Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Conversion of reproduced sound field based on the coincidence of sound pressure and direction of particle velocity

Akio Ando (1)

(1) NHK Science and Technology Research Laboratories, 1-10-11 Kinuta, Setagata, Tokyo, 157-8510, Japan

PACS: 43.38.Md, 43.38.Vk, 43.60.Dh, 43.60.Jn

ABSTRACT

Several new sound systems have been proposed to provide an enhanced spatial impression in comparison with 5.1 surround sound. Such systems require more loudspeakers than the 5.1 system to deliver a superior sound impression, but it may be difficult to set up these systems into the typical home environment. This paper describes a new method for converting the signal of the original sound system into that of an alternative system with a different number of channels, while maintaining the physical properties of sound at the listening point in the reproduced sound field. 22-channel signals of a 22.2 multi-channel sound system without the two low frequency effect channels were converted into 10-, 8-, and 6-channel signals with the method. Subjective evaluations showed that the converted 8-channel sound gave almost the same spatial impressions as the original 22-channel sound, meaning that the proposed method could reproduce the original 22-channel sound field with 8 loudspeakers.

INTRODUCTION

Broadcasting and packaged media have popularized 5.1 multi-channel sound as a home sound system. Research on multi-channel audio has now shifted to advanced systems having more channels to provide enhanced spatial impressions. For example, a 22.2 multi-channel sound system has been developed for ultra-high definition television with 4320 scanning lines [1]. Such advanced sound systems require their own loudspeaker arrangement to bring out the best performance. Whereas loudspeakers can be arranged optimally in a theater, they are difficult to set up in the home environment.

"Down-mixing" is a widely known way of reducing the number of channels in multi-channel audio. Down-mixing from 5.1 to two-channel stereo or monophonic has already been standardized in an ITU-R Recommendation [2] and is equipped on some television receivers. Although such a down-mixing algorithm was reported as having a certain effectiveness [3], it does not work for an arbitrary loudspeaker arrangement[1]. To enable down-mixing between a number of systems, a technology for converting or recreating sound fields is necessary.

This paper refers the sound field reproduced by the original sound system as the "original" sound field and the sound field recreated by an alternative system with a different number of channels as the "reproduction" sound field. The problem of recreating the primary sound field, such as in a concert hall, is a task of a recording engineer and is not dealt with in this paper. Our problem of recreating the original sound field in another room is one of finding a good conversion from the sound signal of the original sound system to that of the reproduction sound system. For that, we aim for faithful reproduction of (i) timbre, (ii) sound localization, and (iii) the sound envelopment of the original sound [4]. Regarding requirement (i), the conversion discussed in this paper should not be a function of frequency, because a frequency-dependent conversion might change the timbre.

Many sound field reconstruction systems have been proposed [5][6][7][8][9][10][11]. Among them, Ambisonics [6][9][10] seems to meet conditions (i) and (iii). Ambisonics, however, may generate a negative conversion coefficient which brings out an opposite-phase signal that would degrade the localization of the reproduced sound [12]. Consequently, we decided to develop a new conversion method satisfying all three requirements.

This paper presents a new approach to conversion that is applicable to automatic down-mixing and up-mixing. The basic idea is to solve the conversion matrix that converts the input signals for loudspeakers in the original sound field into those in the reproduction sound field in such a way that the physical properties of sound coincide at the receiving point (listening point). This problem can be solved analytically and the solution is frequency independent, meaning that the conversion does not change the timbre. In addition, our method does not have any problems with the opposite phase and hence maintains sound localization.

[1] The down-mixing method standardized in [2] uses only three down-mixing coefficients, i.e. 1.0, 0.7071, and 0.0.

