# Pitch Estimation for Musical Sound Including Percussion Sound Using Comb Filters and Autocorrelation Function

YOSHIAKI TADOKORO , TSUTOMU SAITO , YUSUKE SUGA and MASANORI NATSUI*
Dept. of Inf. and Computer Sciences, Dept. of Electrical and Electronic Engineering
Toyohashi University of Technology, Toyota National College of Technology
Tempaku 1-1, Hibarigaoka, Toyohashi, Eisei-cho 2-1, Toyota
JAPAN
http://www.signal.ics.tut.ac.jp

*Abstract:* - This paper proposes a new pitch estimation method using comb filters $H(z)=1-z^{-N}$ and autocorrelation functions (ACFs) for the musical sounds including a percussion sound. We can obtain the ACF of the percussion sound at the comb filter output where the instrument sound with pitch is eliminated. By subtracting this percussion ACF from the ACF of the adjacent comb filter output in the twelve comb filters connected in parallel corresponding to each pitch of one octave, we can also obtain the ACF of the instrument one, and then we can estimate the pitch of the instrument sound from the instrument ACF.

*Key-Words:* - Pitch estimation, Comb filter, Autocorrelation function, Percussion sound

## 1 Introduction

Musical transcription is necessary in the many musical studies, musical retrieval and also a significant problem in artificial intelligence [1]-[3]. In the transcription, the pitch estimation is most important and many studies have been done [4]-[6]. Recent pitch estimation methods use statistic signal processing, auditory scene analysis and sophisticated algorithms [5][6]. On the other hand, we proposed a unique method of the pitch estimation that is based on the elimination of the pitch (fundamental frequency) and its harmonic components using twelve comb filters ($H(z)=1-z^{-N}$) connected in parallel [4]. This comb filter method is simple and one of old methods. In this pitch estimation algorithm, we notice the minimum output in twelve comb filters. However, when the input sound includes percussion sounds, we cannot obtain a good performance by the parallel connected comb filter method.

In this paper, we propose a new pitch estimation method using comb filters and autocorrelation functions (ACFs) for the musical sound including a percussion sound, where we treat percussions without pitches and musical instruments with pitches. This time, we notice the fact that the ACF of the comb filter output eliminated the instrument sound is not periodic, and we can obtain the ACF of the comb filter output for the percussion sound only. Using this ACF of the percussion sound, we can also obtain the ACF of the comb filter output for the instrument sound having a pitch and then we can estimate the pitch of the instrument sound.

We consider four instruments, alto-sax(AS), flute (FL), trumpet (TR), and pipe organ (OR), and six percussions, bass drum (BD), snare drum (SD), low-tom (LT), middle-tom (MT), high-tom (HT) and crush cymbal (CC) that are used in jazz and popular music. We also assume that the instrument sound having a pitch is monophony and in octave 3 to 5, 36 tones. The input sound is made of real sounds of the RWC music database (the Real World Computing Partnership in Japan) [7]. The sampling frequency is $f_s = 44.1\,kHz$ .

## 2 Principle of the Proposed Method

Figure 1 shows the spectrum of the alt-sax C3 (C tone in octave 3) with pitch and six percussion sounds without pitches. To estimate the pitch, we use twelve comb filters ($H(z)_q = 1-z^{-N_q}$) connected in parallel shown in Fig.2 [4] and calculate the ACF of each output of the comb filters. The number $N_q$ of the delay elements of the comb filter $q$ is determined by

$$N_q = [f_s / f_q], \quad [\ ]: \text{integer by rounding}, \qquad (1)$$

where $f_q\ (q=1,2,\cdots,12)$ is the fundamental frequency corresponding to each tone in octave 3 when we treat musical sounds in octave 3 to 5.

