Design and implementation of Auto Tempo Detection algorithm

*Hyo-Moon Cho, **Si-Kong Lee, Snag-Bong Yang, Sang-Bock Cho

School of Electrical Engineering, University of Ulsan
E-mail : hmcho67@hanmail.net

Abstraction

Tempo detection is basic in automatic music processing and in many multimedia applications. This paper presents the implementation algorithm and design for tempo tracking from polyphonic music signals on non-PC system. The method of crossover switching is proposed and the concept of variable threshold for energy flux operation is defined. This system consists of the pre-processing part and main processing part. The preprocessing part is based on an efficient filter- bank separation with logarithmic audio theory. And main processor part is based on energy flux sum or production. This system is evaluated using 100 sample songs from several musical genres. The results rate is about 85%.

Keywords: *beat, tempo, crossover switching, variable threshold.*

I. Introduction

The beat is a fundamental unit of the temporal structure of musical signal. For this reason, automatic beat or tempo tracking is an essential task for many applications such as musical analysis, automatic rhythm alignment of multiple musical systems, audio editing, and beat driven special effects.

The tempo tracking is difficult task for various genres because the frequency characteristics are different depend on the musical genres. [3]

The effective frequency range of hearing is 100Hz~10K Hz although the range of hearing is 20Hz~20K Hz. Therefore, the determination of the crossover frequencies and the select of the number of sub-band bank are very important to decide the algorithm performance. [2, 7, 9]

Traditional tempo detection algorithms focused on symbolic format like as MIDI, CD, etc. In this reason, tempo estimation systems have been implemented on a PC system and have been taken a limitation of usage. [1] To solve these limitations, we used practical music signals which are through microphone and designed non-PC system with FPGA chip.

And early proposed tempo estimation algorithms presented difference results according to genres. Scheirer[12, 15] proposed a method associating a sub-bank with a set of comb-filters. Its efficiency is depends on crossover frequencies and the number of sub-bank. The performance of tempo detection system is increase in out of proportion with the number of sub-bank. But the large number of sub-bank need many operation time and it exert a bad influence.

Recently, the well-established Alonso algorithm uses an eight sub bank, it assumes that the maximum crossover frequency of the effective hearing frequency range is 4K Hz and takes 8K Hz sampling frequency. [1] Therefore, the crossover frequencies are separated from 100Hz to 4K Hz by logarithmic theory in the Alonso algorithm. But, this algorithm could not guarantee for all genres, especially it has bad performance in the polyphonic music such as classic. It is by mis-assigning the crossover frequencies.

We decided hearing frequency bandwidth that guaranteed the same and effective system performance for various musical genres. These crossover frequencies are in from 100Hz to 6.4K Hz by logarithmic theory. And we adopted modified spectral products algorithm. [1]

The performance of our new algorithm is evaluated by using Altera FPGA(FLEX EPF10K70RC240-4) chip and 100 sample songs.

This paper presents several tempo tracking algorithms and describes our new algorithm and its design method. Results of these algorithms compared are presented.

II. Tempo tracking algorithms

The simple method is based on a simple inter-onset time detection of on the traditional autocorrelation method by Seppännen, Tzanetakis and so on.

In this paper, several algorithms for tempo tracking are introduced. These algorithms are based on the association of sub-bank.

Most tempo tracking algorithms are based on the same basic steps. They process several frequency bands separately, detects onset in each sub-band, estimates the periodicity in each sub-band and combine the results in each sub-band to obtain the output tempo signal.

According to these descriptions, the proposed structure of the whole system is shown in fig. 1.

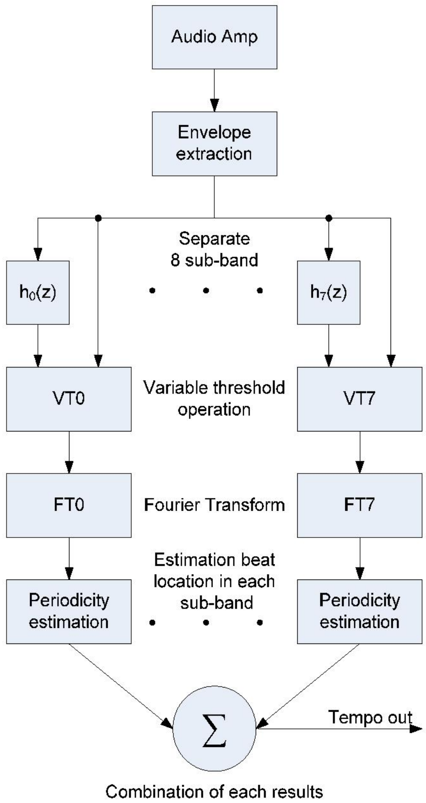


Fig. 1. Block diagram of the proposed auto tempo detecting system.

