SOURCE-ORIENTED LOCALIZATION CONTROL OF STEREO AUDIO SIGNALS BASED ON BLIND SOURCE SEPARATION

Yuuki Haraguchi[†], *Shigeki Miyabe*[†*], *Hiroshi Saruwatari*[†], *Kiyohiro Shikano*[†], *Toshiyuki Nomura*[‡]

† Nara Institute of Science and Technology

{yuuki-h, shige-m, sawatari, shikano}@is.naist.jp

‡ Common Platform Software Research Laboratories, NEC Corporation

t-nomura@da.jp.nec.com

ABSTRACT

We propose methods to analyze and control source localization of stereo audio signals using blind source separation (BSS) based on independent component analysis (ICA). Although an inverse system of separation compensates distortion caused by ICA as reconstruction of stereo spatial characteristics, this technique is insufficient to analyze localization because it achieves compensation of distortion and reconstruction of spatial characteristics simultaneously. Thus we analyze spatial characteristics effectively by dividing the compensation into two steps: monaural-output compensation of distortion and its reconstruction of spatial characteristics. Additionally, we control the localization of each source by modifying the analyzed spatial characteristics. It is shown that the proposed method can be applied to stereo signals consisting of more than two sources.

Index Terms— independent component analysis, blind source separation, sound-localization control, audio object

1. INTRODUCTION

The recent performance advance and price-reduction of DSP have spread to various audio effect systems that can achieve not only simple tone control but also 3D audio effects to control reverberation, width of chamber and many other spatial impressions. However, they are merely modifications of ready-made multichannel audio, and they are not sufficient for user-controllable audio reedit. Our research purpose is to construct a system in which users can reedit each of the sources as if the users can manipulate the mixing console by themselves and achieve

- customizable spatial re-allocation of audio objects,
- selectable enhancement of specific sources, and
- listener's virtual movement in primary sound field.

As one piece of evidence that the user-controllable audio reedit is in strong demand, the ISO/IEC Moving Picture Experts Group (MPEG) has started the Spatial Audio Object Coding (SAOC) project, which aims to standardize user-controllable audio technology [1]. The most attention-getting technology of SAOC is binaural cue coding (BCC) [2], which has been adopted as the basis of MPEG Surround standardized before SAOC. This method can encode multichannel signal with a low bit-rate by parameterizing inter-channel level difference (ICLD), inter-channel time difference (ICTD), and inter-channel coherence (ICC), which are the most important attributes of source localization [3]. However, this method analyzes characteristics of mixed audio signals but not localization of those sources, and is insufficient for the source reedit. Although some researchers have proposed a system to extract and edit vocal and drums parts [4], they utilize characteristics of specific instruments and cannot be applied to general audio signals. Another method is proposed without assumption of specific instrument [5]. However, the quality of this method is degraded because of nonlinear filtering.

In this paper, we propose analysis and modification of source localization. When available information is only the stereo audio signal itself, which is a mixture of multiple sources, we have to extract information on the objective sources to control localization. For this purpose, we focus on blind source separation (BSS), especially that based on independent component analysis (ICA) [6] because of its high-quality separability. In the conventional BSS, optimized ICA outputs monaural distorted estimation of each separated source. To compensate the distortion, the distorted monaural output is reconstructed as stereo source signal with its information on localization recovered. We analyze information of localization by dividing BSS into two steps; the first step is monaural source separation with low distortion, and the second step is reconstruction of the sources' spatial information. Moreover, by modifying ICLD of sources based on the analyzed information, we control localization of each separated source. Because ICA applied to stereo signal separates two dominant sources in each narrow subband and the proposed algorithm has no explicit identification of each source, the proposed localization control can be applied to any stereo signal even with more than two sources. Using stereo music signals with each audio track in the mixing console available, the performance of the proposed method is verified in both objective and subjective evaluation.

2. CONVENTIONAL BLIND SOURCE SEPARATION

2.1. Mixing Model

In this section and Sect. 3, we assume that the number of sound sources is L and the number of audio channels is M, and we deal with the case of $L = M$. Note that the case of $L \neq M$ is discussed in Sect. 4.

