Separation of Mixed Plural Sound Sources in Instrumental Music Using Butterworth Parallel BPF

Yusuke Yamaguchi†, Shigeyoshi Nakajima‡ and Takashi Toriu‡,

Fac. of ENG, Osaka City Univ., 3-3-138 Sugimoto-cho, Sumiyoshi-ku, Osaka City, 5588585 JAPAN

Summary
Many researchers investigated sound source separation as one of themes of music information processing. Sound source separation is a work to separate each sound from a mix of several sound sources such as pianos and violins etc. It is one of very important techniques in automatic music transcription. And also it is assumed as one of early stages in music recognition of human auditory perception and it will take a major role in computer simulations of human perception process. We propose a new method of sound source separation using Butterworth parallel band pass filter(BPF). We show that the proposed method is superior to a recent work. As a result of experiment the proposed method improves fidelity of output comparing an original single instrument sound in input.

Key words:
sound source separation, FFT, RCF, Butterworth filter, automatic music transcription

1. Introduction

Recently popularization of handy music players and high fidelity mobile phones made human lives with richer music during walking time, jogging or other time than a decade ago. The main purchase method changed to download. There are huge number of music data in the world wide web net. People want compile such music information data for searching or recommendation of music. H. Sawada et al.[1] introduced sound source separation methods using Independent Component Analysis (ICA) and using sparseness. ICA used an assumption that several sound sources are independent to other sources and made a filter which makes the separated signals independent to others. This method attracts many attentions because it is a blind method and it uses using independence of sources but doesn't use prior information about source types and mixture types. But N. Ono et al.[2] pointed that ICA needs recorded signals the same number or more than sound sources. But an ordinary normal music recording is almost a stereo recording in these days. So signals are fewer than sound signals in such case. Then ICA is not available for most of music scenes. And sparseness of energies of sound is usual in voice signals but is not usual in music signals. Other researchers worked about monaural signals. H. Sakauchi et al.[3] proposed a notch comb filter method to separate sound sources. T. Tokairin et al.[4] proposed a RCF method to separate sound sources. A notch comb filter method must increase the number of cascade connections of the filter as the number of sound sources increase. But D. Matsuyama et al.[5] pointed that the sound source separation ability of a RCF method comes down as the number of cascade connections of filter increases because noise in low frequency band increases. A RCF method doesn't need cascade connections cf. a notch comb filter method then it avoid the increase of noise in low frequency band. If concentration of amplitude gain of a RCF increases, frequency components other than harmonic components reduces but parts of original sound in the output are removed. On the other hand if the concentration decreases the most of components of the original sound remain but also the components of other sound sources remain in the result of separation.

Most of ordinary comb filter methods assume that a harmonic sound is a set of strict integral multiples of the base frequency. But T. Muraoka and S. Kiriu [6] pointed that each harmonic sound slightly changes from the strict integral multiple, slightly high or slightly low. Such the frequency differences between a comb filter and the real harmonics make separation fail.

We proposed a method to solve such the problem in sound source separation. We detected peaks of frequency distribution using FFT adjusted a set of Butterworth filters [7] to pass real components of single instrumental sound. This adjustment makes a set of frequencies of the filter anharmonic. Butterworth filters pass the high frequency peaks of the same instrument sound but attenuate peaks of other instrument sounds. We measure the ability of the separation of the proposed method and compare it with recent works.

Manuscript received April 5, 2011
Manuscript revised April 20, 2011

In the latter part of this paper Section 2 shows a harmonic structure of an instrumental sound. Section 3 shows an instrument sound data base. Section 4 shows the outline and problem of a RCF method. Section 5 shows the proposed method. Section 6 shows results of experiments and consideration. Section 7 shows the conclusion.

2. Harmonic Structure of Instrumental Sound

In this section we will explain a harmonic structure of a instrumental sound.

The principal parameters of a sound are a volume of the sound, a tone and a set of temporal features. The volume of the sound is the amplitude of the wave. The tone consists of a root and ratios of harmonics power to the root power. A set of temporal features consists of a change of the volume and tone along the time axis.