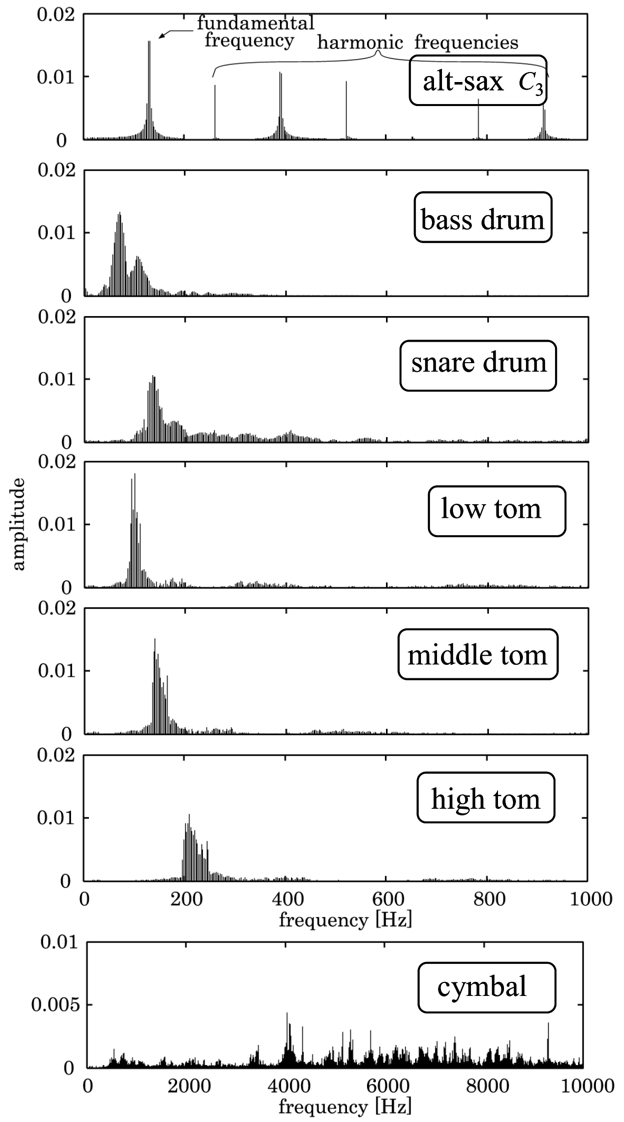



Fig.1 Spectrum of alt-sax sound ($C_3$) with pitch and six percussion sounds without pitches.

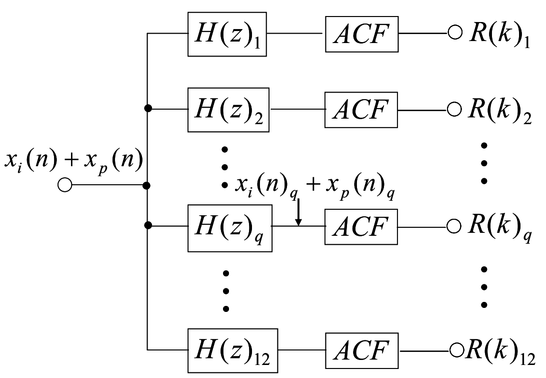
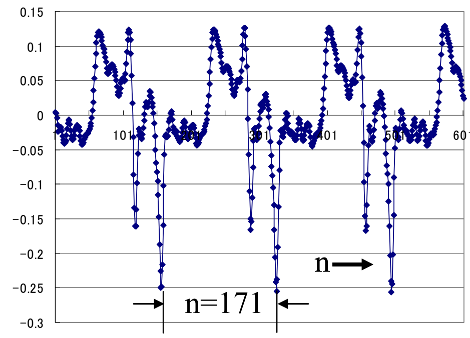
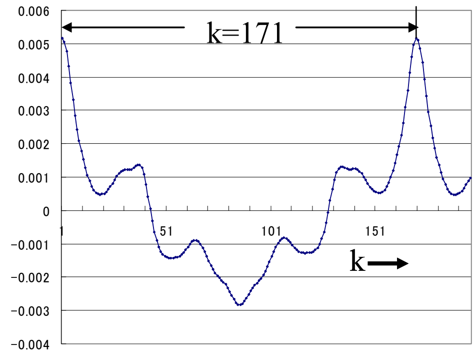



Fig.2 Pitch estimation system using comb filters and autocorrelation functions (ACFs).