Alonso algorithm uses the 8K Hz of sampling frequency and adopted the eight non-overlapping frequency bands. It assume that effective hearing frequency range is 100Hz ~ 4K Hz. Therefore the crossover frequencies are logarithmically distributed in frequency between 100Hz and 4K Hz. This 4K Hz of maxi-mum crossover frequency has some limitations because poly-phonic music like as classic has many classified musical instruments between 2K Hz and 6K Hz. For our experiments, 6.3K Hz of the maximum crossover frequency has the most effective performance.

This crossover frequency is important factor in auto tempo tracking system. It is easy to obtain tempo tracking for the robust music like as rock, reggae, and so on.

In Alonso algorithm, the sequence of tempo estimation system processing is as below;

1) Converts analog-to-digital
2) Separates an eight sub-band
3) Processes FFT (Fast Fourier Transform) in each sub-band
4) Obtains an enveloped signal in each sub-band
5) Operates spectral products in each sub-band
6) Detects onset in each sub-band
7) Estimates periodicity in each sub-band
8) Combine the each results

This sequence is difficult to implement on non-PC system. To easy logic design, we modified processing sequence like as below;

1) Obtains an enveloped signal by using analog circuits
2) Separates an eight sub-band by using analog circuits
3) Converts analog-to-digital
4) Operates spectral products in each sub-band
5) Estimates periodicity in each sub-band
6) Combine the each results

In proposed sequence, onset detection in each sub-band does not needed. Because first two steps includes onset detection. According to our algorithm, the proposed sequence needs an eight ADC. But we must decrease the number of ADCs by reason of considering of chip area overhead. General switching methods includes some loss at switching time and it affects next step operations. So we proposed crossover switching algorithm and its concept diagram is in fig. 2.

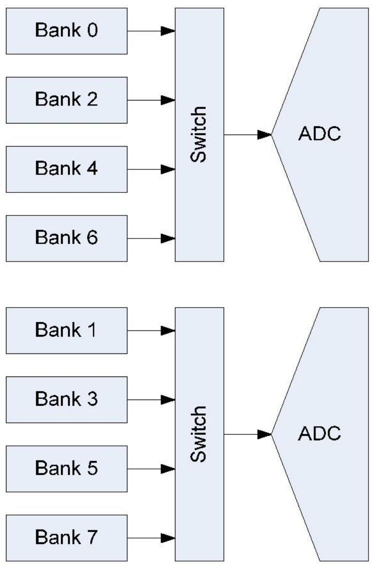


fig. 2. The crossover switching algorithm

The crossover switch algorithm is as follow. Firstly, input audio signal goes to each separated sub-bank. The outputted signals of each sub-band switched to two ADC by switching control signals. Switching time sequences are as below to decrease loss:

Analog switch1: B0-B2-B2-B4-B4-B6-B6-B0-B0-B2-…
Analog switch2: B1-B1-B3-B3-B5-B5-B7-B7-B1-B1-…

And its experimental circuit construction is shown fig. 3, and whole experimental system picture is as fig.4.

