SIGNAL SEPARATION OF INSTRUMENTAL PERFORMANCE AND DEVELOPMENT OF SELECTABLE LISTENING POSITION AUDIO SYSTEM WITH HRTFS

Kenta Niwa†, Mehrdad P. Tehrani‡, Takanori Nishino and Kazuya Takeda†*

† Graduate School of Information Science
‡ Information Technology Center
* Center for Information Media Studies

ABSTRACT

This paper describes the blind signal separation of musical instruments under an ordinary environment and the development of a novel sound-listening system using a head related transfer function. In our study, frequency-domain independent component analysis was used to separate sound signals. Sound signals were recorded using mono-directional and omni-directional loudspeakers because musical instruments have their own directivity. The experiments were conducted using two or three sound sources from instruments (flute, violin, and piano) generated by a synthesizer, and these results indicated that the performance was influenced by the directivity of the instrument and the initial reverberation time. The sound demonstrations using the developed audio system are available at our web site.

1. INTRODUCTION

Blind signal separation (BSS) method based on frequency-domain independent component analysis (ICA) are used in several applications such as speech enhancement, speech recognition, and acoustic signal estimation and so on. Many experiments have been conducted on measurements of speech signals and acoustic signals measured in an anechoic room, but there have been few studies on musical signals or signals measured in an echoic room.

If a musical ensemble is separated into each instrument, we can emphasize our favorite instrument and relocate players as we like (Figure 1). However, a sound source has its own directivity, and there is a reverberation in a room. For example, a trumpet has a strong directivity, while a flute or a violin has less directivity. Of course, a musical performance gives us a fine feeling with a suitable reverberation. Therefore, the directivity of the instrument and ambient reverberation must be considered when we separate musical signals.

This paper describes our evaluations of BSS using musical signals recorded in an echoic rooms and the development of an audio system that can relocate players. We examined the BSS performance using loudspeakers and gave a demonstration through live musical performance. A BSS method based on a frequency-domain ICA was applied for recording signals. The evaluation was conducted using two kinds of loudspeakers: a conventional type with directivity and a dodecahedral type without directivity. The signals of the flute, violin and piano were transduced by loudspeakers located in echoic rooms.

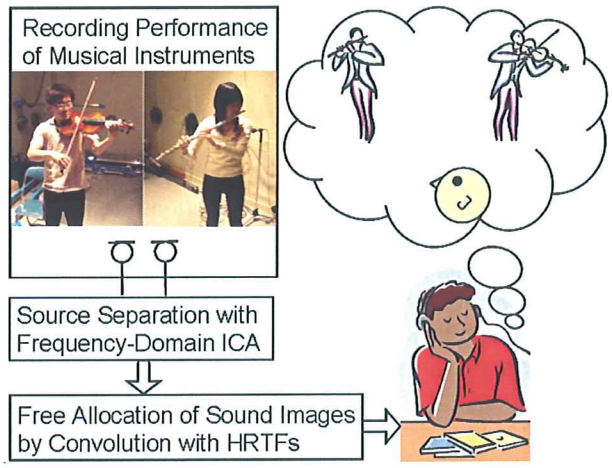

Fig. 1. Selectable listening position (SLP) audio system.

2. SOUND SOURCE SEPARATION METHOD

We used BSS method based on a frequency-domain independent component analysis (FD-ICA), and separated recorded musical signals into each part. In this method, separated signals $\boldsymbol{Y}$ were obtained by multiplying observed signals recorded by microphones and separation filter $\boldsymbol{W}$ in the frequency domain. Figure 2 shows the source separation algorithm based on the FD-ICA. The separation filter was estimated from the observed signals using fast fixed-point algorithm[1] and Kullback-Leibler information minimization [2] for every frequency band. However, since the FD-ICA cannot determine the magnitude and the order of separated signals, we have to solve the scaling problem and the permutation problem.

