PITCH ESTIMATION OF DIFFICULT POLYPHONY SOUNDS OVERLAPPING SOME FREQUENCY COMPONENTS

Yoshiaki Tadokoro, Masanori Natsui, Yasuhiro Seto
Dept. of Information and Computer Sciences, Toyohashi University of Technology,Toyohashi,441-8580 Japan

Michiru Yamaguchi
Dept. of Computer Engineering, Toyama National College of Maritime of Technology, Imizu, 933-0293 Japan

Keywords: pitch estimation, polyphony, unison, octave different tone, three-times tone, transcription.

Abstract: There are some difficult polyphony to estimate these pitches for transcription. This paper proposes two new methods for the pitch estimation of these difficult polyphony. One of them is based on the beat components of the polyphony analyzed by the short-time Fourier transform (STFT). The other is a method noticing the period of the residual signal after the elimination of polyphony components using a comb filter ($H(z) = 1 - z^{-N}$). These methods are based on the fact that there is a small frequency difference between the real sound and the ideal one.

1 INTRODUCTION

Musical transcription is to produce scores from musical sounds and is necessary in the musical field, musical retrieval and also a significant problem in artificial intelligence (Roads, 1985), (Sterian and Wakefield, 2000), (Pollastri, 2002). In the transcription, the pitch estimation is most important and many studies have been done (Roads, 1996), (Tadokoro et al, 2001, 2002, 2003). We also proposed a unique method of the pitch estimation that is based on the elimination of the pitch and its harmonic components using the cascade or parallel connections of the comb filters (Tadokoro et al, 2001, 2002, 2003). But there is a difficult problem in the pitch estimation that has not been solved clearly. That is the pitch estimation of polyphony where the frequency components of each tone overlap completely or partly. For this problem, the methods based on the musical rules (Katano et al, 1996), the assumption of power spectra addition (Ueda and Hashimoto, 1997) and the genetic algorithm (Ono et al, 1997) have been proposed, but they have some problems for the practical use.

Figure 1 shows the spectra of a piano sound (C_3 : tone name C in octave 3) and a violin sound (G_3). Showing in Fig.1, a musical sound is composed of a basic frequency f_p (pitch) and its harmonic components nf_p . The frequency ratio between adjacent tones in the equal temperament of 12 degrees is $2^{1/12}$. And so it is occurred that the

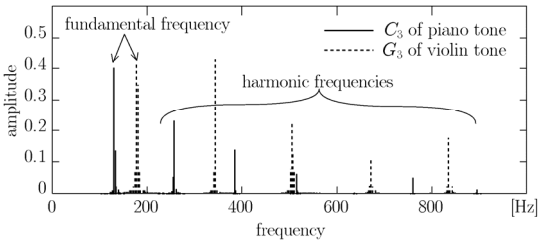


Figure 1: Examples of magnitude spectrum of musical sounds.

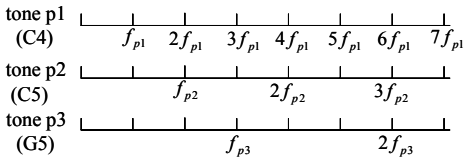


Figure 2: Overlap relation of frequency components of difficult polyphony.

Tadokoro Y., Natsui M., Yamaguchi M. and Seto Y. (2006).
PITCH ESTIMATION OF DIFFICULT POLYPHONY SOUNDS OVERLAPPING SOME FREQUENCY COMPONENTS.
In *Proceedings of the Third International Conference on Informatics in Control, Automation and Robotics*, pages 168-173
DOI: 10.5220/0001206101680173

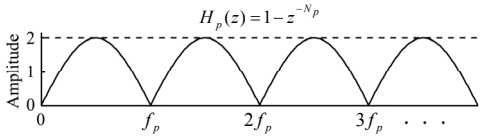


Figure 3: Magnitude characteristic of the comb filter.

frequency components of some polyphony overlap as demonstrated in Fig.2. We have proposed the new pitch estimation method using comb filters as shown in Fig.3. The comb filter to eliminate the tone $p1$ also eliminates the tones $p2$ and $p3$ in Fig.2, and so we cannot estimate the pitches of tones $p2$ and $p3$. The pitch estimation for these polyphony is also an unsolved problem in other pitch estimation methods.