The source signal of the L sources in the time-frequency domain is denoted by an L-dimensional vector $\boldsymbol{S}(f,t) = [S_1(f,t), \ldots, S_L(f,t)]^{\mathrm{T}}$, where f is the index of the frequency bin and t is the index of the analysis frame. In addition, a linear time-invariant transfer system is denoted by an $M \times L$ mixing matrix $\boldsymbol{A}(f) = [A_{ml}(f)]_{ml}$, where $A_{ml}(f)$ is the transfer function from the l-th source to the m-th channel, and $[x]_{ml}$ denotes the matrix that includes the element x in the m-th row and the l-th column. Then, the observed signal $\boldsymbol{X}(f,t) = [X_1(f,t), \ldots, X_M(f,t)]^{\mathrm{T}}$ is written approximately as

$$\boldsymbol{X}(f,t) = \boldsymbol{A}(f)\boldsymbol{S}(f,t). \quad (1)$$

2.2. Frequency Domain ICA

Assuming the source signals are statistically independent mutually and no more than one source is Gaussian, ICA learns the demixing filter in an unsupervised manner. The condition of successful separation with the demixing filter is equivalent to independence among output signals. Here we describe frequency domain ICA (FDICA) [7] where ICA is processed in the frequency domain. In this method, by using the demixing matrix $\boldsymbol{W}(f) = [W_{lm}(f)]_{lm}$, the separated signal $\boldsymbol{Y}(f,t) = [Y_1(f,t), \ldots, Y_L(f,t)]^{\mathrm{T}}$ is given by

$$\boldsymbol{Y}(f,t) = \boldsymbol{W}(f)\boldsymbol{X}(f,t). \quad (2)$$

In addition, $\boldsymbol{W}(f)$ can be optimized by the following iterative updating formula [7]:

$$\boldsymbol{W}^{[i+1]}(f) = \mu\left[\boldsymbol{I} - \langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,t))\boldsymbol{Y}(f,t)^{\mathrm{H}}\rangle_t\right]\boldsymbol{W}^{[i]}(f) + \boldsymbol{W}^{[i]}(f), \quad (3)$$

where $\boldsymbol{I}$ denotes the identity matrix, $\langle\cdot\rangle_t$ denotes the time-averaging operator, H shows conjugate transposition, $[i]$ is used to express the

*Research Fellow of the Japan Society for the Promotion of Science.

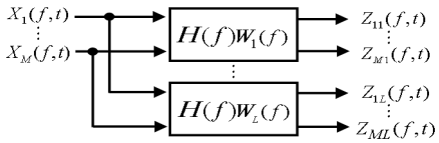

Fig. 1. Separation step based on conventional ICA (FDICA+PB).

value of the i-th step in the iterations, μ is the step-size parameter, and $\mathbf{\Phi}(\cdot)$ is the appropriate nonlinear vector function [8].

2.3. Projection Back

Since the criterion of independence does not specify amplitude and order of signals, FDICA itself is insufficient as filter learning. Ambiguity of amplitude, the so-called *scaling problem*, randomizes spectral characteristics and it results as distortion in the output signals. Ambiguity of order is known as the *permutation problem*, and without identifying correspondence between the sources and separated outputs, broad-band estimation of the source signals cannot be obtained. Here we discuss the solution of the scaling problem using projection back (PB) [9]. Under the assumption that the demixing matrix $\boldsymbol{W}(f)$ separates source components accurately, and permutation ambiguity is aligned by some means [7], $\boldsymbol{W}(f)$ can be expressed as follows:

$$\boldsymbol{W}(f) = \mathrm{Diag}(\boldsymbol{C}(f))\boldsymbol{A}(f)^{-1}, \tag{4}$$

where $\boldsymbol{C}(f) = [C_1(f), \ldots, C_L(f)]^{\mathrm{T}}$ is a constant vector which denotes gain ambiguity of ICA, $\mathrm{Diag}(\cdot)$ is the diagonal matrix whose diagonal component is each element of column vector $\cdot$. To compensate for the effect of $\boldsymbol{C}(f)$, PB applies the inverse matrix of the demixing matrix

$$\boldsymbol{H}(f) = \boldsymbol{W}(f)^{-1}, . \tag{5}$$

The inverse matrix $\boldsymbol{H}(f)$ reconstructs the amplitude of the separated signals at each of the audio channels, and its output signal $Z_{ml}(f,t)$ of the l-th separated signal at the m-th channel can be given as follows:

$$\begin{aligned}[Z_{ml}(f,t)]_{ml} &= \boldsymbol{H}(f)\mathrm{Diag}(\boldsymbol{Y}(f,t)) \\ &= \boldsymbol{A}(f)\mathrm{Diag}(\boldsymbol{S}(f,t)).\end{aligned} \tag{6}$$

Thus, the scaling problem is solved in the form of the reconstruction of the transfer system $\boldsymbol{A}(f)$, referred to as projection back (PB) [9].