To solve the scaling problem, we used the minimal dis-

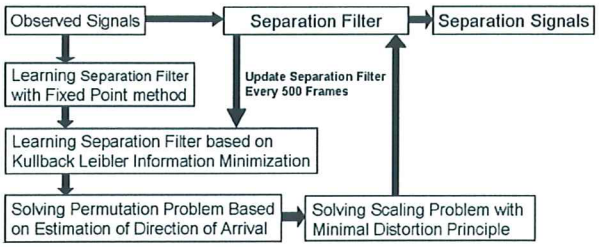

Fig. 2. Flowchart of blind signal separation for SLP audio system.

tortion principal method [3]. This method extracts a diagonal of separation filter matrix $\boldsymbol{W}$, and this diagonal is considered the renewed separation filter matrix $\hat{\boldsymbol{W}}$ (Eq. (1)).

$$\hat{\boldsymbol{W}}(f) = \mathrm{diag}\boldsymbol{W}^{-1}(f)\boldsymbol{W}(f). \tag{1}$$

We modified the sorting method based on estimating the sound source direction with a separation filter[2] to solve the permutation problem. In our study, since the BSS method was applied to the instrumental signal, we implemented our solution's method in the following order: 1) sound source direction, 2) correlation of a harmonic structure and 3) correlation between neighboring frequency bands.

3. EXPERIMENTS

Experiments examining the BSS method for the instrumental signal were conducted using loudspeakers in echoic rooms. Separation performances were evaluated by objective measures.

3.1. Experimental conditions

In our experiments, mixed sound signals were recorded in two echoic rooms. Figure 3 shows the recording environments, and Table 1 lists the recording conditions. The smaller area room is a soundproof chamber, while the other is a lecture room. Sound source signals were transduced by the loudspeaker to simulate an actual musical performance. The number of sound sources is two or three. In the case of two sound sources, we used "G. P. Telemann / Sonata for two flutes" and "G. F. Handel / Bourree." "G. F. Handel / Bourree". The latter piece was also used in the case of three sound sources. Sound signals of a flute, a violin and a piano were made by a synthesizer (YAMAHA, MOTIF6),since original signals are necessary to examine the separation performances. These tones were not played in unison at the same part.

The loudspeakers corresponded to each instrument's transduced sound signals. Since the directivity of instruments is not only mono-directivity, we used conventional loudspeakers having mono-directivity and dodecahedral loudspeakers considered omni-directional sound sources. Figures 4 and 5 show

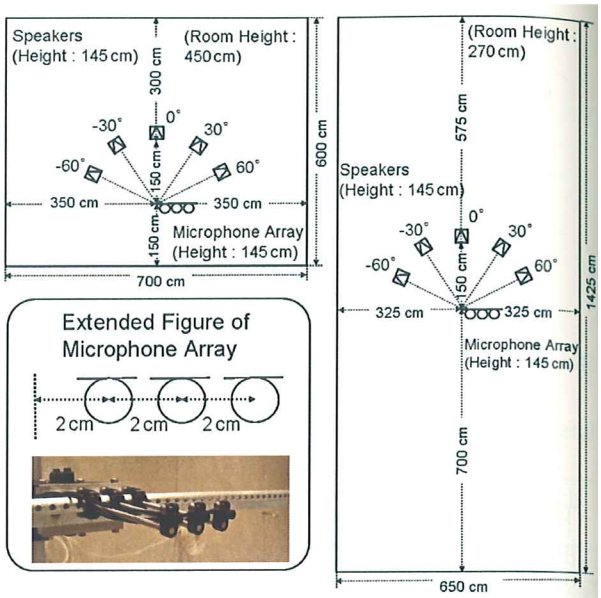

Fig. 3. Recording environment and equipment arrangement.

Table 1. Recording conditions of musical signals

Soundproof chamber	
Background noise level	12.1 dB(A)
Sound pressure level	59.5 dB(A) (1.5 m)
Lecture room	
Background noise level	33.6 dB(A)
Sound pressure level	58.8 dB(A) (1.5 m)

the conditions of loudspeaker arrangement. The azimuth angles are 0° for front microphone, negative for left direction, and positive for right direction. The loudspeakers used to transduce the sound signal were changed when the recordings were conducted, and two kinds of mixed signals were recorded in each loudspeaker arrangement. The microphone array was constructed with three microphones (SONY, ECM-77B). Although original sampling frequency for this recording was 48 kHz, we converted it to 16 kHz because the permutation solving method corresponded to a condition under 16-kHz sampling frequency.