This paper presents two new methods for the pitch estimation of difficult polyphony. One of them is a method based on the beat signals of polyphony frequency components analyzed by the short-time Fourier transform (STFT) and the starting point difference of each tone. Actual polyphony has a small time difference between the starting points of each tone and a small frequency difference between each frequency component, but ideally these start points and frequency components are same. From these starting point difference and the beat signals by the frequency difference, we can estimate the pitches of the difficult polyphony. The other is a method based on the period measurement of the residual signal that is an output signal of the comb filter to eliminate the polyphony frequency components. The residual signal is occurred by the frequency difference between an ideal and a real musical sounds. From the periods of the residual signal, we can obtain the clues of the pitches of the difficult polyphony.

In this paper, we assume that the polyphony is composed of two tones in the octave 4 and 5 of which lower pitch has already estimated by the pitch estimation method described in section 2. And the input polyphony is made from real sounds of the RWC music database (the Real World Computing Partnership in Japan). Each tone has almost same amplitude. The sampling frequency is $f_s = 44.1kHz$.

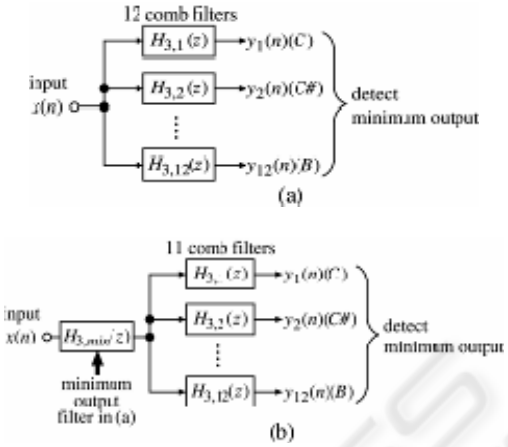


Figure 4: Pitch estimation system using parallel connected comb filters and minimum output.

2 PITCH ESTIMATION METHOD USING COMB FILTERS

Figure 3 shows a frequency characteristic of a comb filter ($H_p(z) = 1 - z^{-N_p}$). The comb filter has zero points at $(f_s / N)n$ where f_s is a sampling frequency. Using a comb filter, we can eliminate all components of one tone of which basic frequency f_p (pitch frequency) is f_s / N. The feature of our pitch estimation method is to eliminate the frequency components of a musical sound using these comb filters. The conventional pitch estimation methods, on the other hand, are based on the extraction of the pitch frequencies.

Figure 4 shows one of the pitch estimation systems that have been proposed by us. This pitch estimation system is composed of twelve comb filters connected in parallel where the lowest notch frequency (zero point) of each comb filter is corresponding to each tone's basic frequency in one octave. If the input sound is monophony, then we can estimate the pitch of its sound by knowing the zero output of the parallel connected twelve comb filters. When the input sound is polyphony, we can detect the pitches by knowing the minimum output of the parallel connected comb filters and then connecting its comb filter to other parallel connected eleven comb filters in cascade and detecting the minimum output of the eleven comb filters as shown in Fig.4. But if the polyphony is difficult one showing in Fig.2, we cannot estimate the pitches other than the lowest pitch, because other tones are eliminated by the comb filter corresponding to the lowest pitch. The pitch estimation method for these

difficult polyphony has not been developed also by other pitch estimation methods.

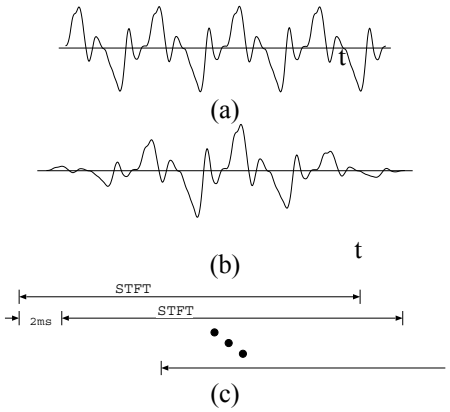


Figure 5: Calculations of STFTs, (a) four periods data, (b) hamming windowed data, (c) STFTs every 2 ms.