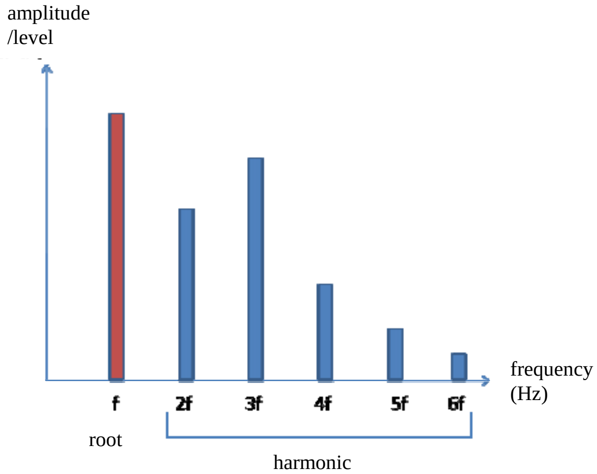


Fig.1 Example of Structure of Harmonics

A base frequency (i.e. a frequency of a root) decide a pitch of a sound. For example a sound with base frequency 440 Hz is A4. Fig.1 shows an example of one of harmonic structure. Two signals from two other instruments with the same root frequency sound different tones because those two harmonics structures are different. There are not only harmonics frequency components but other frequency components in an instrument sound. But two signals sound as the same tone if they have the same structures of the root and the harmonics despite of difference structure of waves outside of root and harmonics bands .

3. RWC Music Database

We used RWC Music Database in our experiments. RWCP(real world computing projects) is one of closed projects of METI (Ministry of Economy, Trade and Industry; successor of Ministry of International Trade and Industry) JAPAN. RWC Music Database is one of the results. It is a free database. We used an instrumental database[8] in it.

4. RCF

4.1 Outline of RCF

We used RWC Music Database. We assume that the number of music source is S, a number of one of the sources is n $(n=1,2,\ldots,S)$, the base frequency of the n^{th} sound source is $f_0^{\,n}(Hz)$, the sampling frequency is $f_s(Hz)$, and the order of the filter N_n is $\lceil f_s/f_0^{\,n} \rceil$ ($\lceil\ \rceil$: ceiling function). Then a RCF for to separate n sound sources from a mixture monaural signal is expressed as below.

$$H_n(Z)=\frac{1-a}{1-a\cdot z^{-N_n}} \qquad (1)$$

Fig.2 shows a block chart of the filter.

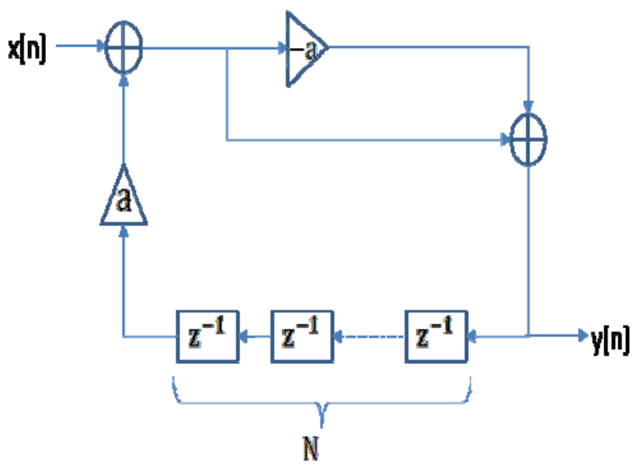


Fig.2 Block Chart of RCF

This filter is an infinite impulse response filter which delays the output $y(n)$ with a delay time N_n and add it to the input. Fig.3 shows the amplitude gain characteristics graph of a RCF.
As shown in Fig.3 a RCF decreases gains of components which frequencies are different from those of the root and

the harmonics. A filter coefficient *a* in Eq.1 varies in a range of $0 \leqq a < 1$. If *a* is near 1, transition bands of an amplitude gain graph become narrow. And if *a* is near 0, transition bands of amplitude gain graph become wide. A transition band means an area between a pass band and a neighbor attenuation band.