(a) waveform of alt-sax C4




(b) ACF of alt-sax C4

Fig.3 (a) Waveform of alt-sax C4 and (b) its ACF

## 2.1 Proposed pitch estimation method

The waveform of alt-sax C4 and its autocorrelation function (ACF) is shown in Fig.3. That is, the period of the ACF corresponds to the pitch of the instrument sound. The ACF of the output ($x_i(n)_q + x_p(n)_q$) of the comb filter $q$ in Fig.2 can be written by

$$
\begin{aligned}
R(k)_q &= \frac{1}{N}\sum_{n=0}^{N-1}(x_i(n)_q + x_p(n)_q)(x_i(n+k)_q + x_p(n+k)_q) \\
&= \frac{1}{N}\sum_{n=0}^{N-1} x_i(n)_q x_i(n+k)_q + \frac{1}{N}\sum_{n=0}^{N-1} x_p(n)_q x_p(n+k)_q \\
&\quad + \frac{1}{N}\sum_{n=0}^{N-1} x_i(n)_q x_p(n+k)_q + \frac{1}{N}\sum_{n=0}^{N-1} x_p(n)_q x_i(n+k)_q \\
&= R_i(k)_q + R_p(k)_q + R_{ip}(k)_q + R_{pi}(k)_q \\
&\approx R_i(k)_q + R_p(k)_q \,, \qquad (2)
\end{aligned}
$$

where $R_i(k)_q$ and $R_p(k)_q$ are the ACFs of the $x_i(n)_q$ (instrument sound) and $x_p(n)_q$ (percussion sound), and $R_{ip}(k)_q$ and $R_{pi}(k)_q$ are the crosscorrelation functions (CCFs) of $x_i(n)_q$ and $x_p(n)_q$, respectively.