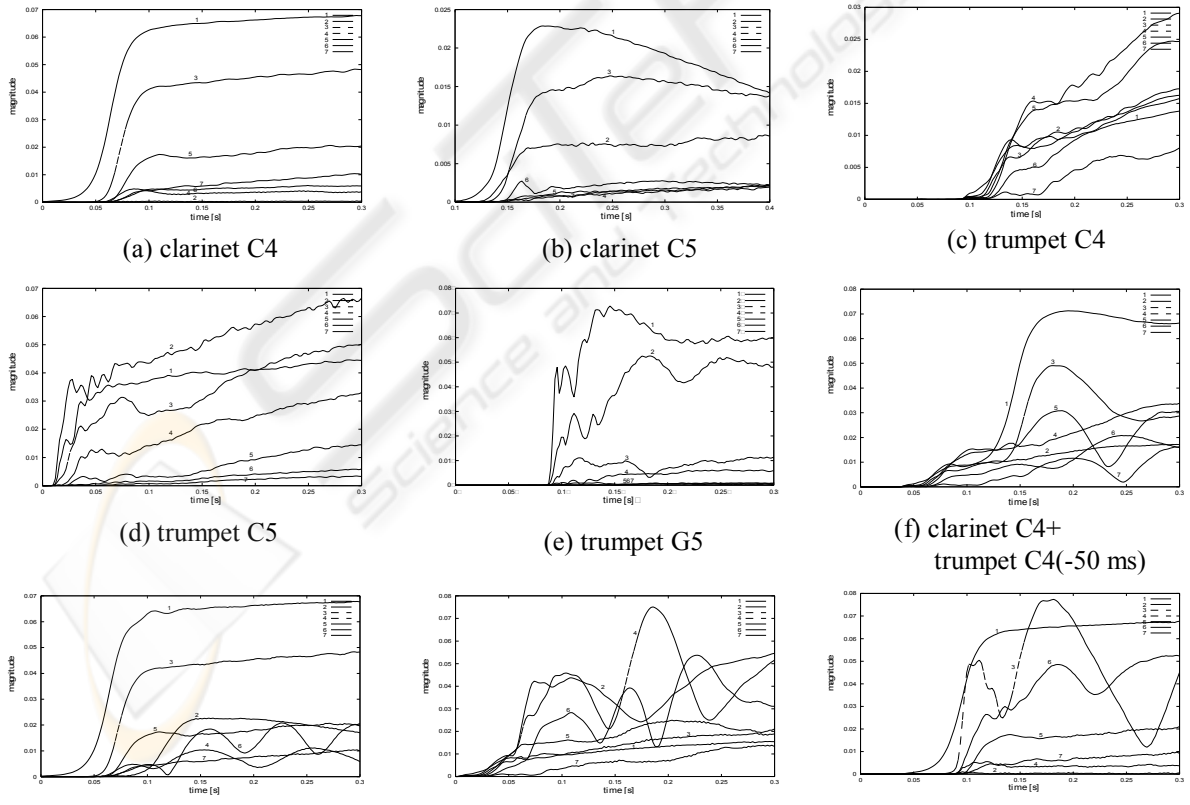
3 SHORT-TIME FOURIER TRANSFORM (STFT) METHOD

3.1 Calculations of STFTs and Their Results

We calculate the STFTs for four periods data of a basic component of a musical sound windowed by the hamming window every 2 ms as shown in Fig.5. Figure 6 shows the time changes of the magnitude characteristics of each frequency component of some musical sounds, where +50 ms or -50 ms means that its tone starts at 50ms after or before the clarinet C4 tone starts, respectively.

3.2 Consideration of Pitch Estimation

As mentioned in section 1, we assume that polyphony is composed of two tones and we know the pitch of the lower tone. From the results in Fig.6, we want to discriminate the following four tones, that is, (1)



(a) clarinet C4 (b) clarinet C5 (c) trumpet C4

(d) trumpet C5 (e) trumpet G5 (f) clarinet C4+ trumpet C4(-50 ms)

(g) clarinet C4 + clarinet C5 (+50 ms) (h) trumpet C4 + trumpet C5 (+50 ms) (i) clarinet C4+ trumpet G5 (+50 ms)

Figure 6: Magnitude spectrum components by STFT.

monophony (C4) or polyphony (C4 + α), (2) polyphony (C4 + C4 : unison), (3) polyphony (C4+C5) :octave), (4) polyphony (C4+G5:three-times tone), where we assume the lower tone is C4 and we denote these tones by (2) unison, (3) octave and (4) three-times tone.