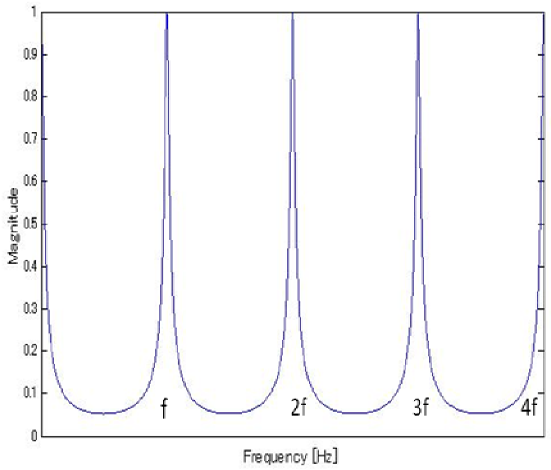


Fig.3 Example of Amplitude Gain Characteristics of RCF

Fig.4 shows a system to separate sound sources using a RCF.

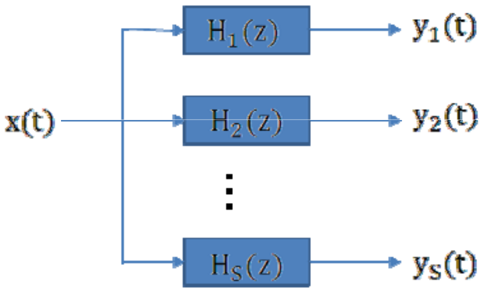
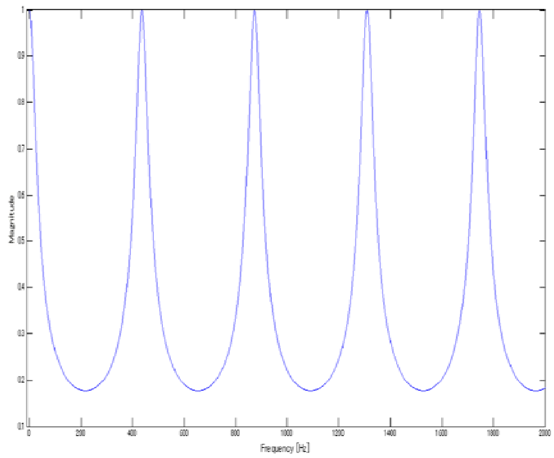


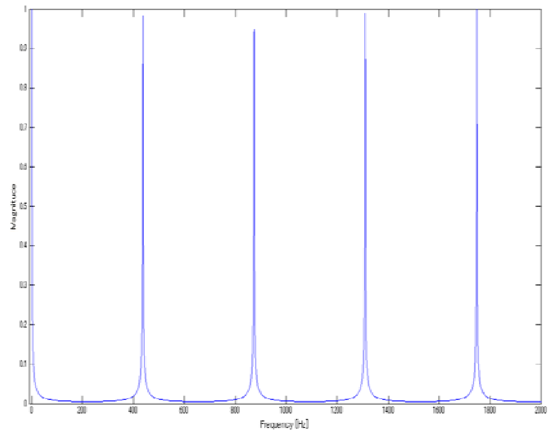
Fig.4 System to Separate Sound Sources using RCF

4.2 Problems of RCF

Fig.5 shows an amplitude gain characteristics of a RCF with different values of *a* 0.7 and 0.99. The cause of the problem is difference between the harmonics frequencies of the comb filter those of real instrumental sound. Usually real harmonic component frequencies exist in transition bands. If we want gains of real harmonics of a target instrument to increase as Fig.5(a) (a= 0.7) gains of attenuation of components of frequencies between harmonics also increase. So components of other instruments increase. If we want suppress components of other instruments and select high value of *a* as Fig.5(b) (a= 0.99) the components of the target instrument decrease. The reason is strictness of integral multiples of the frequencies. So we loosen the strictness slightly and adjust the frequencies of the comb to the frequencies of harmonics of the real target instrument.



(a) a=0.7



(b) a=0.99

Fig.5 Amplitude Gain Characteristics of RCF

5. Method

Fig. 6 shows the block chart of the proposal method. Firstly a time sequence *x(t)* of mixed sound data as input data is transformed into a frequency domain using FFT. Then a base frequency is detected as a root pitch with hamming window described below.

$$w_t = 0.54 + 0.46 \cos (2 \pi t/L) \qquad (2)$$

where t is time and L is the length of data.