3.2. Experimental Results

The separation performances were evaluated by a cepstrum distance given by

$$\mathrm{CD} = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\sum_{k=1}^{D}\left[c_{y,m}(k) - c'_{y,m}(k)\right]^2}, \tag{2}$$

where $c_y(k)$ is the k-th cepstrum of the target signal in the m-th frame and $c'_y(k)$ is the k-th cepstrum of the separation

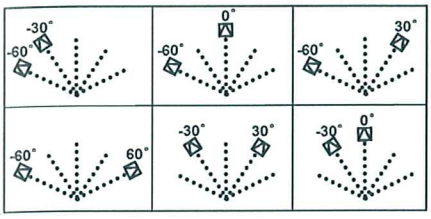

Fig. 4. Loudspeaker arrangement for two sound sources.

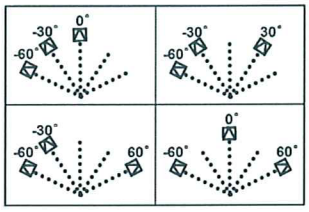

Fig. 5. Loudspeaker arrangement for three sound sources.

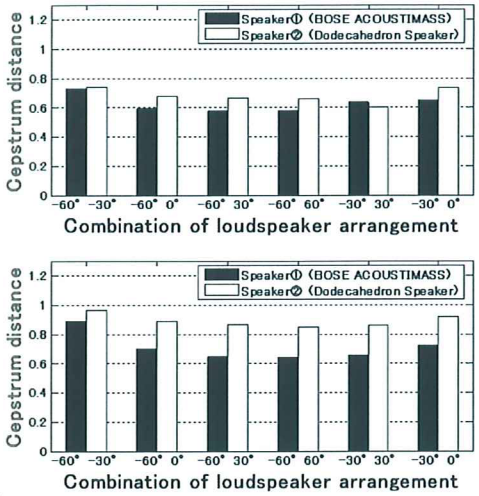

Fig. 6. Cepstrum distance when two sound sources were used; top: results for soundproof chamber, bottom: results for the lecture room.

signal in the m-th frame. D is 2048 for the soundproof chamber and 8192 for the lecture room. M is the number of frames. A small CD score indicates good separation performance.

Figures 6 and 7 show the results of two and three sound sources, respectively. In both figures, the horizontal axis denotes combinations of loudspeaker arrangement and the vertical axis the cepstrum distances. As a result, when the distance between the loudspeakers is far, such as [-60°, 60°] and [-60°, 0°, 60°], separation performance is good.

In the soundproof chamber, there is no difference between the loudspeakers; however, such a difference does exist in the lecture room. The performances using the conventional loudspeaker were better than those using the dodecahedral loudspeaker. The initial reverberation time and the reverberation time calculated using the room impulse responses are shown

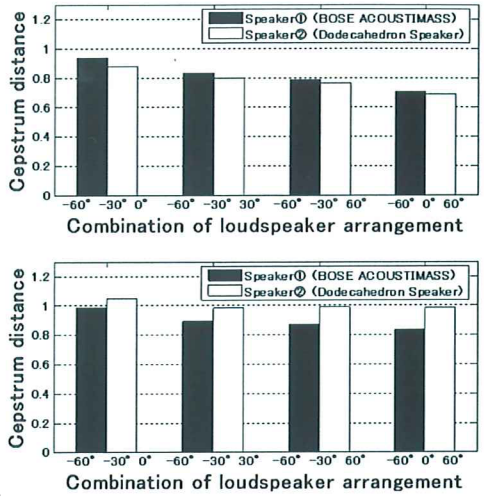

Fig. 7. Cepstrum distance when three sound sources were used; top: results for soundproof chamber, bottom: results for lecture room.

Table 2. Initial reverberation time and reverberation time. Conventional loudspeaker is SP #1, dodecahedral is SP #2

	Initial RT [ms]	RT [ms]
Soundproof chamber (SP #1)	14	138
Soundproof chamber (SP #2)	80	126
Lecture room (SP #1)	13	730
Lecture room (SP #2)	402	648

in Table 2. In the case of the dodecahedral loudspeaker, the initial reverberation time was long when it was compared between two rooms. It was assumed that the increase in initial reverberation time influenced the separation performances.

Figure 8 shows the results of cepstrum distance for each instrument. The CD score of two loudspeakers was better than that of three loudspeakers. Since signals of a piano were only used in the case of three loudspeakers, this signal caused degradation of performance. Figure 9 shows the spectrograms of a flute and a piano. The harmonics of a piano is more complex than that of a flute. Moreover, a piano can play multiple tones simultaneously.