First, we must discriminate if the input sound is monophony or polyphony. If we detect the beat components, then we know the sound may be polyphony. But there are some monophony with beat components like trumpet G5 in Fig.6 (e). Another method to determine if the sound is monophony or polyphony is to notice the time difference of the starting points. For example, we can detect the difference of the starting points in the case of Fig.6 (f).

Next, we must determine if the tone is unison, octave or three-times tone. If we detect that all most components of the sound are beating, we can know the sound may be a union, for example in the case of Fig.6 (f). If even components are beating like Fig.6 (g) and (h), we can know the sound may be an octave. If harmonic components of the third component are beating like Fig.6 (i), we can determine the sound may be a three-times tone.

The pitch estimation method by the beat signals uses some measurement time, about 100 or 200ms. This is a problem to estimate the pitches for shorter sounds.

4 COMB FILTER METHOD

In this method, we process an input sound using a comb filter and the sample data of 2000 (45 ms) to 3000 (68 ms) from the starting point of the input sound. For simplicity, we assume that the lower tone of polyphony is a C4 tone. In this case, we must discriminate the following four tones, (1) monophony C4 or polyphony, (2) polyphony (C4+C4: unison), (3) polyphony (C4+C5:octave), (4) polyphony (C4+G5: three-times tone).

First, an input sound is passed by a comb filter C4. The comb filter C4 means the filter $H_p(z) = 1 - z^{-N_p}$ where p =C4 and $N_p = [f_s / f_p] = [44.1kHz / 261.62\ Hz] = 168$. Ideally the comb filter C4 can eliminate all above four tones, i.e., monophony, unison, octave and three times tone. But we can obtain a small output signal caused by some frequency difference from ideal frequencies. Next, we measure the periods of the output signal of the comb filter C4. From these periods, we can get the clues to discriminate the above four tones.

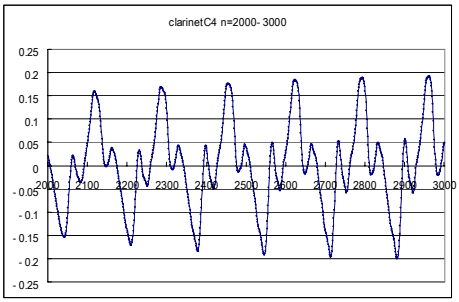
Figure 7 shows the input and output waveforms (sample number, n=2000-3000) of the comb filter C4: input ((a)-(d)) and output ((e)-(i)).

First we use the comb filter C4 of $N_{p0} = 168$ that is a sample number determined from $f_s / f_p = [44100Hz / 261.62Hz]$. When the monophony C4 in Fig.7(a) is filtered by the comb filter C4 of $N_{p0} = 168$, we obtain the output signal in Fig.7(e) of which amplitude is decreased by the factor of 0.04 from the input one. Next we measure the period of the comb filter output signal (Fig.7(e)) and obtain the period of $N_{p1} = 166$. Then we filter the input sound again by the comb filter C4 of $N_{p1} = 166$ and this time we measure the period to be $N_{p2} = 167$ .When we pass the input sound through the comb filter C4 of $N_{p2} = 167$, we obtain the filter output signal having its period $N_{p3} = 166$. These waveforms of the output signals in the comb filters of $N_{p2} = 167$ and $N_{p3} = 166$ are almost same and so we determine that the input sound is monophony C4.

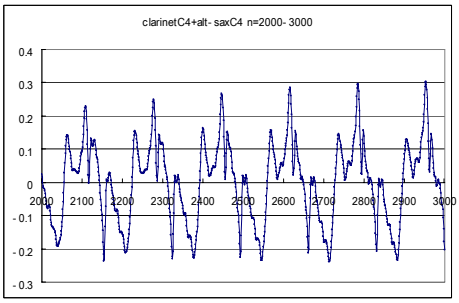
When the input sound of Fig.7(b) is filtered by the comb filter C4 of $N_{p0} = 168$, we obtain the filter output signal in Fig.7(f) of which amplitude is decreased by the factor of 0.16. From the output signal in Fig.7(f), we measure the period to be $N_{p1} = 170$. Then we filter the input signal in Fig.7(b) by the comb filter C4 of $N_{p1} = 170$ and we obtain the output signal in Fig.7(g) with the period of $N_{p2} = 168$ (or 167). The waveforms of Fig.7(f) and (g) are different and so we determine that the input sound of Fig.7 (b) is polyphony (C4+C4: unison).