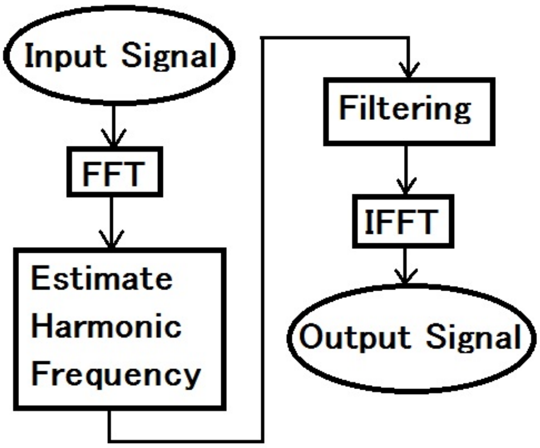


Fig.6 Block Chart of Proposal Method

We assume initial harmonic frequencies as strict integral multiples of the base frequency. We searched the highest peak with ±5Hz width area around the initial frequency. The position of the peak is the real harmonic frequency. Around the real peak we made a Butterworth band pass filter. For example we selected parameters as a ripple of a passband R_p is 1 dB, a stopband attenuation A_s is -40 dB, stopband width F_s is 160 Hz, and the passband width F_q is 30 Hz. An order of a Butterworth band pass filter N is decided as below.

$$N=\frac{\log_{10}[\,(10^{Rp/10}-1)/(10^{As/10}-1)\,]}{2\log_{10}\frac{\omega_s}{\omega_q}} \qquad (3)$$

where $\omega_s=2\pi F_s$ and $\omega_q=2\pi F_q$.

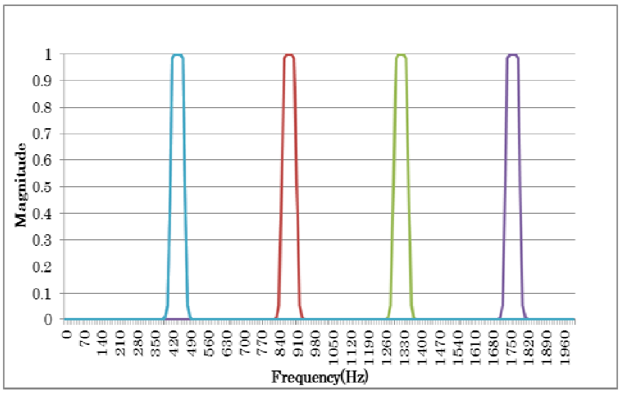


Fig.7 Amplitude Gain Characteristics of RCF

Fig.7 shows an amplitude gain characteristics of the proposed filter. There are 4 Butterworth band pass filters. The BPF around the lowest frequency passes root pitch. Other higher ones pass harmonics respectively. We call the method AH-P-BPF (anharmonicity parallel band pass filter).

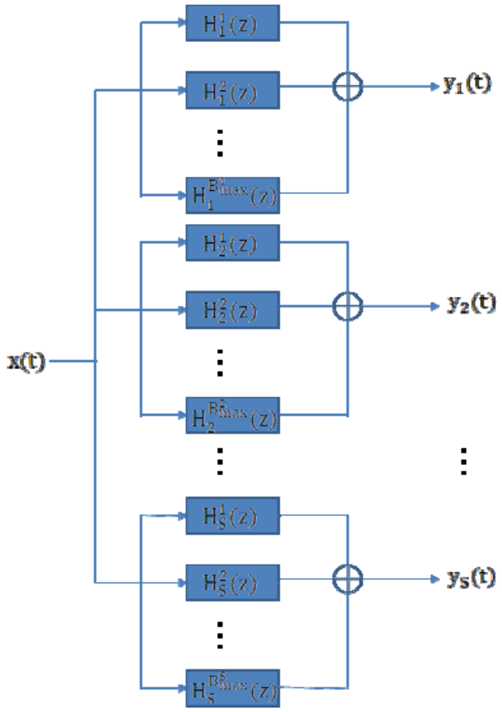


Fig.8 Block Chart of Real Time Model of Proposed Filter

Fig.8 shows the block chart of the real time model of proposed filter. Each $H_i^j(z)$ means a real time Butterworth band pass filter. This real time model is our future target. Block chart of Fig.6 is not a real time model and it is a model to evaluate the quality of the proposed method.