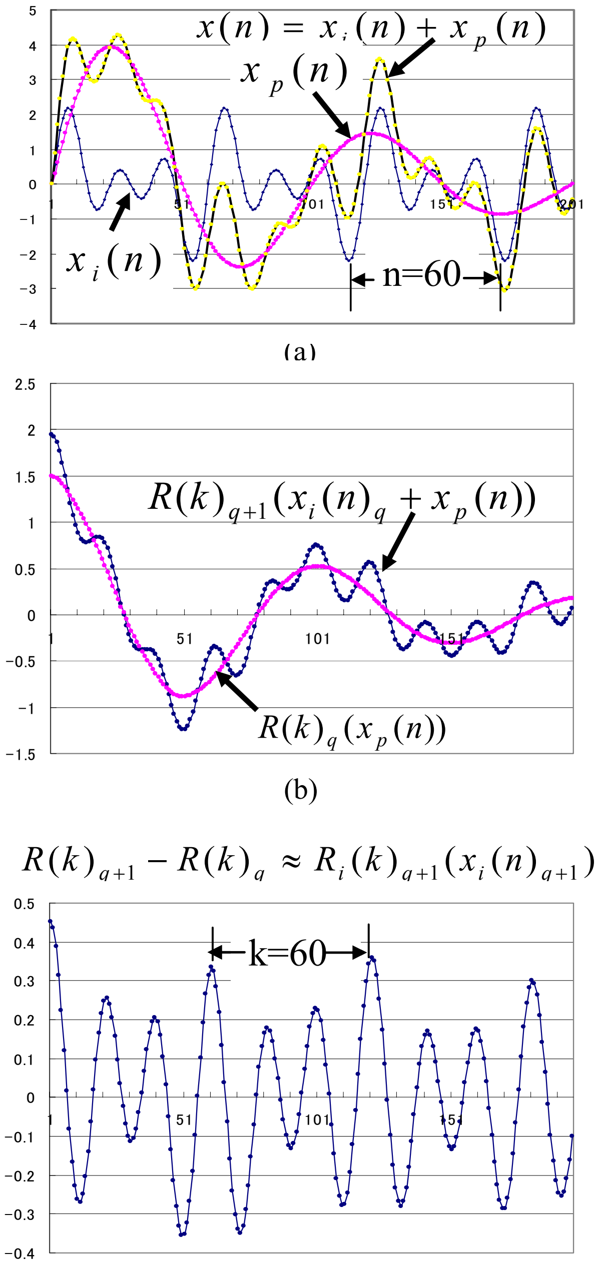



Fig.4 (a) Waveforms of $x_i(n), x_p(n)$ and $x(n) = x_i(n) + x_p(n)$ , (b) ACFs of $x_p(n)_q$ and $x_i(n)_q + x_p(n)_q$ ,and (c) ACF of $R_i(k)_{q+1}$ .

Generally, $R_{ip}(k)_q$ and $R_{pi}(k)_q$ are small comparing with $R_i(k)_q$ and $R_p(k)_q$ . If comb filter $q$ can eliminate the instrument sound, then $R(k)_q$ is represented by

$$R(k)_q \approx R_p(k)_q \qquad (3)$$

The ACF of the output of the comb filter ($q+1$) (or ($q-1$)) that is an adjacent comb filter of the comb filter $q$ is written by

$$R(k)_{q+1} \approx R_i(k)_{q+1} + R_p(k)_{q+1} \qquad (4)$$

If we can assume that $R_p(k)_{q+1} \approx R_p(k)_q$ satisfying $N_{q+1} \approx N_q$ , we can obtain the ACF of the instrument sound in the following equation:

$$R(k)_{q+1} - R(k)_q \approx (R_i(k)_{q+1} + R_p(k)_{q+1}) - R_p(k)_{q+1} = R_i(k)_{q+1} \qquad (5)$$

From $R_i(k)_{q+1}$ , we can estimate the pitch of the instrument sound.

## 2.2 Simulation of the proposed method

We assume an instrument sound $x_i(n)$ and a percussion sound $x_p(n)$ in the following form:

$$x_i(n) = 0.5\sin(2\pi f_i / f_s')n + \sin(4\pi f_i / f_s')n + \sin(6\pi f_i / f_s')n \qquad (6)$$

$$x_p(n) = 5(0.98)^n \sin(2\pi(f_p / f_s')n \quad , \qquad (7)$$

where $f_i = 100\,Hz, f_p = 60\,Hz$ and $f_s' = 6\,kHz$.

Figure 4 (a) shows the waveforms of $x_i(n)$ , $x_p(n)$ and $x(n) = x_i(n) + x_p(n)$ , where the power ratio of $\sum(x_i(n)^2 / x_p(n)^2)(n = 0-199) = 0.4$ . We pass the signal $x(n)$ through the comb filters $H(z)_q(N_q = 60)$ and $H(z)_{q+1}(N_{q+1} = 57)$ , where $H(z)_q$ eliminates $x_i(n)$, and calculate $R(k)_q$ and $R(k)_{q+1}$ of the each output of the comb filters $H(z)_q$ and $H(z)_{q+1}$ ( $(N_q - N_{q+1}) / N_q = 0.05$) by the following equations:

$$R(k)_q = \frac{1}{1000}\sum_{n=0}^{1000-1} x_p(n)_q x_p(n+k)_q \qquad (8)$$

$$R(k)_{q+1} = \frac{1}{1000}\sum_{n=0}^{1000-1} (x_i(n)_{q+1} + x_p(n)_{q+1}) \cdot (x_i(n+k)_{q+1} + x_p(n+k)_{q+1}) \qquad (9)$$

