MUSICAL INSTRUMENT ESTIMATION FOR POLYPHONY USING AUTOCORRELATION FUNCTIONS

Yoshiaki Tadokoro
Dept. of Information and Computer Sciences, Toyohashi University of Technology,Toyoahashi,441-8580 Japan

Koji Tanishita
Dept. of Information and Computer Sciences, Toyohashi University of Technology,Toyoahashi,441-8580 Japan

Keywords: musical instrument estimation, polyphony, autocorrelation function.

Abstract: This paper proposes a new musical instrument estimation of polyphony using autocorrelation functions. We notice that each musical instrument has each autocorrelation function. Polyphony can be separated into each monophony using comb filters ($H(z) = 1 - z^{-N}$). We can obtain the autocorrelation functions for the outputs of comb filters from the autocorrelation functions of the monophony. By the pattern patching between the autocorrelation functions for the output signals of the comb filters and ones calculated from monophony of each instrument, we can estimate the musical instruments for polyphony.

1 INTRODUCTION

Musical transcription is necessary in the musicology field, musical retrieval and also a significant problem in machine perception (Roads, 1985), (Sterian and Wakefield, 2000), (Pollasri, 2002). In the transcription, the pitch estimation is most important and many studies have been done (Roads, 1996), (Tadokoro el 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 el al, 2001, 2002, 2003). On the other hand, there are not many studies for the instrument estimation(Brown and cooke, 1994), (Abe and Ando, 1996), (Zhang, 2001), (Lee and Chun, 2002), (Krishan and Steenivas, 2004), (Jincahita, 2004), although the instrument estimation is also necessary in the transcription. Most of old studies are for monophony and based on the spectrum analysis of a musical sound. In the recent studies, the new technologies such as neural network, fuzzy logic (Zhang, 2001), hidden morkov model (Lee and Chun, 2002) and independent subspace analysis (Jincahita, 2004).

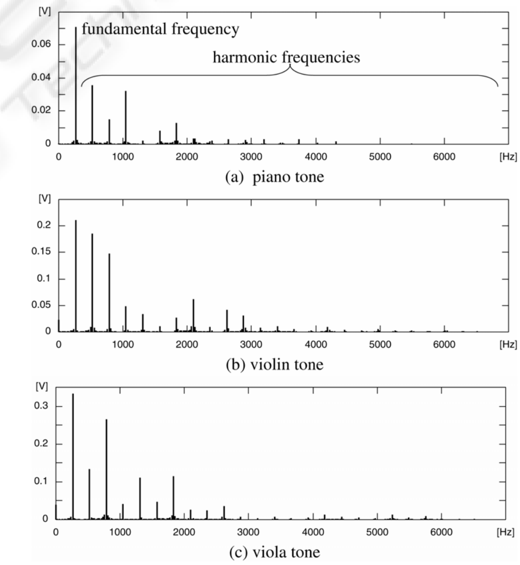

Figure 1: Spectra of C_4 tones, (a)piano, (b)violin and (c)viola

Tadokoro Y. and Tanishita K. (2005).
MUSICAL INSTRUMENT ESTIMATION FOR POLYPHONY USING AUTOCORRELATION FUNCTIONS.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Signal Processing, Systems Modeling and Control*, pages 143-148
DOI: 10.5220/0001175001430148