In general, $Z_{ml}(f,t)$ is obtained directly by the filter $\boldsymbol{H}(f)\boldsymbol{W}_l(f)$ instead of obtaining $\boldsymbol{Y}(f,t)$ where $\boldsymbol{W}_l(f)$ denotes the demixing matrix replacing all coefficients by zero except the l-th row of $\boldsymbol{W}(f)$ (see Fig. 1).

3. ANALYSIS OF SOUND LOCALIZATION

In this section, we propose an analysis method of sound localization. Since sound localization is determined individually for each of the sound sources, analysis of sound localization is inextricably linked to source separation.

3.1. Extraction of Sound-Localization Information

From Eq. (1), the transfer system $\boldsymbol{A}(f)$ contains all information on sound localization for each of the source signals in $\boldsymbol{S}(f,t)$. Assuming $\boldsymbol{A}(f)$ is known, the following processing is possible by using $\boldsymbol{A}(f)$.

First, from Eq. (1), the source separation is entirely achieved as follows:

$$\begin{aligned}\boldsymbol{Y}(f,t) &= \boldsymbol{A}(f)^{-1}\boldsymbol{X}(f,t) \\ &= \boldsymbol{S}(f,t).\end{aligned} \tag{7}$$

Second, the same sound localization that the l-th sound source has can be given to another monaural sound source $R(f,t)$ as

$$\tilde{\boldsymbol{X}}(f,t) = \boldsymbol{A}(f)[Y_1(f,t), \ldots, Y_{l-1}(f,t), R(f,t), Y_{l+1}(f,t), \cdots, Y_L(f,t)]^{\mathrm{T}}, \tag{8}$$

where $\tilde{\boldsymbol{X}}(f,t)$ is the signal replacing $Y_l(f,t)$ with $R(f,t)$. We call such a substitution of the sources **punch in**.

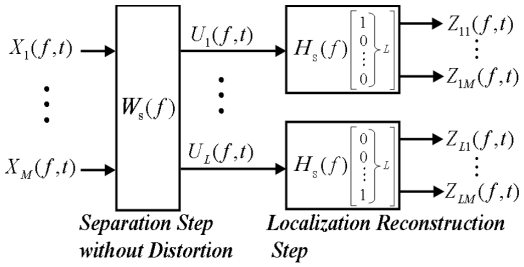

Fig. 2. Procedure step of proposed method

Third, we can control sound localization individually and freely by modifying the relation among the channels of $\boldsymbol{A}(f)$. Using modified transfer system $\hat{\boldsymbol{A}}(f)$, the localization-controlled audio signal $\hat{\boldsymbol{X}}(f,t) = [\hat{X}_1(f,t), \ldots, \hat{X}_M(f,t)]^{\mathrm{T}}$ is given by

$$\hat{\boldsymbol{X}}(f,t) = \hat{\boldsymbol{A}}(f)\boldsymbol{Y}(f,t). \tag{9}$$

However, the transfer system $\boldsymbol{A}(f)$ is generally unknown in practical situations and should be estimated by some means.

3.2. Problem of Conventional Projection Back

From Eqs. (4) and (5), the inverse matrix $\boldsymbol{H}(f)$ of $\boldsymbol{W}(f)$ estimated by ICA is given as follows:

$$\boldsymbol{H}(f) = \boldsymbol{A}(f)\mathrm{Diag}(\boldsymbol{C}(f))^{-1}. \tag{10}$$

If the punch-in process was implemented by using $\boldsymbol{H}(f)$, another monaural sound source $R(f,t)$ would be affected not only by $\boldsymbol{A}(f)$ but also by $\mathrm{Diag}(\boldsymbol{C}(f))^{-1}$. Therefore, $\boldsymbol{H}(f)$ is inadequate to substitute for the transfer system $\boldsymbol{A}(f)$. However, it is very difficult for the deconvolution to be achieved without information on source signals.