6. Results of Experiments and Consideration

We used AH-P-BPF in our experiments. Also we used H-P-BPF (harmonicity parallel band pass filter) for comparison. H-P-BPF is a set of Butterworth filters without adjustment of harmonics frequencies. And we used several RCFs (a= 0.99, 0.95, 0.90, 0.80, 0.70) for comparison.

We evaluate the result with GDL(generalized distortion level) [8]. The value of $x(t)$ is the input time sequence. After processing there are separated output signal $y_i(t)$ for single instrument sound. Each $y_i(t)$ corresponds to one of harmonics or the root. The original signal in the input corresponding to $y_i(t)$ is $\hat{y}_i(t)$. The separated signal $y_i(t)$ is affected by the gain of the filter. So normalization coefficient G makes it normal.

$$G=\frac{\Sigma\hat{y}_i(t)/L}{\Sigma y_i(t)/L} \qquad (4)$$

where L is the data length.

The separated $y_i(t)$ is multiplied by G, convoluted by a hamming wind, transformed by FFT and becomes $Y_i(t)$. Also $\hat{y}_i(t)$ becomes $\hat{Y}_i(t)$ in the same manner. Then GDL_i is

$$GDL_i=10\log_{10}\frac{\Sigma\{\hat{Y}_i(t)-Y_i(t)\}^2}{\Sigma\{\hat{Y}_i(t)\}^2} \qquad (5)$$

If there are much signal powers in $y_i(t)$ other than the original signal $\hat{y}_i(t)$ GDL_i becomes high . Then a low GDL_i means that a filter is good at separation and output $y_i(t)$ has high fidelity of comparing the original single sound $\hat{y}_i(t)$.

We processed sound data of a mixture of two instruments with different root pitches. We prepared 6 musical instruments, an acoustic guitar (AGAPM), an electric guitar (EGLPM), a flute (FLNOM), a piano (PFNOM), and a trumpet (TRNOM). There are 5 root pitches, C4, E4, F#4, A4, and B4 for each instrument. C4 = 262Hz, E4= 330Hz, F#4= 370Hz, A4= 440Hz, and B4= 494Hz. If there are plural music instrument sounds in a time sequence of data each root pitch of each instrument is different to another. The number of sets of two root pitches different to each other is 20 ($={}_5C_2=5\times 4/2$). The number of sets of two instruments is 36 ($=6\times 6$). A set of two instruments of the same type is available. So there are 360 data variation.

6.1 Experimental Result

We processed 360 mixture data with AH-P-BPF, H-P-BPF, and some RFC with variations of *a*. Table 1 shows the average *GDL* of the experiment. AH-P-BPF is the best and H-P-BPF is near.

Table 1: Average *GDL* of Experiment

AH-P-BPF	-13.41
H-P-BPF	-13.30
RFC 0.99	-4.62
RFC 0.95	-8.54
RFC 0.90	-10.04
RFC 0.80	-10.25
RFC 0.70	-9.50

Table 2: *GDL* of Trumpet(F#4)

AH-P-BPF	20.74
H-P-BPF	-20.64
RFC 0.99	-4.37
RFC 0.95	-5.71
RFC 0.90	-7.67
RFC 0.80	-11.39
RFC 0.70	-13.96

Table 2 shows the *GDL* of a trumpet with its root pitch F#4. Also AH-P-BPF is the best and H-P-BPF is near.

Table 3: *GDL* of Acoustic Guitar(C4)

AH-P-BPF	-15.47
H-P-BPF	-16.11
RFC 0.99	-4.84
RFC 0.95	-15.43
RFC 0.90	-18.94
RFC 0.80	-16.89
RFC 0.70	-14.16

Table 3 shows the *GDL* of an acoustic guitar with its root pitch C4. RFC 0.90 is the best. Sometimes H-P-BPF is superior to AH-P-BPF as shown in Table 3.