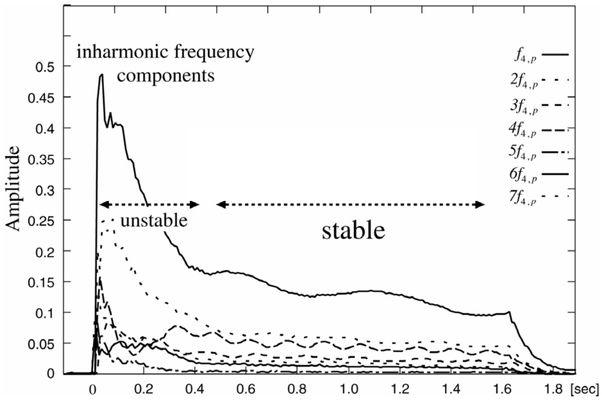

Figure 2: STFT result of piano tone C_4

The spectrum of each musical instrument has different frequency components as shown in Fig.1. Therefore the instrument estimation based on the spectrum analysis is reasonable. But there are some problems in the instrument estimation based on the spectrum analysis. One of them is that the spectrum for the signal just after the instrument is played is unstable. Figure 2 shows the result of the short-time Fourier transform (STFT) of piano tone C_4 (C of octave 4). In the range from 0.0 to 0.4 s, the each harmonic component is changing irregularly. But, we must estimate the instrument in a short duration signal like about 100ms, because a sixteenth note is 125 ms at the tempo of a quarter note =120. Another is that tones in lower octaves have lower fundamental frequencies and to separate polyphony into each monophony and obtain these spectra by the FFT method, we must use a longer signal duration necessarily. For an example, to distinguish two tones of C_2 $(f_{c2} = 65.41\,Hz)$ and $C_{2\#}(f_{C2\#} = 69.30\,Hz)$, we must use at least a signal duration of 257 ms, because the frequency difference between these two tones is 3.89 Hz. That is, the method based on the DFT must use the longer signal duration to obtain a higher frequency resolution. On the other hand, the method based on the parametric model such as the linear prediction method (LPM) can calculate the spectrum from a smaller data. Then we considered the instrument estimation for monophony musical sound using the LPM that could be applied to the sounds of the shorter duration like about 50 ms (Tadokoro et al, 2004). But, the LPM method has the problem that it has many computations and the prediction coefficients are sensitive to the change of a signal waveform.

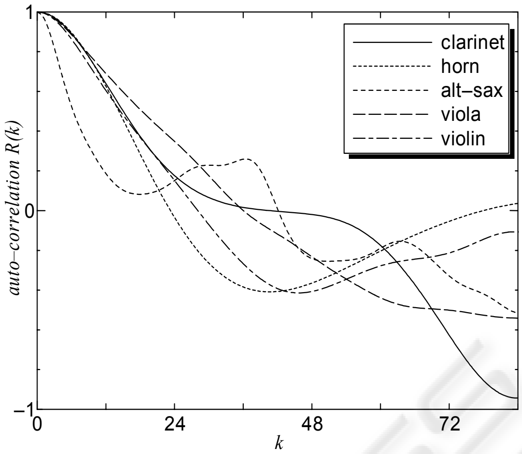

Figure 3: Autocorrelation functions $R(k)$ of C_4 tones for five instruments

Table 1: Accumulated differences between the autocorrelation functions ($R_p(k), R_q(k)$) of two instruments

C_4	clarinet	horn	alt-sax	viola	violin
clarinet	0	24.91	19.25	12.50	19.56
horn		0	23.20	22.71	7.82
alt-sax			0	17.19	19.74
viola				0	16.17
violin					0

In this paper, we consider the instrument estimation using each autocorrelation function for each instrument. The proposed method has a smaller computation than the LPM, because the p-order LPM must use $p \times p$ autocorrelation functions and solve the Yule-Walker equation. Furthermore, we consider the instrument estimation for polyphony musical sound that may be suitable to the pitch estimation method using comb filters that we proposed.

We assume that the polyphony is composed of two different tones of which pitches have already estimated by the pitch estimation method. And the input sounds are real sounds of five instruments (clarinet, horn, alto-sax, viola and violin) and are in octave 3 to 5. These database (RWC music database) are made by the Real World Computing Partnership in Japan. The sampling frequency is $f_s = 44.1\,kHz$. The playing method is moderate but not piano or forte.

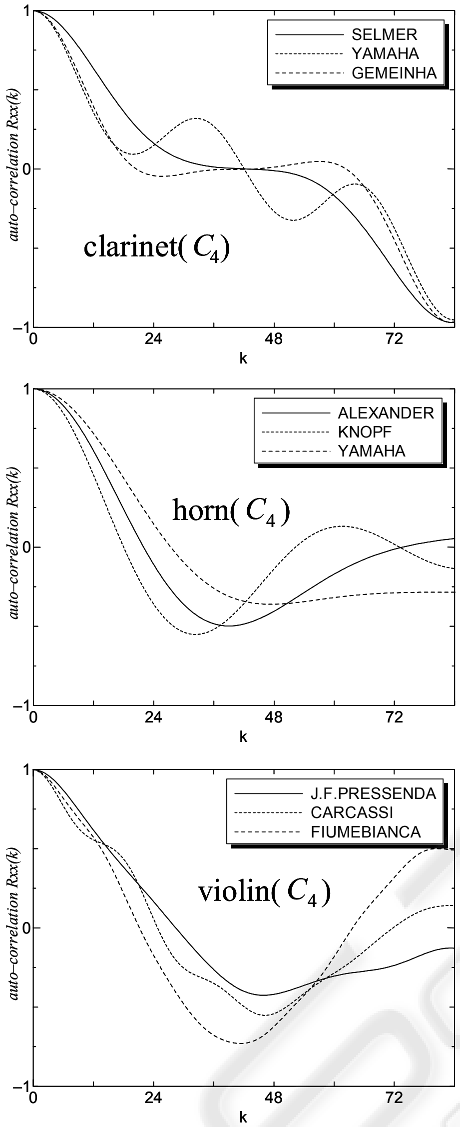

Figure 4: Autocorrelation functions of some instrument makers

Table 2: Accumulated differences between auto-correlations functions of some instrument makers

