International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:12, 2009

Enhanced Parallel-Connected Comb Filter Method for Multiple Pitch Estimation

Taro Matsuno, Yuta Otani, Ryo Tanaka, Kaori Ikezaki,
Hitoshi Yamamoto, Masaru Fujieda and Yoshihisa Ishida

Abstract**—** This paper presents an improvement method of the multiple pitch estimation algorithm using comb filters. Conventionally the pitch was estimated by using parallel -connected comb filters method (PCF). However, PCF has problems which often fail in the pitch estimation when there is the fundamental frequency of higher tone near harmonics of lower tone. Therefore the estimation is assigned to a wrong note when shared frequencies happen. This issue often occurs in estimating octave 3 or more. Proposed method, for solving the problem, estimates the pitch with every harmonic instead of every octave. As a result, our method reaches the accuracy of more than 80%.

Keywords**—**music transcription, pitch estimation, comb filter, fractional delay.

I. Introduction

In this paper we propose a new algorithm to estimate pitch for automatic music transcription (AMT). The AMT has a long history as a research theme going back to the 1970s [9]. In recent years Kashino *et al.* have proposed the OPTIMA processing architecture which utilizes auditory scene analysis based on a probabilistic model [6]. Goto has proposed PreFEst, which estimates relatively dominant fundamental frequencies by assuming that all single tones sound simultaneously [7]. Kameoka *et al.* have proposed a method using harmonic clustering (HC), which does not use the prior information [3]. This method has been expanded into harmonic temporal structured clustering (HTC) [4] and Harmonic-Temporal-Timbral clustering (HTTC) [5]. Kulapuri have proposed a method for sequentially determining the components in a sound mixture by repeatedly estimating the predominant fundamental frequency and removing its harmonic components [8]. Morita *et al.* have proposed the Parallel-Connected comb filters method (PCF) [1][2]. In this method the pitch is estimated by comb filters which remove the fundamental frequency and harmonics.

We regard the method of the PCF, because it can estimate pitch with lower calculation amount than other methods. However the method has the problem that the accuracy is poor in octave 3 or more, because in the method comb filters remove frequencies with every octave. Proposed method configures comb filters which remove frequencies with every harmonic instead of every octave. This paper shows the accuracy of the pitch estimation is improved, and shows the availability for the music transcription.

Authors are with the Department of Science and Technology, Meiji University 1-1-1 Higashi-mita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan. Phone: +81-44-934-7307 E-mail:ce81057@isc.meiji.ac.jp

II. Conventional Method with Comb Filters

In this section we explain the theory of the pitch estimation with comb filters and its problem.

A. *Fractional delay comb filters* [10]

This paper uses fractional delay [11] comb filters. Let us define octave number and note number as q and p. The comb filter $H_{q,p}(z)$ is written as the following:

$$H_{q,p}(z) = 1 - z^{-m} A(z) \tag{1}$$

$$A(z) = \frac{a_N + a_{N-1}z^{-1} + ... + a_1 z^{-(N-1)} + z^{-N}}{1 + a_1 z^{-1} + ... + a_{N-1} z^{-(N-1)} + a_N z^{-N}} \tag{2}$$

$$D = \frac{fs}{f_{q,p}} \tag{3}$$

$$N - 0.5 \le d \le N + 0.5 \tag{4}$$

$$D - d = m \qquad (m;\text{Integer}) \tag{5}$$

where D denotes the desired fractional delay, and it is represented by using sampling frequency fs and fundamental frequency $f_{q,p}$ as equation (3). Fig.1 shows the fractional delay comb filter.

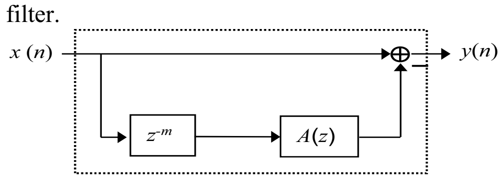

Fig.1 The fractional delay comb filter

B. *Theory of pitch estimation with comb filter*

Comb filters have zero points nearly regularly in frequency domain (Fig.2).

Note number is estimated by using 12 comb filters $H_{2,1}(z)$ ~ $H_{2,12}(z)$.

