction. But the wraped linear prediction implements the critical band auditory resolution for spectral modeling. Linear prediction filter implements the uniform frequency resolution. So we can reduce the order of filter that is used for the following steps.

Before that the input mixed signal is divided into frames. Hamming window is used to divide the signal into frames. Frame size of 23.2ms is used.so pre-whitening is done after framing the sound. Pre-whitening removes the short time correlations between the signal with its delayed version. So it gives the better resolution in peaks after periodicity detection.

B. Channel separation

The linear prediction coefficients obtained for the input signal is a single channel signal. In this step we divide the single channel into four channels using wavelet packet transform. Each channel represents either lowpass or high pass filtering. The half of the signal channel is half wave rectified and lowpass filtered to separating the low channel. We used Daubechies 3 wavelets in the design as it gave better results compared to other wavelets available in Matlab wavelet toolbox.

C. Periodicity detection

The periodicity of the each channel is computed separately. First the window function is determined using hamming window. Then the window coefficients are summed with the each of the channel separately. Take the DFT of the signal,then take the logarithm of each channel finally performs the inverse fourier transform of the signal. Sum the inverse fourier transform of the each channel. Now get the pitch frequency of complete sound signal. Pitch frequency gives the pitches of the sound signal. But from this summed coefficients we can't extract the correct pitches of the signal. So process these pitches further by following steps.

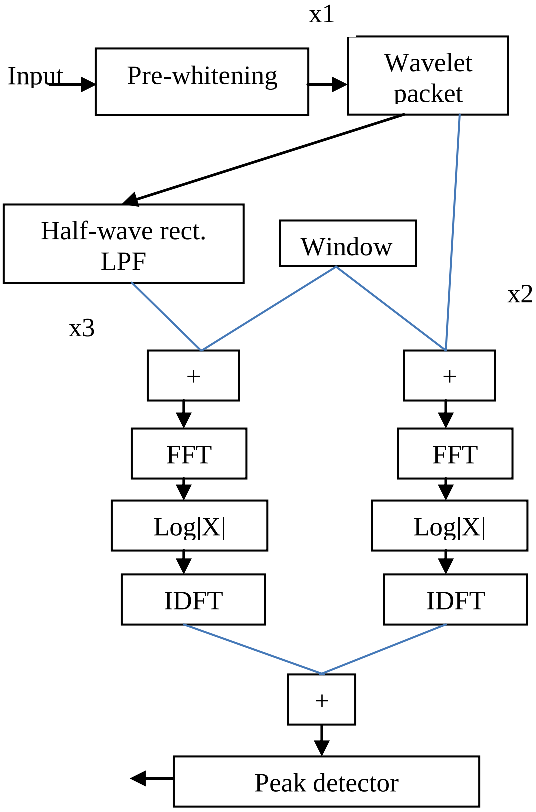

Figure.3: Block diagram of proposed pitch estimation method

$$x2 = IDFT\big(log(DFT(\text{x2} + window))\big) + IDFT\big(log(DFT(\text{x3} + window))\big) \qquad (2)$$

D. Peak detector

Here the signal is clipped to positive values only. Then the time is expanded by time scaling by a factor 2. Now the signal is substracted from the original negative clipped signal. After the substraction it may also contains negative, positive values. So the negative pitches are removes by using clipping to positive values only.

Now to extract the pitch frequencies of the two input signal separately find out the maximum peak and minimum peak amplitudes of the signal. And then compute pitch frequency of the respective maximum and minimum amplitude. The maximum amplitude is correspond to dominant signal and minimum amplitude is correspond to interference signal. Then finally compare the proposed model with the existing pitch estimation model by root mean square error.

IV. EXPERIMENTAL RESULTS

The proposed model is verified for different input mixed signals like male-male signal, female-female signal, male and female with male dominant, male and female with female dominant etc.