	clar.1	clar.2	clar.3	hor.1	hor.2	hor.3	viol.1	viol.2	viol.3
clar.1	0	14.8	10.5	28.6	32.3	18.5	19.6	29.1	42.3
clar.2		0	11.6	26.5	30.9	21.1	21.5	27.7	38.4
clar.3			0	27.4	23.4	23.5	24.3	29.8	39
hor.1				0	14	14.5	11.9	8.49	13.8
hor.2					0	38.2	55.5	11.4	34.1
hor.3						0	4.1	12.7	27.2
viol.1							0	9.7	23.9
viol.2								0	34.8
viol.3									0

Table 3: Instrument estimation results for C_4 tone

sound \ est.	clarinet	horn	alt-sax	viola	violin
clarinet	97	0	0	3	0
horn	0	100	0	0	0
alt-sax	0	0	100	0	0
viola	3	0	0	97	0
violin	0	0	0	2	98

2 INSTRUMENT ESTIMATION FOR MONOPHONY USING AUTOCORRELATIO FUNCTIONS

2.1 Autocorrelation Function of Monophony

We calculate the autocorrelation function of a signal $x(n)$ by

$$R(k) = \frac{1}{M}\sum_{n=0}^{M-1} x(n)x(n+k) \tag{1}$$

Figure 3 shows the autocorrelation functions $R(k)$ of C_4 tones for five instruments calculated by using the signals of 50 ms duration in the beginning part of the sounds. Table 1 represents the accumulated differences of $R(k)$ between two instruments showing in Eq.(2)

$$AD(p-q) = \sum_{k=0}^{r} \left| R_p(k) - R_q(k) \right| \tag{2}$$

From these results, we can realize that we can estimate the instruments by the autocorrelation functions $R(k)$.

But we have one problem that the autocorrelation functions for instruments are different depending on the instrument makers. Figure 4 shows the autocorrelation functions for some instruments of some instrument makers. Table 2 represents the accumulated differences $AD(p-q)$ between these autocorrelation functions. From these results, we must prepare the templates for each instrument maker.

2.2 Instrument Estimation for Monophony

We made some experiments for the instrument estimation under the following conditions: The template of autocorrelation function $R_q(k)$ of each instrument is made at the point of 20 ms in the beginning part of a sound, and 100 autocorrelation functions $R_p(k)$ are made randomly in the range from15 ms to 25 ms in the beginning part of the sound. We made some instrument estimations for G_3, C_4 and F_5 tones. Table 3 shows the estimation results for C_4 tone. We could obtain the mean estimation error of 0.8 % for these tones (G_3, C_4, F_5).

3 INSTRUMENT ESTIMATION FOR POLYPHONY USING AUTOCORRELATIO FUNCTIONS

3.1 Separation of Polyphony Using Comb Filter

The comb filter $H_p(z)$ is written by Eq.(3) and its block diagram and the frequency characteristic are shown in Fig.5. We can separate polyphony into each monophony using the comb filters as shown in Fig.6. The comb filter $H_p(z)$ can eliminate one tone corresponding to the its period $T_p = N_p / f_s$ where one delay $z^{-1} = 1/f_s$.

$$H_p(z) = 1 - z^{-N_p} \tag{3}$$

Because the instrument sound with pitch f_p is composed of a fundamental frequency (pitch) f_p and its harmonic ones $nf_p (n = 2,3,\cdots)$. But the amplitudes of $x'_p(n)$ and $x'_q(n)$ of each monophony separated by the comb filters are changed by the amplitude characteristics of the comb filters $H_q(z)$ and $H_p(z)$, respectively.