The ACFs of $R(k)_q$ ,and $R(k)_{q+1}$ are shown in Fig.4 (b). From Fig.4 (c), we can realize that the difference of $R(k)_{q+1}$ and $R(k)_q$ is almost the ACF of $x_i(n)_{q+1}$,

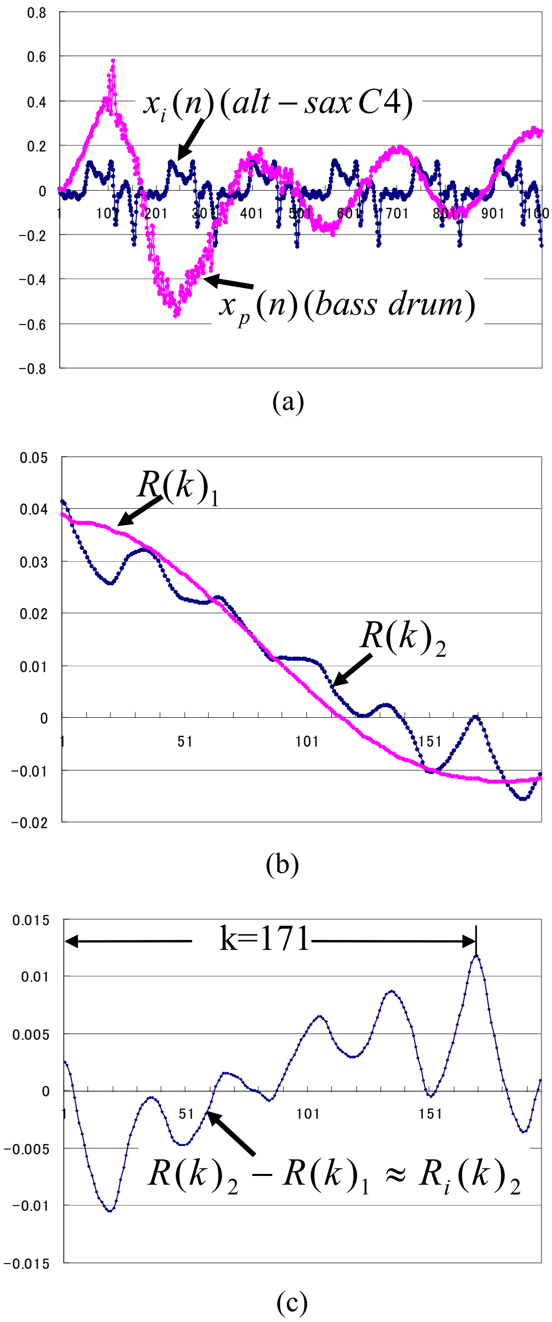



Fig.5 (a) Waveforms of $x_i(n), x_p(n)$ $x(n)=x_i(n)+x_p(n)$, (b) $R(k)_1\,(x_p(n)_1)$ and $R(k)_2$ $(x_i(n)_2+x_p(n)_2)$, and (c) $R(k)_2-R(k)_1 \approx R_i(k)_2$.

and then we can estimate the pitch period $(60\times(1/6000))=0.01\,\text{sec})$.

## 3 Experimental Results

In this experiment, we use 4 instruments (AS, FL, TR and OG) and 6 percussions (SD, BD, HT, MT, LT and CC). In Fig.2, first, we detect the pitch name from the minimum $R(0)_q$ or $R(k)_q$ not having a periodic component. Second, we estimate the pitch from the period of $(R(k)_{q+1}-R(k)_q)$, where $R(k)_q$ is the ACF of the comb filter output not including the instrument sound.