FORMULATION OF SOUND FIELD CONVERSION

The proposed method converts the original signal $\boldsymbol{s}(t)$ into the converted signal $\boldsymbol{q}(t)$ through matrix operation

$$\boldsymbol{q}(t) = W\boldsymbol{s}(t) \,, \tag{1}$$

where

$$\boldsymbol{s}(t) = \begin{pmatrix} s_1(t) \\ \vdots \\ s_n(t) \end{pmatrix}, \quad \boldsymbol{q}(t) = \begin{pmatrix} q_1(t) \\ \vdots \\ q_m(t) \end{pmatrix} .$$

The matrix W should be solved such that the same physical properties at the receiving point in the original and reproduction sound field are the same. Figure 1 shows the block diagram of the proposed method.

The following assumptions are made: (1) each loudspeaker can be modeled as a point source; (2) the sound pressure at a unit distance from a loudspeaker is in proportion to the input to the loudspeaker (the proportionality coefficient is denoted as G); (3) only the outgoing wave from the loudspeaker is considered; and (4) reflected sound can be neglected in the original and converted field. Furthermore, (5) $k\sigma_{\min} \ll 1$ is assumed, where k is the wave number and $\sigma_{\min}$ is the minimum distance between the loudspeakers and receiving point. Assumption (5) is valid except for the low frequency sound that does not contribute to perception of sound localization.

It is well known that sound can be described by two physical properties: sound pressure and particle velocity [13]. The sound pressure is a scalar variable of the time and location. On the other hand, the particle velocity is, in general, a three-dimensional variable. Note that both physical properties are linear functions of the source signal. Based on the assumptions mentioned above, if a loudspeaker whose input signal is $s(t)$ is located at $\boldsymbol{\xi}=(\ \xi_x\ \xi_y\ \xi_z)^T$, the Fourier transforms of sound pressure and particle velocity at the receiving point $\boldsymbol{r}$ $=(x\ y\ z)^T$ can be written as

$$p(\omega) = G\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\xi}|}}{|\boldsymbol{r}-\boldsymbol{\xi}|}s(\omega) \tag{2}$$

and

$$\begin{aligned} \boldsymbol{u}(\omega) &= G\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\xi}|}}{|\boldsymbol{r}-\boldsymbol{\xi}|}\left(1+\frac{1}{ik|\boldsymbol{r}-\boldsymbol{\xi}|}\right)\frac{1}{|\boldsymbol{r}-\boldsymbol{\xi}|}\begin{pmatrix} x-\xi_x \\ y-\xi_y \\ z-\xi_z \end{pmatrix} s(\omega) \\ &\approx G\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\xi}|}}{|\boldsymbol{r}-\boldsymbol{\xi}|^2}\begin{pmatrix} x-\xi_x \\ y-\xi_y \\ z-\xi_z \end{pmatrix} s(\omega) \end{aligned} \tag{3}$$

respectively.

Suppose there are n loudspeakers in the original field, whose locations are $\boldsymbol{\xi}^{(j)}=(\ \xi_x^{(j)}\ \xi_y^{(j)}\ \xi_z^{(j)})^T$, j=1,…,n, and m loudspeakers in the reproduction, whose locations are $\boldsymbol{\zeta}^{(j)}=(\ \zeta_x^{(j)}\ \zeta_y^{(j)}\ \zeta_z^{(j)})^T$, j=1,…,m. Let the input signal of loudspeakers in the original space be $s_j(t)$, j=1,…,n, and those in the reproduction space be $q_j(t)$, j=1,…,m; the Fourier transforms of sound pressure and particle velocity at the receiving point in the original field are represented as

$$p(\omega) = G\sum_{j=1}^{n}\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\xi}^{(j)}|}}{|\boldsymbol{r}-\boldsymbol{\xi}^{(j)}|}s_j(\omega) \tag{4}$$

$$\boldsymbol{u}(\omega) = G\sum_{j=1}^{n}\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\xi}^{(j)}|}}{|\boldsymbol{r}-\boldsymbol{\xi}^{(j)}|^2}\begin{pmatrix} x-\xi_x^{(j)} \\ y-\xi_y^{(j)} \\ z-\xi_z^{(j)} \end{pmatrix} s_j(\omega) \ , \tag{5}$$