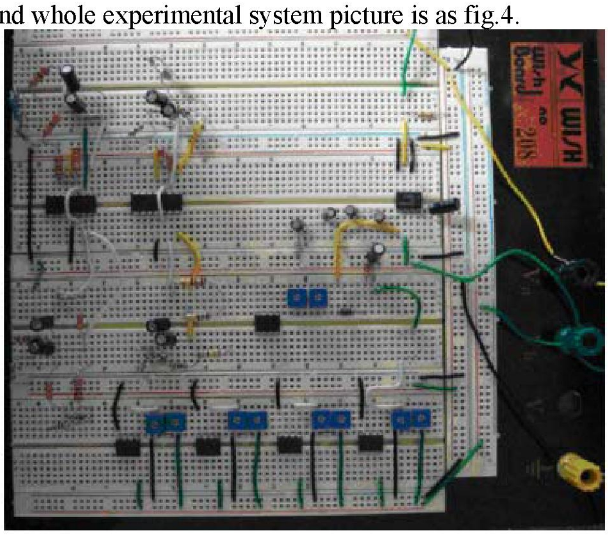


fig. 3. The picture of the analog circuit part

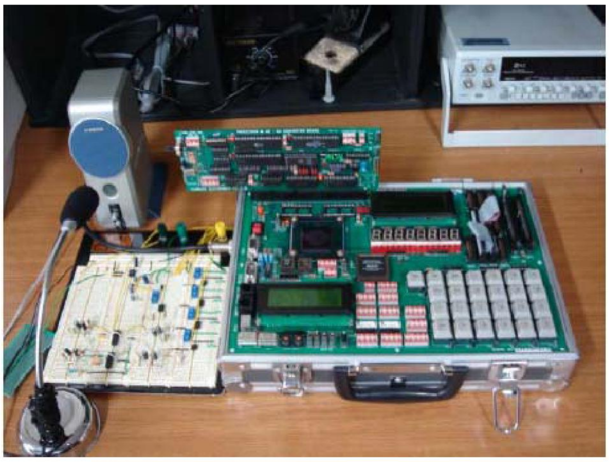


fig. 4. The picture of the whole experimental system

To obtain energy flux of the input audio signal, FFT(Fast Fourier Transform) is used and its expression is as eq. (1);

$$\widetilde{X}(f,m) = \sum_{n=0}^{N-1} w(n)x(n+mM)e^{-j2\pi fn} \qquad (1)$$

Where x(n) denotes the audio signal, w(n) is the finite analysis window of size N in sample. M is the hop size in sample, m the frame index and f is the frequency. And we can get its energy flux as eq. (2);

$$E(f,m) = H(e^{j2\pi f}) * \widetilde{X}(f,m) \qquad (2)$$

Where $\widetilde{X}(f,m)$ means the value is obtained by variable threshold processed value. To implement eq. (2), we used VTG(Variable Threshold Generator) and its basic concept is as below;

$$A.B \quad (A \geq 0,\ 0 \leq B < 1) \text{ then} \qquad (3)$$
$$A^2 >> A \text{ and } B^2 << B$$

It means the threshold values is equal to a decimal point, so the square value of a real number part is more larger than original number and the square value of fraction part is more smaller than it.
Where, B is decided by variable threshold generator and threshold value is obtained by as below;

ID: digitalized input signal
IB: digitalized sub-band output signal

The threshold value of banked signal is decided by

$$|IB| \geq 0.7|ID| \qquad (4)$$

And its block diagram is shown fig. 5.

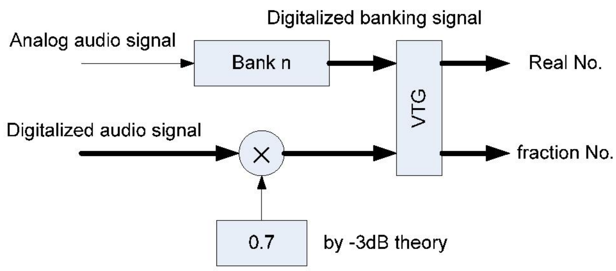


fig. 5. The VTG(variable Threshold Generator) block diagram

These processes are to easy implement for audio signal spectral energy flux. The frequency domain transformation by a Fourier transform is needed to audio signal analysis. Short signal segments are extracted at regular time intervals, multiplied by an analysis window and FFT. This leads to eq. (1).

We obtained beat position in each sub-band and its periodicity by using simple energy flux algorithm. To implement this, we used 500K Hz sampling frequency and designed FFT and 8-bit ALU logic. This whole system block diagram is shown as fig.6.