3.2 Autocorrelation Function of the Output of a Comb Filter

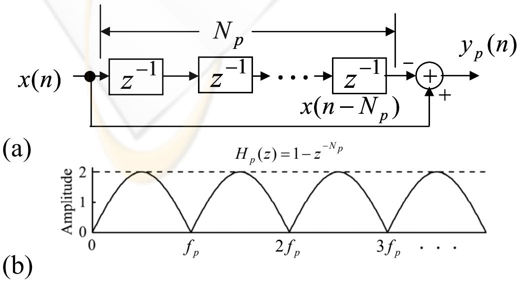

Figure 5: (a) Block diagram of comb filter and (b) its Frequency characteristic

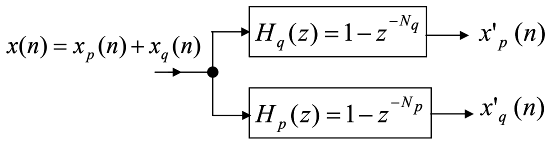

Figure 6: Separation of polyphony into each monophony

The output of the comb filter $H_p(z)$ is written by

$$y_p(n) = x(n) - x(n - N_p) \tag{4}$$

The autocorrelation function $R_{po}(k)$ of $y_p(n)$ can be calculated by using the autocorrelation functions for the monophony as shown in Eq.(5).
By using Eq.(5), we have only the same number of autocorrelation functions for the templates as the number of instruments per each tone. We confirmed that the autocorrelation function of the output of the comb filter can be calculated by Eq.(5). Figure 7 shows two autocorrelation functions, one of them is one calculated by using the output of the comb filter $H_p(z)$, and the other is calculated by Eq.(5) using the autocorrelation function $R(k)$ for monophony

$$\begin{aligned} R_{po}(k) &= \frac{1}{N}\sum_{n=0}^{N-1} y_p(n)y_p(n+k) \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\{x(n)-x(n-N_p)\}\cdot \\ &\qquad \{x(n+k)-x(n+k-N_p)\} \\ &= \frac{1}{N}\sum_{n=0}^{N-1} x(n)x(n+k) \\ &\quad + \frac{1}{N}\sum_{n=0}^{N-1} x(n-N_p)x(n-N_p+k) \\ &\quad - \frac{1}{N}\sum_{n=0}^{N-1} x(n)x(n+k-N_p) \\ &\quad - \frac{1}{N}\sum_{n=0}^{N-1} x(n-N_p)x(n+k) \\ &= 2R(k) - R(N_p+k) - R(N_p-k) \end{aligned} \tag{5}$$

when two tones of the polyphony are a clarinet C_4 and a horn E_4 and the comb filter $H_p(z)$ eliminates the horn E_4 . These autocorrelation functions are almost same. Figure 8 shows the autocorrelation functions of the output $x_p'(n)$ of the comb filter $H_q(z)$ for five instruments in the same condition as Fig.7. Table 4 shows the values of Eq.(2) when the input sound is composed of one tone (C_4) of five instruments and a horn tone (E_4), and the E_4 tone is eliminated by

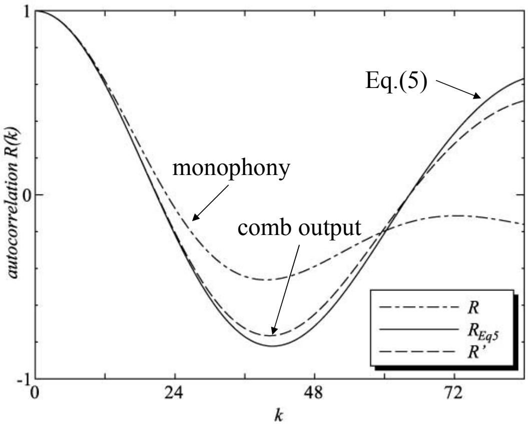

Figure 7: Comparison between autocorrelation function of output (C_4') of the comb filter $H_q(z)(E_4)$ and one calculated by Eq.5 using autocorrelation function $R(k)$ of monophony (C_4:horn, E_4:clarinet)

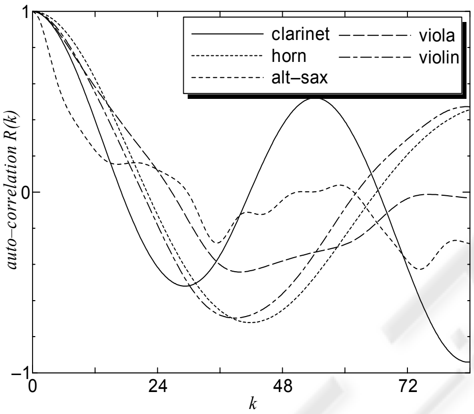

Figure 8: Autocorrelation functions of output of the comb filter $H_q(z)(E_4)$ for five instruments

Table 4 Accumulated differences between the autocorrelation functions ($R_p(k), R_q(k)$) of two instruments when the input sound is $C_4 + E_4$ and the E_4 tone is eliminated by the comb filter $H_q(z)(E_4)$.