$H_{2,p}(z)$ removes all harmonic components in octave 2~5. Zero output detection, $y_{2,p}(n)$=0, decides a note number p. And then an octave number is estimated by using $H_{q,p}(z)$ as follows:

$$\text{octave number} = \begin{cases} 2 & y_{3,p}(n) \neq 0 \\ 3 & y_{3,p}(n) = 0, y_{4,p}(n) \neq 0 \\ 4 & y_{4,p}(n) = 0, y_{5,p}(n) \neq 0 \\ 5 & y_{5,p}(n) = 0 \end{cases} \quad (6)$$

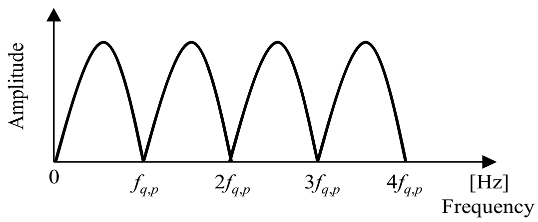

Fig.2 Comb filter's amplitude characteristic

Practically the zero output is estimated by comparison to a threshold value as follows:

$$Y_{q,p}(n) = E(|y_{q,p}(n)|) \quad (7)$$

$$y_{q,p}(n) \begin{cases} = 0 & (Y_{q,p}(n) \leq TH) \\ \neq 0 & (Y_{q,p}(n) \geq TH) \end{cases} \quad (8)$$

where *TH* is the threshold value, which is derived empirically. The output which falls below *TH* is considered zero output.

C. *Parallel-connected comb filter method*

Procedure based on above theory is described. The process has 5steps. The maximum number of concurrent sounds is known. If zero output is not obtained the process finishes as impossible estimation. Fig.3 shows the flow of the process.

Step.1 Input signal $x(n)$.

Step.2 Filter $x(n)$ through 12 comb filters and compare all output $y_{q,p}(n)$.

Step.3 The first note is set the note number p_1 when the output becomes smallest.

Step.4 If the minimum output falls below the threshold value *TH*, go to Step.5. If it does not, go to Step.1.

Step.5 Extract the targeted single by using comb filters $H_{2,p_1}(z),\dots,H_{2,p_r}(z)$ where maximum r is the number of concurrent sounds. Octave is estimated every note by the equation (6).

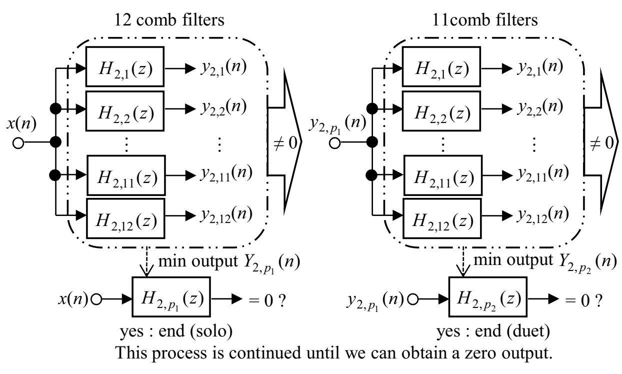

Fig.3 The PCF method algorithm for the pitch estimation

D. *PCF method's problem*

As previously explained, comb filters have zero points nearly at regular intervals in frequency domain. Therefore two or more different comb filters remove an input undesirably. (Fig4-A, 4-B). This often happens in octave 3 or more.

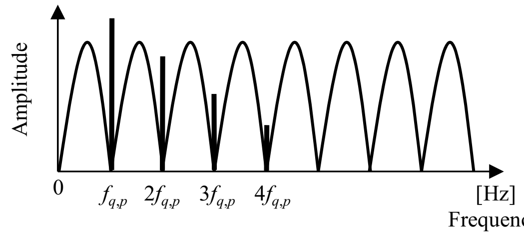

Fig.4-A The PCF method achieves the correct estimation

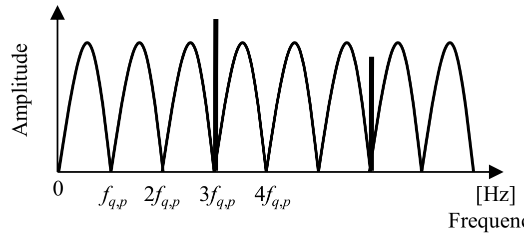

Fig4-B The PCF method often fails to estimate

III. An Enhanced PCF method

This section shows an enhanced PCF method (EPCF method). Fig.5 shows the flow of the process which has 5steps. In this paper the second zero point's frequency of comb filters is defined as fundamental frequency of comb filters. Fundamental frequency of the comb filter $H^k_{q,p}(z)$ is k times fundamental frequency of $H_{q,p}(z)$ (k is integer). The output of comb filters $H^k_{q,p}(z)$ is represented as $y^k_{q',p'}(n)$.