and those in the reproduction field are

$$\tilde{p}(\omega) = G\sum_{j=1}^{m}\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\zeta}^{(j)}|}}{|\boldsymbol{r}-\boldsymbol{\zeta}^{(j)}|}q_j(\omega) \tag{6}$$

$$\tilde{\boldsymbol{u}}(\omega) = G\sum_{j=1}^{m}\frac{e^{-ik|\boldsymbol{r}-\boldsymbol{\zeta}^{(j)}|}}{|\boldsymbol{r}-\boldsymbol{\zeta}^{(j)}|^2}\begin{pmatrix} x-\zeta_x^{(j)} \\ y-\zeta_y^{(j)} \\ z-\zeta_z^{(j)} \end{pmatrix} q_j(\omega) \ . \tag{7}$$

To obtain an analytic solution in which the physical properties of sound coincide at the receiving point, the proposed method finds a local solution creating a phantom source of an original loudspeaker at the corresponding position in the reproduction field. The phantom source is generated by three loudspeakers adjacent to the source position. Figure 2 shows this step of the method. The method then creates a global solution W by summing up such local solutions for all loudspeakers of the original field.

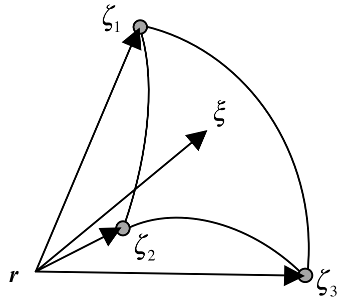


Figure 2. Generation of a phantom source at $\boldsymbol{\xi}$ (position of a loudspeaker in the original field) with three loudspeakers located at $\boldsymbol{\zeta}_1$, $\boldsymbol{\zeta}_2$ and $\boldsymbol{\zeta}_3$ in the reproduction field

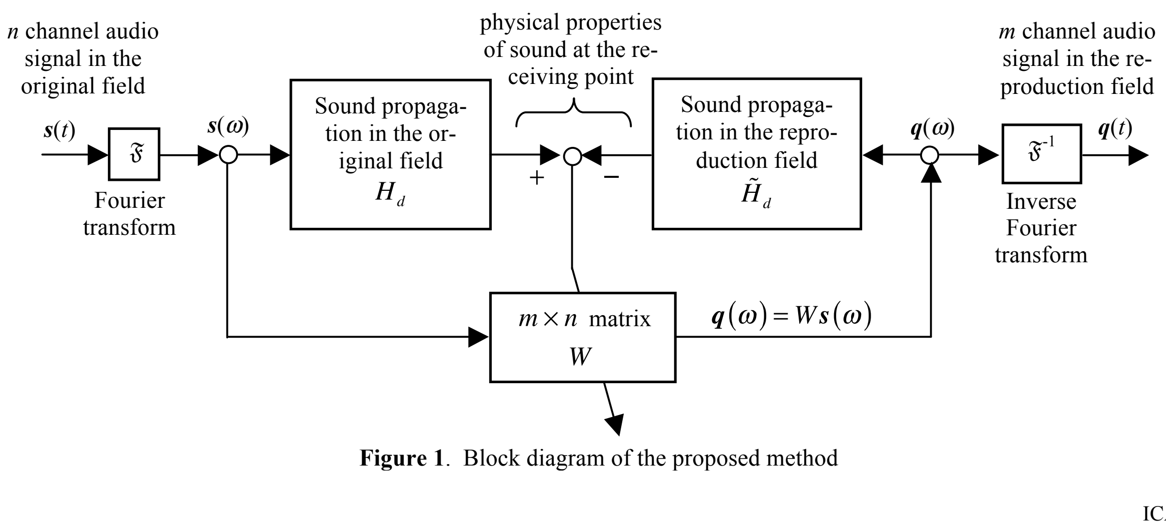


Figure 1. Block diagram of the proposed method

ANALITIC SOLUTION

Solution for generating a phantom source

We shall use polar coordinates originating at the receiving point. Assuming that a loudspeaker is located at ($\sigma\ \theta\ \phi$) in the original space and its input signal is $s(t)$, the Fourier transform of the particle velocity observed at the receiving point is

$$\boldsymbol{u}(\omega) = -\frac{G}{\rho c}\boldsymbol{h}s(\omega) \tag{8}$$

where

$$\boldsymbol{h} \triangleq \frac{e^{-ik\sigma}}{\sigma}\begin{pmatrix}\cos\theta\cos\varphi\\ \sin\theta\cos\varphi\\ \sin\varphi\end{pmatrix}.$$