3.3. Proposed Algorithm

If the separation is achieved without distortion, its inverse filter plays only the role of reconstructing localization. Then the inverse filter can be used as an approximation of the transfer system $\boldsymbol{A}(f)$, and can achieve punch in without distorting the substituted source. Thus our strategy is to divide the separation process into two steps: a separation step without distortion and a localization reconstruction step (see Fig. 2). The localization control can be achieved by modifying the localization reconstruction step.

3.3.1. Monaural separation step without distortion

In this section, to separate the observed signals into each of the monaural source signals without distortion, we obtain the demixing filter $\boldsymbol{W}_{\mathrm{s}}(f)$, which intentionally scales each of its separated signals to an average value of the channels.

It is easy to obtain the average value of the channels with respect to each sound source at the audio channels by using PB. Furthermore, it can be said that the average value of the channels is a monaural signal with little distortion. By using Eq. (6), the channel-averaged source estimation is given by

$$\begin{aligned}&\frac{1}{M}\cdot\left[\sum_{m=1}^{M} Z_{m1}(f,t), \ldots, \sum_{m=1}^{M} Z_{mL}(f,t)\right]^{\mathrm{T}} \\ &= \frac{1}{M}\cdot\mathrm{Diag}([\boldsymbol{H}(f)]^{\mathrm{T}}[\underbrace{1, \ldots, 1}_{M}]^{\mathrm{T}})\boldsymbol{W}(f)\boldsymbol{X}(f,t).\end{aligned} \tag{11}$$

Therefore, the demixing filter $\boldsymbol{W}_{\mathrm{s}}(f)$ is defined as follows:

$$\boldsymbol{W}_{\mathrm{s}}(f) = \frac{1}{M}\cdot\mathrm{Diag}\Big([\boldsymbol{H}(f)]^{\mathrm{T}}[\underbrace{1, \ldots, 1}_{M}]^{\mathrm{T}}\Big)\boldsymbol{W}(f). \tag{12}$$

Thus, by using $\boldsymbol{W}_{\mathrm{s}}(f)$, the average value of the channels with respect to each sound source $\boldsymbol{U}(f,t) = [U_1(f,t), \ldots, U_L(f,t)]^{\mathrm{T}}$ is given by

$$\boldsymbol{U}(f,t) = \boldsymbol{W}_{\mathrm{s}}(f)\boldsymbol{X}(f,t). \tag{13}$$

3.3.2. Localization reconstruction step

Here, $\boldsymbol{H}_{\mathrm{s}}(f) = \boldsymbol{W}_{\mathrm{s}}(f)^{-1}$ is defined as the localization reconstruction filter. This filter only takes charge of reconstructing sound localiza-

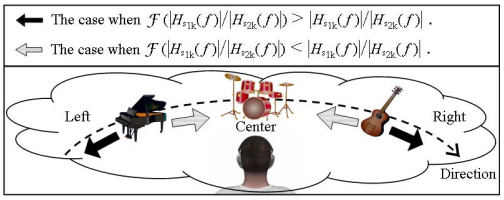

Fig. 3. Configuration of sound-localization control

tion to the separated signal $\boldsymbol{U}(f,t)$.

By applying $\boldsymbol{H}_{\rm s}(f)$ to $\boldsymbol{U}(f,t)$, the output signals are equivalent to the output signals of PB as follows:

$$\boldsymbol{H}_{\rm s}(f){\rm Diag}(\boldsymbol{U}(f,t)) = [Z_{ml}(f,t)]_{ml}\,. \tag{14}$$

This indicates that $\boldsymbol{H}_{\rm s}(f)$ reconstructs the inter-channel level and phase differences to the monaural separated signal $\boldsymbol{U}(f,t)$, which has sound reverberation caused by the transfer system $\boldsymbol{A}(f)$. Therefore, $\boldsymbol{H}_{\rm s}(f)$ can be approximated to play only the role of reconstructing sound localization of $\boldsymbol{U}(f,t)$.