Step.1 Estimate temporary note numbers $\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_r$ by the PCF Step.1~4.

Step.2 Extract the targeted single note $\tilde{p}_1$ by using comb filters $H_{2,\tilde{p}_2}(z), \dots, H_{2,\tilde{p}_r}(z)$.

Step.3 Eight comb filters $H^k{}_{2,\tilde{p}_1}(z)$ (k=1,…,8) detect zero output.

Step.4 Select the number k which becomes the maximum in zero outputs. Then the first octave number and the first note number are fixed at Q_1 and P_1.

Step.5 The second or more note number and octave number are estimated by repeating the same process of the EPCF Step.2~4 about the targeted single note numbers $\tilde{p}_2, \dots, \tilde{p}_r$.

The input of octave 6 or more is not target in this algorithm. And when k=9 or more comb filters have zero points of different position from the frequency of the equal temperament scale.

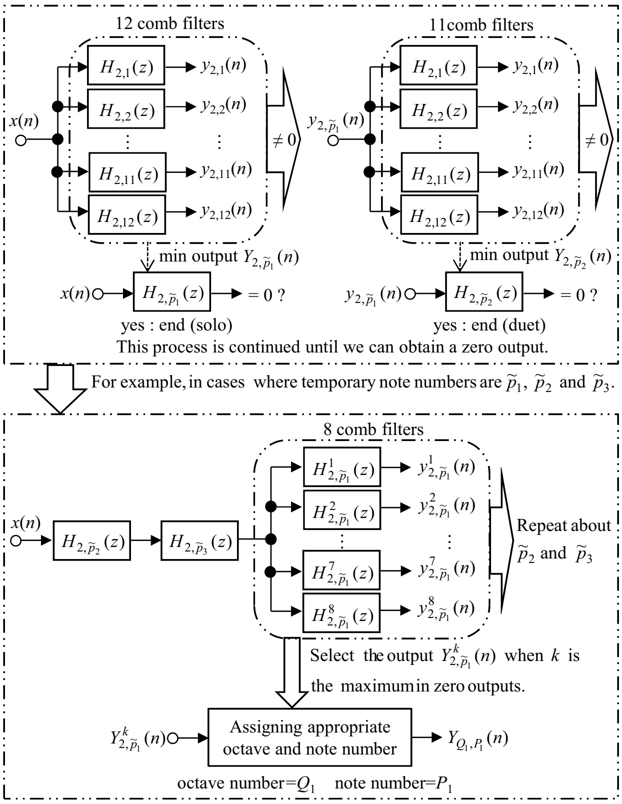

Fig.5 The enhanced PCF algorithm for the pitch estimation

IV. SIMULATION

This section compares results of EPCF method to that of the PCF method. Fig.6-A and Fig.6-B show the results by the PCF method, and Fig.7-A and Fig.7-B show those by EPCF method. Fig.6-A and Fig.7-A show the results of the example A (TABLE I). Fig.6-B and Fig.7-B show those of the example B (TABLE I). Sampling frequency *fs* is 16(kHz). Order of fractional delay comb filters N is 10. Input signals $x(n)$ are made on the computer with sin waves and the number of concurrent sound is three. M-th order harmonics has $1/M$ (M=1,…,10) as much amplitude as the fundamental frequency. Length of one sound is 2.0(seconds). Both the window size and the window shift size are 0.128(seconds). The window function is the rectangular window. TABLE II shows the accuracy of the estimation as follows:

$$\text{the estimation accuracy} = \frac{\text{(correct plotted points)}}{\text{(all plotted points)}} \qquad (9)$$

TABLE I
Description of the input signal $x(n)$

Example	octave number	Triadic input signal $x(n)$		
Example A	octave2	C+E+G	D+F+A	D+E+B
	octave3	C+A+B	D+E+F	D+F+A
Example B	octave4	C+E+G	C+D+F	C+D+B
	octave5	D+E+B	D+F+A	C+D+A

Square plotted points show the correct results. Black circular plotted points show the results of the simulation (Fig.6, Fig.7).