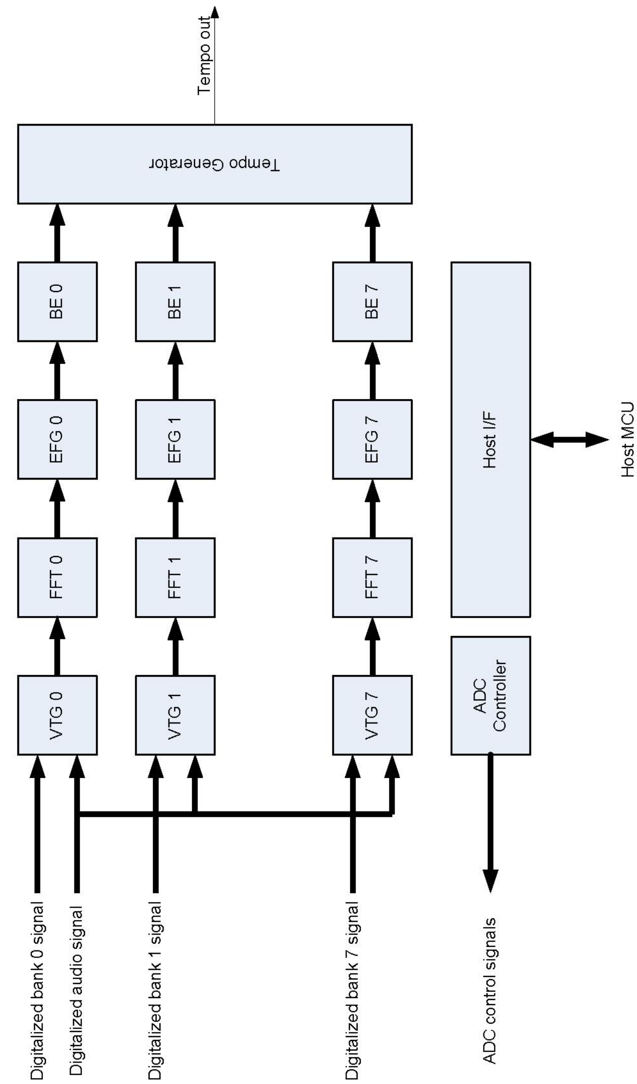


fig. 6. Block diagram of whole logic

The operation frequency of each bank is 12.6K Hz in sampling theory. According to our proposed crossover switching algorithm and advanced performance of algorithm, we used 500K Hz sampling frequency. Thus, VTG(Variable Threshold generator) , RFG(Energy Flux generator), BE(Beat location estimator) logics in each bank are operated by about 40K Hz. So this operation frequency is sufficient for calculation time. The gate count of our algorithm is about 40K gates. The beat location in each bank is obtained by calculating periodic estimation logic. Finally, main processor combines each beat location and calculates output tempo signal.

The logic schematic of FFT on Altera FPGA chip is shown in fig. 7. We have evaluated our algorithm by using 100 sample songs which are in several music genres. Our sample songs include music instruments and singer voice except classic genre.