Above we showed an example to discriminate the monophony and polyphony in the comb filter method. But we think that a more effective method using a comb filter is to use two input sounds obtained from two different points and measure the amplitude ratio of the output/input signals of the comb filter C5 or G5. If the input sound is monophony, then the output/input ratio does not change for two input sounds. If the input sound is polyphony, the waveforms of two input sounds change for the phase relation of two tones and so the output/input ratio also change. We can determine if an input sound is monophony or polyphony by noticing the change of the output/input ratio for two input sounds.

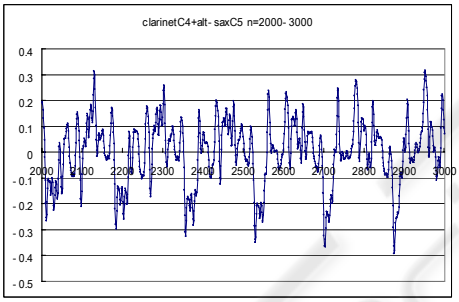
Next we consider the polyphony of octave and three-times tone. When the input sound of Fig.7(c) is passed through the comb filter C4 of $N_{p0} = 168$, we



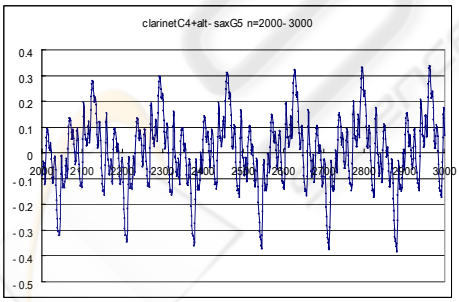
(a) input:clarinet C4



(b) input:clarinet C4+alt-sax C4



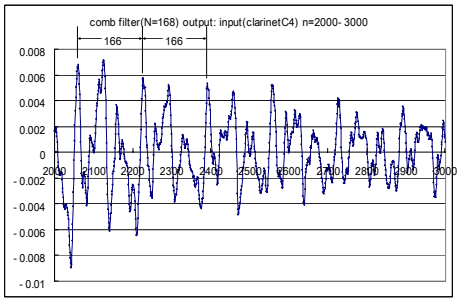
(c) input:clarinet C4+alt-sax C5



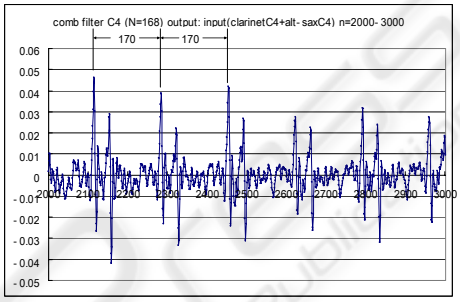
(d) input:clarinet C4+alt-sax G5

Figure 7.1: Input waveforms for comb filter C4.

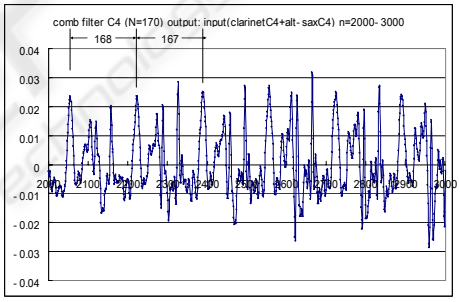
obtain the output signal of Fig.7 (h) of which amplitude is decreased by the factor 0.5 and we can measure the period of $N_{p1} = 86(87)$. The period



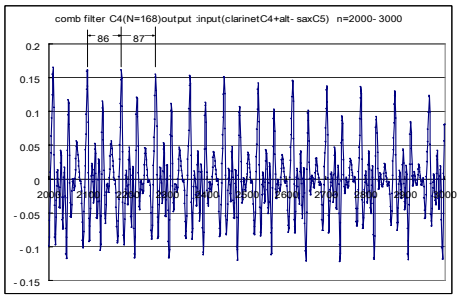
(e) comb filter (N=168) output:input (clarinet C4)



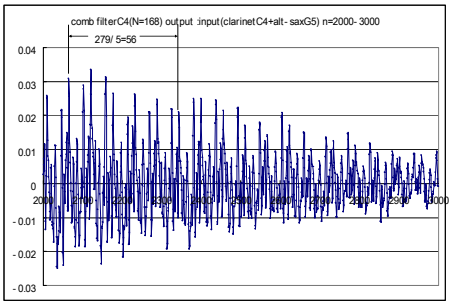
(f) comb filter (N=168) output :input (clarinet C4+ alt-sax C4)