Figure 5 shows the effect of the proposed method using the real sounds of alt-sax C4 and bass drum, where the power ratio of the instrument sound to the percussion one is about 0.1 (-10dB) in the signal period of n=0-999 (1000 samples). We detect the pitch name C from the minimum $R(0)_1$. That is, $R(k)_1$ at the comb filter C3 shows the ACF of $x_p(n)_1$ (bass drum). On the other hand, $R(k)_2$ at the comb filter C#3 ( $(N_1-N_2)/N_1=(337-318/337=0.056$ ) includes the components of the alt-sax C4 and bass drum sounds. However, we cannot detect the period of the instrument sound in $R(k)_2$ causing the large percussion component. If we subtract $R(k)_1\,(x_p(n)_1)$ from $R(k)_2(x_i(n)_2+x_p(n)_2)$, we can detect the period of the instrument sound showing in Fig.5 (c). Then, we can estimate the pitch C4 of the instrument sound $(x_i(n)_1)$, i.e., the period=171 (the number of samples, $171\times(1/44.1kHz)=3.87\,ms \approx 1/261.62\,Hz)$.

Table 1 Detection results of the pitch name from the minimum $R(0)_q$ when power ratio of signal/noise=1/2 (%).

| | AS | FL | TR | OR |
|---|---|---|---|---|
| SD | 84.8 | 100.0 | 54.8 | 19.4 |
| BD | 100.0 | 100.0 | 74.1 | 88.8 |
| HT | 96.9 | 100.0 | 70.9 | 55.5 |
| MT | 96.9 | 100.0 | 70.9 | 52.7 |
| LT | 87.8 | 100.0 | 54.8 | 44.4 |
| CC | 93.9 | 100.0 | 51.6 | 36.1 |
| mean | 93.4 | 100.0 | 62.9 | 49.5 |

Table 1 shows the detection results of the comb filter corresponding to the pitch name from the minimum $R(0)_q$ when the power ratio of the instrument and percussion sounds is 1/2. Now we do not have good results for trumpet (TR) and organ (OR). However, we could estimate the pitch name with the accuracy of 93.5% using other type comb filter ( $H(z) = 1 + z^{-N}$ ). Concerning the pitch estimation from the pitch name, we have not done the pitch estimation for all the combinations of 4 instruments and 6 percussions, but we could obtain the pitch estimation with the accuracy of 100% for (AS+SD), (AS+BD), (OR+CC), and 61.1% for (OR+SD), 58.0% for (OR+LT).

## 4 Conclusion

We proposed a new pitch estimation method for monophony instrument sounds including a percussion one. We can obtain the ACF of the percussion sound from the comb filter output eliminating the instrument one. Using this ACF of the percussion sound, we can also obtain the ACF of the instrument one, and then we can estimate the pitch of the instrument sound. We could obtain the aspect of a good pitch estimation even if the power ratio of the instrument and percussion sounds is small. However, we have not done a satisfying experiment.

As a future work, we would like to improve the estimation accuracy of the pitch name for trumpet and organ and consider the pitch estimation for polyphony sounds including some percussion sounds.

*References:*

[1] A.P.Klapuri and M.Davy, *Signal Processing Methods for Musical Transcription,* Springer, 2006.

[2] A.Sterian and G.H.Wakefield, Music transcription system:from sound to symbol, *Proc.of AAAI-2000 workshop on artificial intelligence and music,2000*

[3] C.Roads, Research in music and artificial intelligence, *ACM computing survey*, Vol.17, No.2, 1985, pp.163-190

[4]Y.Tadokoro, T.Morita and M.Yamaguchi, Pitch detection of musical sounds noticing minimum output of parallel connected comb filters, *IEEE TENCON 2003*, tencon-72, 2003.

[5]A.P.Klapuri, Multiple fundamental frequency estimation based on harmonicity and spectral smoothness, *IEEE Trans. on Speech and Audio Processing*, Vol.11, No.6, 2003, pp.804-816.

[6]M.Goto, A predominant-F0 estimation method for polyphonic musical audio signals, *Proc. of ICA2004*, 2004, pp.II-1085-1088.

[7]M.Goto, Development of the RWC music database, *Proc. of ICA2004*, 2004, pp.I-553-556.