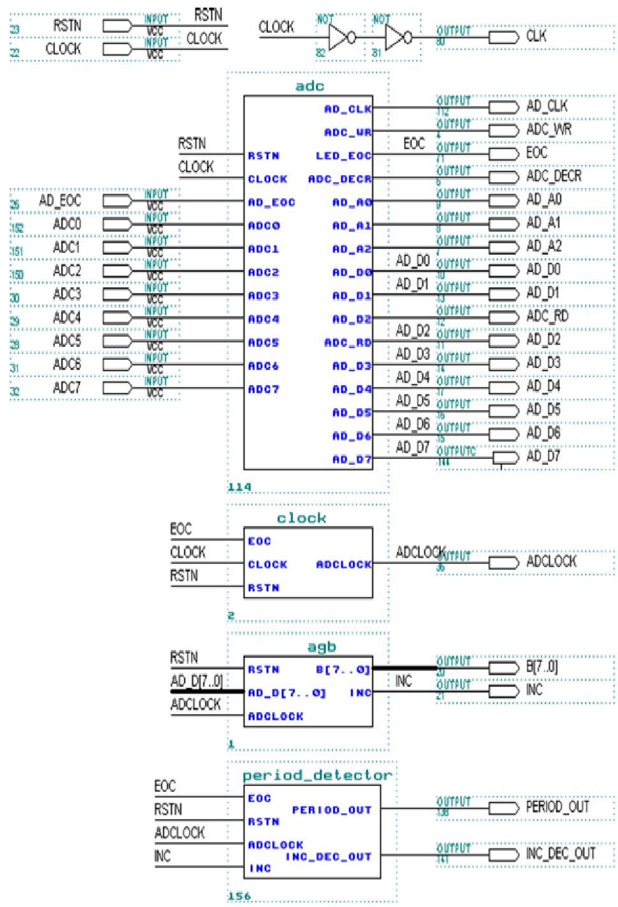


fig. 7. FFT logic schematic

III. Performance Analysis and Results

The proposed algorithm and system was evaluated using 100 songs it taken from commercial CD recordings, and MP3 format on internet sites. These pieces were selected to cover as many characteristics: various tempi in 50 to 200 BPM range, old/recent recordings, many instruments are using. Its genre distribution of the sample songs are table 1.

Genre	Pieces	Percentage
Ballard	20	20%
Dance	10	10%
HipHop	10	10%
Reggae	10	10%
Rock	10	10%
R&B	10	10%
Classic	30	30%
Total	**100**	**100%**

table 1. Genre distribution of the sample song

The procedure for estimating the tempo of each musical piece is as following:

- taps the tempo, while listening to a sample songs
- The tapping signal is recorded and its intervals are measured
- compared between our tapping signal and output signal of our implemented system

The beat is a perceptual concept that human feel in music, thus it is difficult to find “the exact beat location” in general. Thus, we adopted a 2% tolerance rate from manually annotated tempo signal for our auto tempo tracking system.

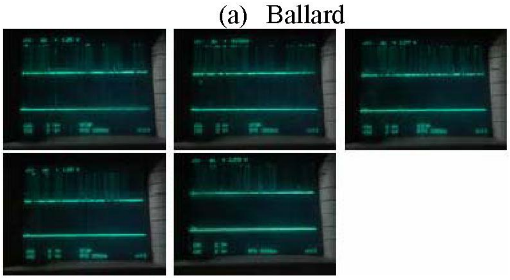
The recognition rate is about 85% and its value is similar for all genres. This rate is some low than Alonso algorithm (89.9%), but our result is form non-PC system. Our future work will upgrade the recognition rate by improvement of algorithm and design. And also, we will apply to general people speech and we will find speech rhythm of people speech habit.

Table 2 is shown to recognition rate for each genre and fig. 8 is for its output waveform by using oscilloscope.

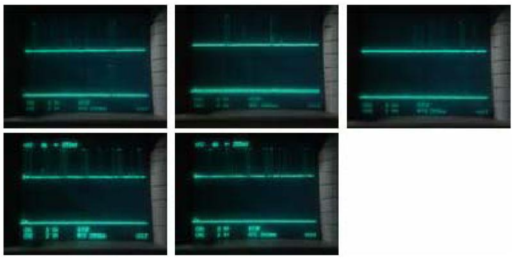
Genre	Average Recognition rate
Ballard	85%
Dance	83%
HipHop	85%
Reggae	85%
Rock	87%
R&B	87%
Classic	83%
Total	**100%**