(g) comb filter (N=170) output :input (clarinet C4+alt-sax C4)



(h) comb filter (N=168) output :input(clarinet C4+alt-sax C5)



(i) comb filter (N=168) output:input (clarinet C4 +alt-sax G5)

Figure 7.2: Output waveforms of comb filter C4.

of $N_{p1} = 86$ corresponds to the basic period of the C5 tone and we can determine that the input sound is polyphony of the octave (C4+C5). In this case, we know that the basic frequency of the C5 tone is different from nf_p with some extend, because the output amplitude of the comb filter is not decreased largely.

When we pass the input sound of Fig.7 (d) through the comb filter of $N_{p0} = 168$, we get the output signal of Fig.7 (i) of which amplitude is decreased by the factor of 0.1. From the output signal of Fig.7 (i), we measure the period of $N_{p1} = 56$ and its period corresponds to the G5 tone. We can clearly measure the period of the output signal using the autocorrelation function as shown in Fig.8. Then we can determine that the input sound is polyphony of the three-times tone (C4+G5).

By the method mentioned above, we can discriminate the difficult four tones (monophony or polyphony, unison, octave and three-times tone).

5 CONCLUSIONS

We proposed two new methods to estimate the pitches of the difficult polyphony where all or some frequency components of the tones overlap, i.e. unison, octave and three-times tones. One of them is the method using the beat signals of the spectrum components analyzed by the STFT that are happened for a small frequency difference between two tones. This method has some measurement time of about 100 or 200 ms for the detection of the beat signals. The other is the method using a comb filter. We can obtain a small output signal of the comb filter for a small frequency difference between the ideal and real tones. We can discriminate these difficult four

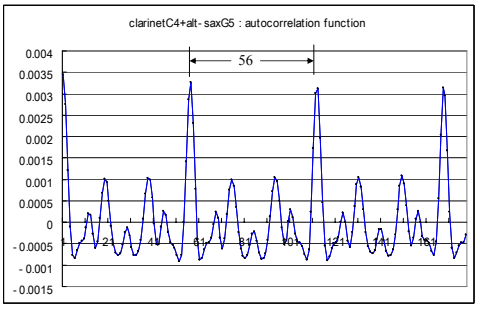


Figure 8: Autocorrelation function of Fig.7-2 (i).

tones by the measurements of the periods of the filter output signals. The measurement time of the comb filter method is about 50 ms.

As a feature research, we want to test the proposed two methods for many musical sound data.

REFERENCES

Kashino,K., Kinoshita,T., Nakadai,K. and Tanaka,H., 1996. "Chord recognition mechanisms in the OPTIMA processing architecture for music scene analysis," Trans. IEICE of Japan, vol.J79-D-II, no.11, pp.1771-1781.

Ono, T., Saito,H. and Ozawa,S., 1997. "Mixed tones estimation for transcription using GA," Trans.SICE of Japan, vol.33, no.5, pp.417-423.

Pollastri,E., 2002. "A pitch tracking system dedicated to process singing voice for musical retrieval," Proc. of IEEE Int. Conf. on Multimedia and Xpo, ICME2002.

Roads,C.,1985. "Research in music and artificial intelligence," ACM computing Survey, vol.17, no.2, pp.163-190.

Roads,C.,1996. "The Computer Music Tutorial," MIT Press.

Sterian,A., Wakefield,G.H.,2000. "Musical transcription system : From sound to symbol," Proc. AAAI-2000 Workshop.

Tadokoro,Y.and Yamaguchi,M., 2001. "Pitch detection of duet song using double comb filters," Proc. of ECCTD'01, I, pp.57-60.

Tadokoro, Y., Matsumoto, W. and Yamaguchi,M.,2002. "Pitch detection of musical sounds using adaptive comb filters controlled by time delay," ICME2002, P03.

Tadokoro,Y., Morita, T. and Yamaguchi,M., 2003. "Pitch detection of musical sounds noticing minimum output of parallel connected comb filters, " IEEE TENCON2003, tencon-072.

Ueda,M. and Hashimoto,S., 1997. "Blind decomposition alagorithm for the sound separation," Trans. IPS of Japan, vol.38, no.1, pp.146-157.