6.2 Consideration

The difference between two Butterworth parallel BPF and 5 RCF is apparent in Table 1. Sometimes RFC is superior as Table 3. An anharmonicity filter (AH-P-BPF) is slightly superior to a harmonicity one (H-P-BPF). We think that the difference of frequencies between other root pitches (e.g. C4 vs E4) is too wide. So the effect of Butterworth is very significant and the effect of frequency adjustment is a little significant in such the cases.

7. Conclusion

We proposed a new method to separate sound data to each sound of single instrument using an anharmonicity Butterworth parallel band pass filter and a harmonicity Butterworth parallel band pass filter. We showed an effect of Butterworth parallel band pass filter comparing with RCFs. Low GDL of proposed filter means high fidelity of the output comparing an original single sound. An anharmonicity filter is slightly superior to a harmonicity filter. In future work we will process data which needs anharmonicity seriously.

References

[1] H. Sawada, S. Araki and S. Makino " Recent Advances in Audio Source Separation Techniques," *Journal of IEICE* (*Institute of Electronics, Information and Communication Engineering JAPAN)* , Vol.91, No.4, pp.292-296, 2008.

[2] N. Ono, H. Hosoya, Y. Izumi, and S. Sagayama, "Blind Separation of Music Signals by Harmonic Component Analysis," *Technical Report* of *IEICE EA (Engineering Acoustics)*, 108(70), pp.47-52, 2008.

[3] H. Sakauchi, M. Natsui, Y. Tadokoro, "Realization of Automatic Musical Transcription System Based on Comb Filters," *Technical Report of IPSJ (Information Processing Society of JAPAN) SIG (Signal),* Vol.2007, No.81, pp.13-18, 2007.

[4] T. Tokairin, S. Kitani and R. Muroya , "A Separation of Polyphonic Musica1 Sound and Estimation of Musical Instruments Using Resonator Comb Filters," *Memoirs of Hakodate National College of Technology*, Vol. 43, pp.57-62, 2009.

[5] D. Matsuyama, M. Natsui and Tadokoro Yoshiaki, " Consideration of Pitch Estimation Method for Piano Chords Consisting of Many Notes Using Cascaded Seven Comb Filters," *Technical Report of IPSJ (Information Processing Society of JAPAN) SIG (Signal),* Vol.2007, No.81, pp.167-172, 2007.

[6] T. Muraoka and S. Kiriu, " Reduction of frequency searching processes for Generalized Harmonic Analysis(GHA) ," *Technical Report* of *IEICE DSP (Digital Signal Processing)*, 103(146), pp.1-6, 2003.

[7] S. Takahashi and M. Ikehara, "Digital Filter, " Baifukan, Tokyo JAPAN, 1999.

[8] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC Music Database : Music Genre Database and Musical Instrument Sound Database ," *Technical Report of IPSJ MUS (music)* , Vol.2002, No.40, pp.19-26, 2002.

[9] M Mitsuya, " Digital Filter Design, " Shokodo, Tokyo JAPAN, 1987.

Yusuke Yamaguchi received the B.E. degree in Information Engineering from Osaka City University in 2011. His research interests include signal processing and music data.

Shigeyoshi Nakajima received the B. E. and M. E. degree in Electric Engineering from Kyoto University, Kyoto, Japan in 1982 and 1984 respectively. He received the Ph.D. degree in Information Engineering from Osaka City University, Osaka, Japan in 1997. He is now an Associate Professor in Osaka City University, Japan. His research interests include signal processing, image processing, medical engineering and optimization algorithm and so on. He is a member of IEEE, IEICE and IPSJ.

Takashi Toriu received the B.Sc. in 1975, M.Sc. and Ph.D. degree in physics from Kyoto University, Kyoto, Japan, in 1977 and 1980, respectively. He was a researcher in Fujitsu Laboratories Ltd. from 1982 to 2002, and now he is a Professor of Osaka City University. His research interests are in the areas of image processing, computer vision, and especially in modeling of human visual attention. He is a member of IEEE, IEICE, IPSJ, ITE and IEEJ.