4. SELECTABLE LISTENING POSITION AUDIO SYSTEM

To evaluate the effectiveness of the selectable listening position (SLP) audio system, we recorded the mixed signals of a live instrumental performance and separated these signals into each instrumental signal. Table 3 shows the recording programs and conditions. A piano tone (synthesizer) was transduced by a loudspeaker (YAMAHA, MS-10).

We did not evaluate the separation performances objec-

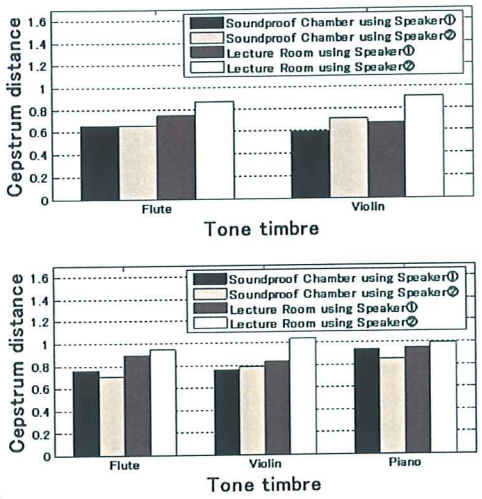

Fig. 8. Cepstrum distance for each instrument; top: results for two sound sources, bottom: results for three sound sources.

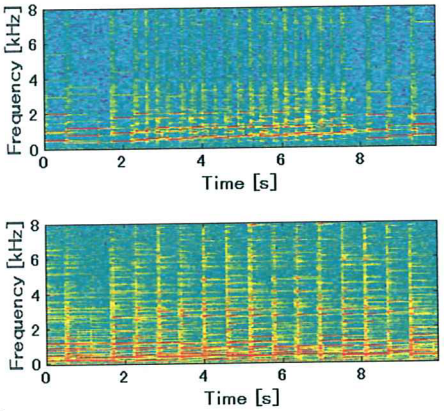

Fig. 9. Sound spectrogram; top: flute, bottom: piano. Colors correspond to the magnitude: red is strong and blue is weak.

tively because it was difficult to record the original signals. The performances were evaluated by preliminary subjective tests, and fine performances were obtained in the case of a soundproof chamber. The trumpet could be separated in the case of a lecture room because it has strong mono-directivity. We assume that the signal of an instrument with mono-directivity is easy to separate.

The SLP audio system convolves the separation signals with a head related transfer function (HRTF) and rearranges each sound image. An HRTF is an acoustic transfer function between the sound source and the ear canal, and it's sometimes used in spatial audio. We used the HRTF database [4] that measured with a head and torso simulator (B&K, 4128). The HRTFs in the directions not covered by this database do not include were obtained using a linear interpolation method [5]. By reproducing the sound signals with the HRTFs, the user can freely arrange the location of sound images. Furthermore, since the SLP audio system re-mixes signals, good performance is obtained even if the signals do not separate perfectly.

Table 3. Recording programs and conditions

Program	Location	Instruments
Sonata for two flutes	-60° 60°	Violin Flute
Bourree (Two sources)	-60° 60°	Violin Flute
Bourree (Three sources)	-60° 0° 60°	Violin Piano (Synthesizer) Flute
Aida	-60° 60°	Trumpet Piano (Synthesizer)

The reader can listen to demonstrations of the SLP audio system can be listened at this site:
http://www.sp.m.is.nagoya-u.ac.jp/~niwa/

5. CONCLUSIONS AND FUTURE WORKS

We examined the signal separation of musical instruments in echoic rooms and developed a selectable listening point audio system. As a result, good performance could be obtained when the sources were at a distance from each other, and it was confirmed that the directivity of instruments influences the separation performance in an echoic room. The developed SLP audio system gives us the freedom to arrange players as we like. Future works include improving the signal separation performance and developing a real-time SLP audio system.

6. REFERENCES

[1] E. Bingham and A. Hyvarinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, Feb. 2000.

[2] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *Proc. of International symposium on ICA*, pp. 505–510, Apr. 2003.

[3] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. of International symposium on ICA*, pp. 722–727, Dec. 2001.

[4] http://www.sp.m.is.nagoya-u.ac.jp/HRTF/.

[5] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolation of the head related transfer function on the horizontal plane," *J. Acoust. Soc. Jpn.*, vol. 55, no. 2, pp. 91–99, Feb. 1999 (in Japanese).