table 2. The results of recognition rate for each genres

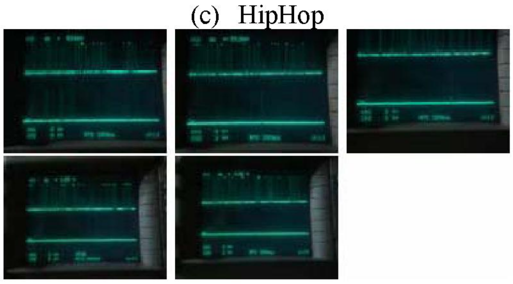
(a) Ballard



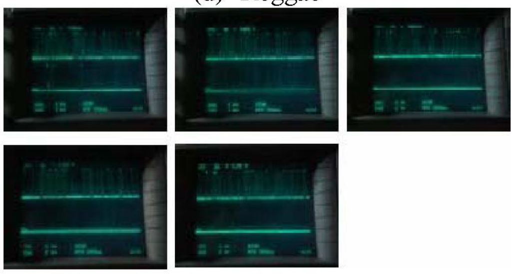
(b) Dance



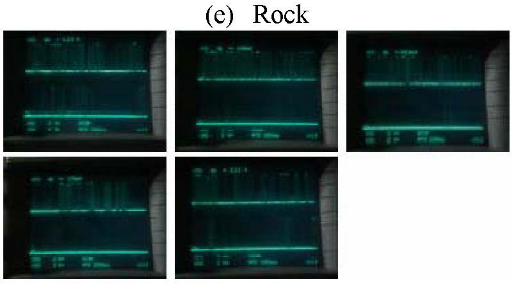
(c) HipHop



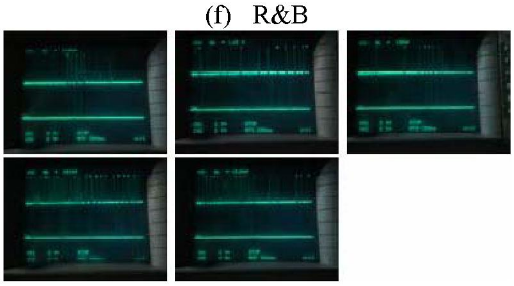
(d) Reggae



(e) Rock



(f) R&B



(g) Classic

fig. 8. The waveforms of test results

This paper was supported by BK21 project of MOE, NARC(Network-based Automation Research Center)/ IDEC(IC Design Education Center) of MOCIE, IT-SoC project of MIC and Univ. of Ulsan.

References

[1] Alonso M., David B. and Richard G., “A Study of Tempo Tracking Algorithms from Polyphonic Music

Signals", Proceedings of the 4th. COST 276 Workshops, Bordeaux, France. March 2003.
[2] Dixon, M. "Automatic Extraction of Tempo and Beat from Expressive Performances", Journal of New Music Research, vol. 30 No. 1, pp. 39-58 2001.
[3] Duxbury C., Sandler M. and Davies M., "A Hybrid Approach to Musical Note Onset Detection," Proceedings of the 5th. Int. Conf. on Digital Audio Effects (DAFx), pp. 33–38, Hamburg Germany, September 2002.
[4] Goto M. and Muraoka Y., "Issues in Evaluating Beat Tracking Systems," Working Notes of the IJCAI-97Workshop on Issues in AI and Music, August 1997.
[5] Goto, M. , "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds", Journal of New Music Research, vol. 30, No. 2, pp. 159-171, 2001.
[6] Klapuri, A., "Musical meter estimation and music transcription," Cambridge Music Processing Colloquium, Cambridge University, UK, March 2003.
[7] Klapuri, A., "Sound Onset Detection by Applying Psychoacoustic Knowledge," Proceedings IEEE Int., Conf., Acoustics Speech and Sig., Proc. (ICASSP), pp. 3089-3092, Phoenix AR, USA March 1999.
[8] Laroche, J., "Estimating, Tempo, Swing and Beat Locations in Audio Recordings," Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 135-138, New Paltz, NY, USA, October 2001.
[9] Laroche, J., "Efficient Tempo and Beat Tracking in Audio Recordings," J. Audio. Eng. Soc., vol. 51, No. 4, pp. 226-233, April 2003.
[10] Paulus J. and Klapuri A., "Measuring the Similarity of Rhythmic Patterns," Proceedings of the International Symposium on Music Information Retrieval, Paris, France, 2002.
[11] Proakis, J. G. and Manolakis, D., Digital Signal Processing: Principles, Algorithms and Applications. Prentice Hall, New York, 1995.
[12] Scheirer, E. D., "Tempo and Beat Analysis of Acoustic Music Signals," J. Acoust., Soc. Am., vol. 104, pp. 588-601, January 1998.
[13] Seppa anen, J., "Tatum grid analysis of musical signals," Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA October 2001.
[14] Todd, N. P. McA., "The Auditory Primal Sketch: A Multi-scale model of rhythmic grouping," Journal of New Music Research, vol. 23, No. 1, pp. 25 - 70, 1994.
[15] Eric D. Scheirer. Tempo and beat analysis of acoustic music signals. J. Acoust. Soc. Am., Vol. 103(1): pp. 588-601, Jan. 1998.
[16] Jouni Paulus and Anssi Klapuri.: Measuring the similarity of rhythmic patterns. The 3rd. International Conference on Music Information Retrieval (ISMIR), 2002.
[17] Anssi Klapuri. Sound onset detection by applying psychoacoustic knowledge. In IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), pages pp. 3089-3092, 1999.