C'$_4$	clarinet	horn	alt-sax	viola	violin
clarinet	0	44.51	25.90	37.24	41.57
horn		0	34.62	16.72	6.22
alt-sax			0	20.20	33.90
viola				0	18.96
violin					0

Table 5: Instrument estimation results when the input sound is composed of one of five instruments (C_4) and horn (E_4), and the E_4 tone is eliminated by the comb filter $H_q(z)$(E_4).

	clarinet	horn	alt-sax	viola	violin
clarinet	100	0	0	0	0
horn	0	85	0	0	15
alt-sax	0	0	100	0	0
viola	0	0	0	100	0
violin	0	0	0	0	100

Table 6: Instrument estimation errors.

instrument	sound range	error(%)
clarinet	$D_3 - B_5$	4.6
horn	$C_3 - F_5$	12.7
alt-sax	$D_3 - A_5$	2.2
viola	$C_3 - B_5$	4.9
violin	$G_3 - B_5$	6.0
mean error		6.0

the comb filter $H_q(z)$(E_4). From these results, we can realize that each autocorrelation function of the output of the comb filter for each instrument is different each other.

3.3 Instrument Estimation for Polyphony

Using the combination of five instruments, we made the instrument estimation when two tones are C_4 and E_4. Like in the case of monophony, we made each 100 autocorrelation functions in the range from 15 ms to 25 ms for two outputs of the comb filters in Fig.6. Then we calculated Eq.(2) between the autocorrelation function of the output of the comb filter and the templates calculated by Eq.(5) for five instruments. Table 5 shows one example of the instrument estimation results under the same condition as Table 4. Table 6 shows the each instrument estimation error for two tones that are made by all the combinations of five instruments. We could obtain the mean estimation error of 6% for five instruments.

4 CONCLUSIONS

We proposed a new musical instrument estimation of polyphony using autocorrelation functions. Polyphony can be separated into each monophony using the comb filters. Using the autocorrelation functions of the outputs of the comb filters, we can estimate the instrument by comparing with the autocorrelation functions of the templates that can be calculated from the autocorrelation functions of monophony. We could obtain the mean estimation error of 6% for five instruments.

As a future work, we'd like to reduce the number of templates considering the analogous autocorrelation functions of neighbour tones.

REFERENCES

M.Abe and S.Ando, "Application of loudness/ pitch/timbre decomposition operators to auditory scene analysis," Proc. of ICASP.pp.2646-2649, 1996.

G.J.Brown and M.Cooke, "Perceptual grouping of musical sounds: a computational model," Journal of New Music Research, vol.23, no.2 pp.107-132, 1994.

P.Jincahita, "Polyphonic instrument identification suing independent subspace analysis," IEEE Int. Conf. on Multimedia and Xpo (ICME), vol.2, pp.1211-1214, 2004

A.G.Krishna, and T.V.Sreenivas, "Music instrument recognition: From isolated notes to solo phrases," Proc. Of ICASSP, pp.IV265-268, 2004

J.Lee and J.Chun, "Musical instrument recognition using hidden morkov model," Conference record of the Asilomar Conference on Signals, Systems and Computers, vol.1, pp.196-199, 2002

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

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

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

A.Sterian and G.H.Wakefield, "Music transcription systems: from sound to symbol," Proc. of AAAI-2000 workshop on artificial intelligence and music, July 2000.

Y.Tadokoro and T.Miwa, "Musical pitch and instrument estimation of polyphony using comb filters for transcription," 4th World Multiconference on Circuits, Systems, Communications and Computers (CSCC2000), Advance in Physics, Electronics and Signal Processing Applications, pp.315-319, 2000.

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

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

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

Y.Tadokoro, K.Tanishita and M.Yamaguchi, "Musical instrument estimation using linear prediction method," Proceedings of the 11th International Workshop on Systems, Signals and Image Processing(IWSSIP'04), pp.2077-210, September 13-15, 2004.

T.Zhang, "Instrument classification in polyphonic music based on timbre analysis," Proc. of SPIE, vol.4519, pp.136-147, 2001