4. PROPOSED SOURCE-LOCALIZATION CONTROL

4.1. Motivation

In this section, by changing the inter-channel gain difference of the localization reconstruction filter $\boldsymbol{H}_{\rm s}(f)$, we control the direction of the virtual image of each source, as its configuration is shown in Fig. 3. The inter-channel gain difference between the left and right channels of $\boldsymbol{H}_{\rm s}(f)$ concerned with each separated signal is nearly in one-to-one correspondence to the direction of the source. Thus, by modifying the inter-channel gain difference between the left and right channels of $\boldsymbol{H}_{\rm s}(f)$, the direction of each separated signal can be controlled.

In general, the number of sources must be estimated in advance in the BSS. Additionally, high-quality separation is difficult with many sources. Nevertheless, the proposed method can deal with an arbitrary number of sources because of the following reasons. First, since the proposed control of localization is a simple conversion of the inter-channel gain difference and the explicit identification of the sources is unnecessary, we need not solve the permutation, which is difficult to solve with an unknown or large number of sources. Second, as discussed in the following section, two-input two-output ICA can analyze localization of stereo signal consisting of an arbitrary number of sources sufficiently.

4.2. Behavior of Localization Analysis with Many Sources

In this section, we discuss the behavior of two-input two-output ICA against stereo signal consisting of many sources. Assuming sparseness among sources [10], it can be expected that the number of dominant sources often decreases in each narrow subband. Sparseness among sources is an assumption that the magnitudes of the sources are distributed sparsely in the time-frequency domain and no two dominant source components share the same time-frequency grid.

First, in the time-frequency bin where the number of dominant sources is below two throughout all the frames, the analysis of sound localization can be achieved successfully as discussed in Sect. 3.3. Next, in the time-frequency bin where more than two dominant sources exist, ICA separates two dominant sources to maximize the difference in statistical behaviors between the separated signals. As a result, ICA estimates two clusters of sources and the localization reconstruction filter plays the role of reconstructing sound localization to the separated monaural source clusters. Thus, ICA can sufficiently analyze sound localization information of stereo signals consisting of more than two sources.

4.3. Algorithm

By changing the inter-channel gain difference between the left and right channels of $\boldsymbol{H}_{\rm s}(f)$ with its total power maintained, only the direction can be controlled without affecting perception of distance. Here, sound-localization control of the individual sources is achieved approximately by converting the inter-channel gain difference of each separated signal with some function as

$$\frac{|\hat{H}_{{\rm s}1k}(f)|}{|\hat{H}_{{\rm s}2k}(f)|} = \mathcal{F}\left(\frac{|H_{{\rm s}1k}(f)|}{|H_{{\rm s}2k}(f)|}\right), \tag{15}$$

where $H_{{\rm s}mk}(f)$ denotes the unprocessed coefficient of the localization reconstruction filter concerned with the k-th separated signal at the m-th channel, $\hat{H}_{{\rm s}mk}(f)$ denotes its modified version, and $\mathcal{F}(\cdot)$ is an arbitrary function to modify the inter-channel gain difference. In addition, various control is possible according to the design of this function $\mathcal{F}(\cdot)$. Using $H_{{\rm s}mk}(f)$, the modified coefficient $\hat{H}_{{\rm s}mk}(f)$ can be written as

$$\hat{H}_{{\rm s}mk}(f) = \sqrt{\frac{\sum_m |H_{{\rm s}mk}(f)|^2}{|H_{{\rm s}mk}(f)|^2\left\{\left(\mathcal{F}\left(\frac{|H_{{\rm s}1k}(f)|}{|H_{{\rm s}2k}(f)|}\right)\right)^{2(-1)^m}+1\right\}}}\,H_{{\rm s}mk}(f) \quad \text{for } m=1,2. \tag{16}$$

Using $\hat{\boldsymbol{H}}_{\rm s}(f) = [\hat{H}_{{\rm s}ml}]_{ml}$ obtained above, the signal in which the controlled direction of each sound source $\hat{\boldsymbol{X}}_{\rm prop}(f,t) = [\hat{X}_{1_{\rm prop}}(f,t), \hat{X}_{2_{\rm prop}}(f,t)]^{\rm T}$ can be given as

$$\hat{\boldsymbol{X}}_{\rm prop}(f,t) = \hat{\boldsymbol{H}}_{\rm s}(f)\boldsymbol{U}(f,t). \tag{17}$$