TABLE I
ROOT MEAN SQUARE ERROR OF PROPOSED METHOD

	Root Mean Square Error			
	Dominant		***Interference***	
	Autcrltn	***Cepstrum***	***Autcrltn***	***Cepstrum***
A	76	82	70	53
B	90	89	82	60
C	63	64	82	67
D	76	85	68	59

TABLE II
ROOT MEAN SQUARE ERROR OF EXISTING METHOD

	ROOT MEAN SQUARE ERROR			
	Dominant		***Interference***	
	Autcrltn	***Cepstrum***	***Autcrltn***	***Cepstrum***
A	122	111	140	92
B	134	137	99	69
C	123	113	99	72
D	132	131	140	98

A-Male-male signal
B-Female-female signal
C-Male female with male dominant
D-Male female with female dominant

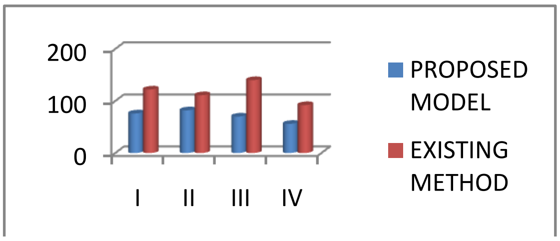

Figure 4.performance comparison of existing and proposed methods

I-Dominant RMSE of autocorrelation method
II- Dominant RMSE of cepstrum method
III-Interference RMSE of autocorrelation method
IV-Interference RMSE of cepstrum method

V. CONCLUSION AND DISCUSSION

In this paper we proposed a new approach for estimating the pitches of a sound signal based on wavelet packet transform and cepstrum method. The model has been developed as a compromise for computational efficiency. Pre-whitening based on wraped linear prediction removes the short time correlation between the signals. The proposed model provides the better root mean square error and reduced computational complexity.

REFERENCES

[1] R. C. Maher and J.W. Beauchamp, “Fundamental frequency estimation of musical signals using a two-way mismatch procedure,” *J. Acoust. Soc.Amer.*, vol. 95, pp. 2254–2263, Apr. 1994.
[2] K. Kashino and H. Tanaka, “A sound source separation system with the ability of automatic tone modeling,” *Int. Computer Music Conf.*, pp. 248–255, 1993.
[3] A. Klapuri, “Number theoretical means of resolving a mixture of several harmonic sounds,” in *Proc. Signal Processing IX: Theories Applications*, vol. 4, 1998, p. 2365.
[4] K. D. Martin, “Automatic transcription of simple polyphonic music: Robust front end processing,” Mass. Inst. Technol., Media Lab Perceptual Computing, Cambridge, Tech. Rep. 399, 1996.
[5] D. P. W. Ellis and B. L. Vercoe, “A wavelet-based sinusoid model of sound for auditory signal separation,” in *Int. Computer Music Conf.*, 1991, pp. 86–89.
[6] R. Meddis and L. O’Mard, “A unitary model for pitch perception,” *J. Acoust. Soc. Amer.*, vol. 102, pp. 1811–1820, Sept. 1997.
[7] M. Slaney and R. F. Lyon, “A perceptual pitch detector,” *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. 357–360, 1990.[28] U. K. Laine, M. Karjalainen, and T. Altosaar, “Warped linear prediction(WLP) in speech and audio processing,” *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. III.349–III.352, 1994.
[8] U. K. Laine, M. Karjalainen, and T. Altosaar, “Warped linear prediction (WLP) in speech and audio processing,” *Proc. IEEE Int. Conf. Acoustics,Speech, and Signal Processing*, pp. III.349–III.352, 1994.
[9] Anssi P. Klapuri“Multipitch estimation and sound separationBy the spectral smoothness principle”*Tampere University of Technology, P.O.Box 553, FIN-33101 Tampere, Finland.*
[10] Li Tan and MontriKarnjanadecha“Pitch detection algorithm: autocorrelation method and amdf”*Department of Computer Engineering Faculty of Engineering* Prince of Songkhla UniversityHat Yai, SongkhlaThailand, 90112
[11] TeroTolonen, “A Computationally Efficient Multipitch Analysis Model” *Student Member, IEEE*, and MattiKarjalainen, *Member, IEEE*