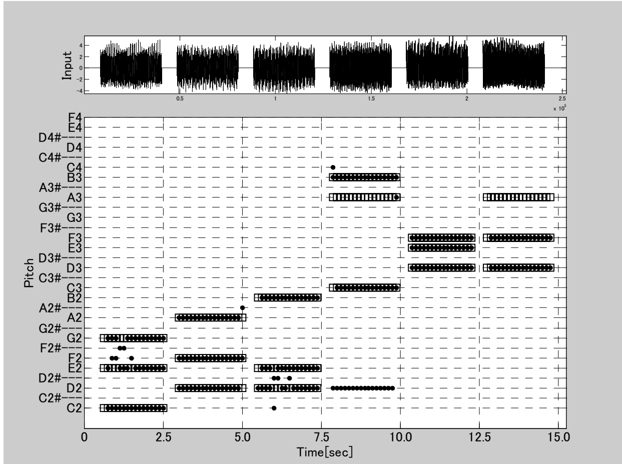

Fig.6-A The result of the pitch estimation by the PCF method

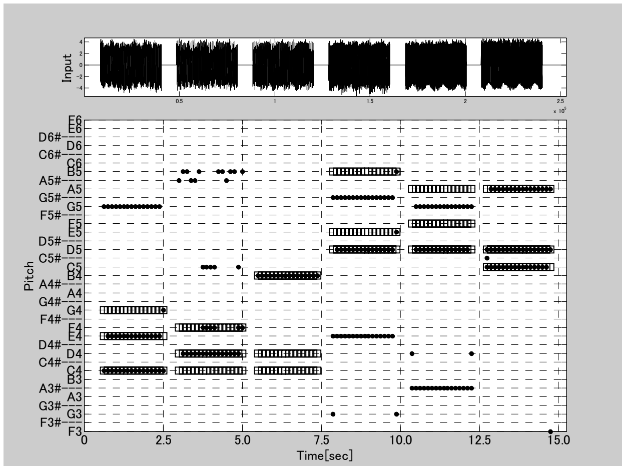

Fig.6-B The result of the pitch estimation by the PCF method

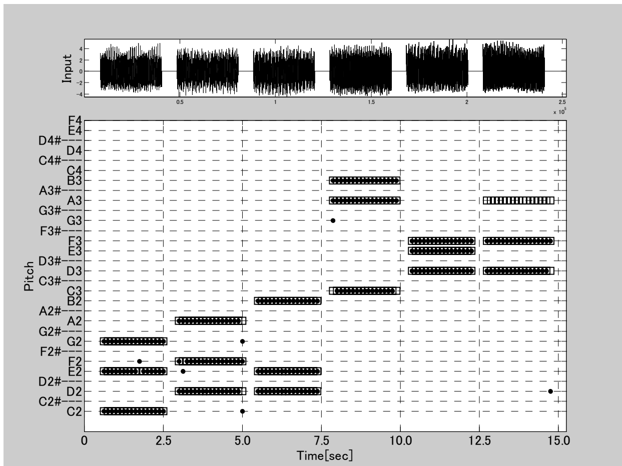

Fig.7-A The result of the pitch estimation by the EPCF method

International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:12, 2009

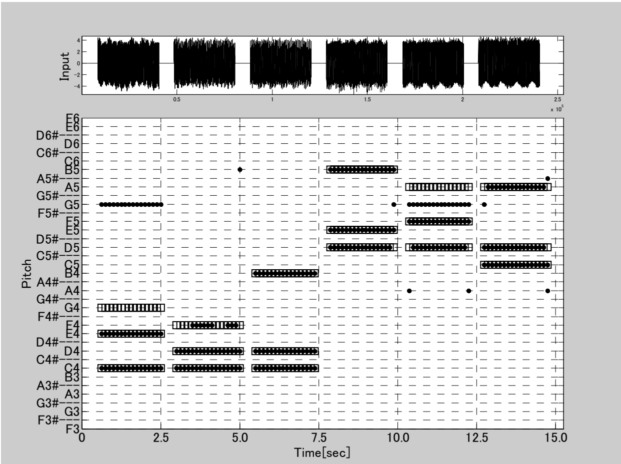

Fig.7-B The result of the pitch estimation by the EPCF method

TABLE II
The accuracy of each method

	PCF	EPCF
Example A	82%	92%
Example B	49%	84%

V. Conclusions

In this paper we have described how to improve a method called PCF which estimates multiple pitch with comb filters. We have shown the simulation results with the enhanced parallel-connected comb filter method for multiple pitch estimation. The proposed method has been able to improve the accuracy more than the conventional method. Future tasks are further improvements of the accuracy and are to make the system which estimates the pitch of real music. Nonetheless, our method will expand possibilities for music transcription with comb filters.

References

[1] T. Morita, M. Yamaguchi and Y. Tadokoro, "Pitch Estimation Method for Transcription Using Parallel-Connected Comb Filters and Noticing Their Output Values," *Journal of IEICE*, D-II, Vol.J87-D-II, No.12, pp.2271-2279, Dec. 2004.