5. EVALUATION EXPERIMENT

5.1. Experimental Condition

In this section, we verify the efficiency of the proposed analysis and the process of the localization information using ICA by comparing the performance of the proposed method and competitive methods. The comparison is conducted in both the subjective and objective evaluations. In this experiment, to simplify the discussion, we used the gain-difference-conversion function denoted in Eq. (15) to control the range of the localized directions given by

$$\mathcal{F}\left(\frac{|H_{{\rm s}1k}(f)|}{|H_{{\rm s}2k}(f)|}\right) = \left(\frac{|H_{{\rm s}1k}(f)|}{|H_{{\rm s}2k}(f)|}\right)^{\alpha}. \tag{18}$$

With this function, the gain difference is converted in proportion to α in the log domain.

Here we describe two competitive methods.

Competitive method 1: This method is a control of localization based on fixed filtering. In this method, the inter-channel averaged level difference of the analysis frames of the stereo channels $X_1(f,t), X_2(f,t)$ is modified to its α-th power.

Competitive method 2: This method is a control of localization based on time-varying filtering. In this method, the inter-channel level difference of the stereo channels $X_1(f,t), X_2(f,t)$ is modified to its α-th power without changing the total power of the channels in each of the time-frequency grids.

In both the subjective and objective evaluations, we used six stereo recordings of music. Each of the stereo signals consists of three instruments, and each panned stereo signal of each source track is available separately and is used in the evaluation of the signal-to-noise ratio (SNR). All of them are recorded and edited by professional musicians. They are recorded in sampling frequency 44.1 kHz with quantization of 16 bit. For each of the stereo signals, we made six processed signals by all three methods with two settings of the parameter, i.e., setting $\alpha = 10$ to spread the width of the spacial image and setting $\alpha = 1/10$ to narrow the width. The length of the filter is 1024 taps.

5.2. Objective Evaluation

We compare the controllability of the conventional and proposed methods in the objective evaluation. By filtering the stereo signal of separated source track $s_{lm}(n)$ in each of the methods, we obtain the processed stereo signal $\hat{s}_{lm}(n)$ of each sources, where l denotes

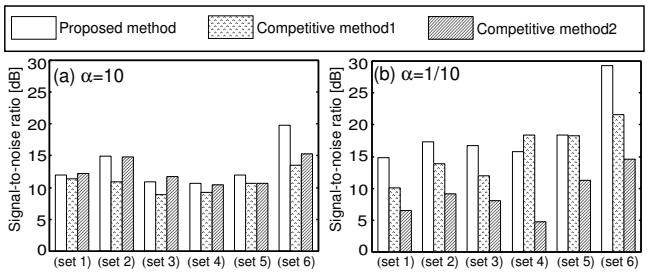

Fig. 4. Result of objective evaluation

Table 1. Rating Scheme

Score	Impairment
0	Imperceptible
-1	Perceptible but not annoying
-2	Slightly annoying
-3	Annoying
-4	Very annoying

index of the sources, $m = 1, 2$ denotes index of stereo channels and n is the index of samples. In addition, by modifying the amplitude of each separated track in each channel, the target processed signal $t_{lm}(n)$ is obtained. As an evaluation score, we used the SNR of each source evaluating the power ratio of the target and the error of processing given by

$$\mathrm{SNR}_l = 10\log_{10} \sum_n \frac{|t_{l1}(n)|^2 + |t_{l2}(n)|^2}{|t_{l1}(n) - \hat{s}_{l1}(n)|^2 + |t_{l2}(n) - \hat{s}_{l2}(n)|^2}. \quad (19)$$

We evaluated the averaged SNR of the sources.

The result of the objective evaluation is shown in Fig. 4. In both of the parameter settings, the performance of the proposed method shows the best performances. In contrast, the performance of the conventional methods changes depending on the parameters. Thus the proposed method can achieve stable controllability of localization for any parameter setting.

5.3. Subjective Evaluation

We evaluated the ability of desired control from the viewpoints of source localization ability and sound quality in the subjective evaluation.