[2] T. Saito, T. Matsui, H. Honda and Y. Tadokoro, "Real-time Realization of Scale Detection Based on Comb Filters Using DSPs," *Journal of SICE,* Vol. 34, pp.504-509, 1998.

[3] H. Kameoka, T. Nishimoto and S. Sagayama, "Separation of Harmonic Structures Based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds," *ICASSP '04*, pp.297-300, 2004.

[4] H. Kameoka, T. Nishimoto and S. Sagayama, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Tras. On Audio*, *Speech*, *and Language Processing*, Vol. 15, No.3, pp. 982-994, 2007.

[5] K. Miyamoto, H. Kameoka, T. Nishimoto, N. Ono and S. Sagayama, "Harmonic-Temporal-Timbral Clustering (HTTC) for the analysis of multi-instrument polyphonic music signal," *ICASSP '08*, pp. 113-116, 2008

[6] K. Kashino, T. Kinoshita, K. Nakadai and H. Tanaka, "Chord Recognition Mechanisms in the OPTIMA Processing Architecture for Music Scene Analysis," *The transactions of the Institute of Electronics, Information and Communication Engineers*, J79-D-2, 1996, 1762-1770 (in Japanese).

[7] M. Goto, "A predominant-F0 estimation method for CD recordings: MAP estimation using EM algorithm for adaptive tone models," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* Salt Lake City, UT, 2001, 3365-3368

[8] Anssi P. Klapuri, "Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness," *IEEE Trans. on Speech and Audio Processing*, Vol.11, No.6, 804-816, 2003.

[9] M. Goto and K. Hirata, "Recent Studies on Music Information Processing," *Acoustical Science and Technology*, 25(4), pp. 419-425, 2004.

[10] T. Matsuno, M. Fujieda and Y. Ishida, "Pitch Estimation Using Comb Filter with Fractional Delay," *Proc. IEICI, Fundamentals of Electronics, Communications and Computer Sciences*, 2009, 196 p., (in Japanese)

[11] Timo I. Laakso, V. Valimaki, M. Karjalainen and Unto K. Laine "Splitting the Unit Delay," *IEEE, Signal Processing Magazine*, pp. 30-60, 1996.

Taro matsuno was born in Tokyo, Japan, on July 18, 1984. He received the B.E. degree in Department of Science and Technology form Meiji University, Kawasaki, Japan, in 2008. He is currently working toward the M.E. degree at Graduate School of Electrical Engineering, Meiji University. His research interests include signal processing. He is member of the **IEICE** of Japan.

Yuta Otani was born in Japan, 1988. He is currently working toward the B.E. degree in Department of Electronics and Bioinformatics, school of Science and Technology form Meiji University, Kawasaki, Japan. His research interests include signal processing.

Ryo Tanaka was born in Japan, 1987. He is currently working toward the B.E. degree in Department of Electronics and Bioinformatics, school of Science and Technology form Meiji University, Kawasaki, Japan. His research interests include signal processing.

Kaori Ikezaki was born in Japan, 1987. He is currently working toward the B.E. degree in Department of Electronics and Bioinformatics, school of Science and Technology form Meiji University, Kawasaki, Japan. His research interests include signal processing.

Hitoshi Yamamoto was born in Japan, 1983. He received the B.E. degree in Department of Science and Technology form Meiji University, Kawasaki, Japan, in 2007. He is currently working toward the M.E. degree at Graduate School of Electrical Engineering, Meiji University. His research interests include signal processing.

Masaru Fujieda was born in Tokyo, Japan, 1983. He received the B.E. degree in Department of Science and Technology form Meiji University, Kawasaki, Japan, in 2006 and the M.E. degree in Graduate School of Electrical Engineering, Meiji University, in 2008. He is currently working toward the Ph.D. at Graduate School of Electrical Engineering. His research interests include signal processing. He is member of the **IEEE**, and the **IEICE** of Japan.

Yoshihisa Ishida was born in Tokyo, Japan, on February 24, 1947. He received the B. E., M. E., and Dr. Eng. Degrees in Electrical Engineering, Meiji University, Kawasaki, Japan, in 1970, 1972, and 1978 respectively. In 1975 he joined the Department of Electrical Engineering, Meiji University, as a Research Assistant and became a Lecturer and an Associate Professor in 1978 and 1981, respectively. He is currently a Professor at the Department of Electronics and Bioinformatics, Meiji University. His current research interests include signal processing, speech analysis and recognition, and digital control. He is a member of the **IEEE**, and the **IEICE** of Japan.