In the evaluation of localization, the two stimuli selected from different methods are presented in random order, and the subjects select the better one to fit the purpose of the processing. In the evaluation of sound quality, the processed signals are presented in a random order followed by the presentation of the unprocessed signals, and the subjects evaluate the degradation of the sound quality. The stimuli are given with headphones. The subjects consisted of eight males and a female. Table 1 shows the rating scheme. We show the results of the subjective evaluation in Fig. 5.

The filter design of competitive method 1 with a single fixed filter coefficient in a frequency bin assumes the existence of only a single source in a frequency bin through all the analysis frames. Thus, in the frequency subbands where multiple sources exist, this method cannot modify the inter-channel level difference of each of the sources separately. In addition, the application of the single-channel filter for each of the channels causes colorization to degrade the quality.

The time-varying filtering of competitive method 2 assumes the existence of only a single source in each time-frequency grid, which is often satisfied. However, in the time-frequency grids where the assumption is not satisfied, this method causes musical noise similar to the Wiener filter and time-frequency binary masking [11] and the degradation of quality is more significant than in competitive method 1.

In contrast, as discussed in Sect. 4.2, the proposed method has a mechanism to analyze the localization information sufficiently even when the sparseness assumption does not hold. Thus the proposed method can control sound localization without degrading sound quality. The proposed method shows the best performance in controllability in both settings, and the degradation of sound quality is not significant. As a result, it is ascertained that the proposed method can control sound localization of stereo audio signals with multiple sources sufficiently.

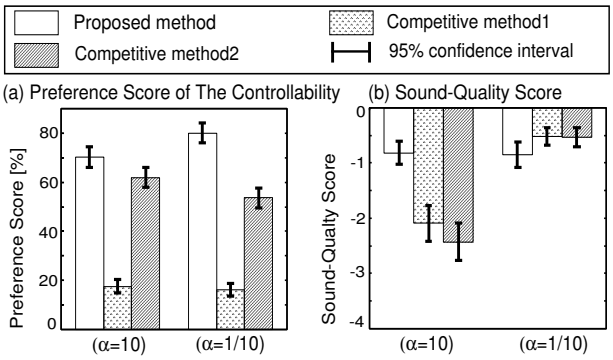

Fig. 5. Results of subjective evaluation

6. CONCLUSION

In this paper, first, we proposed a localization information analysis method with low distortion. Next, we proposed a localization control method of stereo audio signal consisting of multiple sources. The efficacy of the proposed method is ascertained in the objective and subjective evaluations.

The processing of the proposed method and the punch in described in Sec. 3.1 is demonstrated in the following URL.
`http://spalab.naist.jp/database/Demo/slc/`

7. REFERENCES

[1] J. Herre, S. Disch J. Hilpert, and O. Hellmuth, "From SAC to SAOC – Recent developments in parametric coding of spatial audio," *Proc. AES 22nd UK Conf.,* 2007.

[2] C. Faller and F. Baumgarte, "Binaural Cue Coding—Part II: Schemes and Applications," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 6, pp. 520–531, 2003.

[3] J. Blauert, Spatial Hearing, MIT Press, Cambridge, MA, 1997.

[4] O. Gillet and G. Richard, "Extraction and remixing of drum trucks from polyphonic music signals," *Proc. WASPAA*, pp. 315–318, 2005.

[5] C. Avendano, "Frequency-domain source identification and manipulation in stereo mixes for enhancement, suppression and repanning applications," *Proc. WASPAA,* pp. 55–58, 2003.

[6] P. Comon, "Independent component analysis—A new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[7] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, "Blind source separation combining independent component analysis and beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1135–1146, 2003.

[8] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency domain blind source separation," *IEICE Trans. Fundam.*, vol. E86-A, no. 3, pp. 590–596, 2003.

[9] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. NOLTA'98*, pp. 923–926, 1998.

[10] P. Bofill, "Underdetermined blind separation of delayed sound sources in the frequency domain," *Neurocomputing*, vol. 55, pp. 627–641, 2003.

[11] S. Ben Jebara, "A perceptual approach to reduce musical noise phenomenon with wiener denoising technique," *Proc. ICASSP*, pp. III-49–III-52, 2006.