Here, σ is the distance between the source and receiving point, θ the azimuthal angle, and ϕ the elevation angle. On the other hand, if there are three loudspeakers in the reproduction space, each of whose input signal and location are $w_l s(t)$ and $(\tilde{\sigma}_l\ \tilde{\theta}_l\ \tilde{\varphi}_l)$, respectively ($l$=1,..,3), the Fourier transform of the observed particle velocity is

$$\tilde{\boldsymbol{u}}(\omega) = -\frac{G}{\rho c}\tilde{H}\boldsymbol{w}s(\omega) \tag{9}$$

where

$$\tilde{H} \triangleq \begin{pmatrix} \frac{e^{-ik\tilde{\sigma}_1}}{\tilde{\sigma}_1}\cos\tilde{\theta}_1\cos\tilde{\varphi}_1 & \frac{e^{-ik\tilde{\sigma}_2}}{\tilde{\sigma}_2}\cos\tilde{\theta}_2\cos\tilde{\varphi}_2 & \frac{e^{-ik\tilde{\sigma}_3}}{\tilde{\sigma}_3}\cos\tilde{\theta}_3\cos\tilde{\varphi}_3 \\ \frac{e^{-ik\tilde{\sigma}_1}}{\tilde{\sigma}_1}\sin\tilde{\theta}_1\cos\tilde{\varphi}_1 & \frac{e^{-ik\tilde{\sigma}_2}}{\tilde{\sigma}_2}\sin\tilde{\theta}_2\cos\tilde{\varphi}_2 & \frac{e^{-ik\tilde{\sigma}_3}}{\tilde{\sigma}_3}\sin\tilde{\theta}_3\cos\tilde{\varphi}_3 \\ \frac{e^{-ik\tilde{\sigma}_1}}{\tilde{\sigma}_1}\sin\tilde{\varphi}_1 & \frac{e^{-ik\tilde{\sigma}_2}}{\tilde{\sigma}_2}\sin\tilde{\varphi}_2 & \frac{e^{-ik\tilde{\sigma}_3}}{\tilde{\sigma}_3}\sin\tilde{\varphi}_3 \end{pmatrix}$$

and

$$\boldsymbol{w} = \begin{pmatrix} w_1\\ w_2\\ w_3\end{pmatrix}.$$

A virtual source of the original loudspeaker is localized at (σ θ ϕ) in the reproduction space if

$$\tilde{H}\boldsymbol{w} = \boldsymbol{h} \tag{10}$$

holds. A local solution $\hat{\boldsymbol{w}}_u$ can be then obtained as

$$\hat{\boldsymbol{w}}_u = \tilde{H}^{-1}\boldsymbol{h} = \begin{pmatrix} \frac{\tilde{\sigma}_1}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_1)}\frac{\tilde{D}_1}{\tilde{D}} \\ \frac{\tilde{\sigma}_2}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_2)}\frac{\tilde{D}_2}{\tilde{D}} \\ \frac{\tilde{\sigma}_3}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_3)}\frac{\tilde{D}_3}{\tilde{D}} \end{pmatrix}, \tag{11}$$

where

$$\begin{aligned}\tilde{D} =& \sin(\tilde{\theta}_3-\tilde{\theta}_2)\cos\tilde{\varphi}_3\cos\tilde{\varphi}_2\sin\tilde{\varphi}_1 + \sin(\tilde{\theta}_1-\tilde{\theta}_3)\cos\tilde{\varphi}_1\cos\tilde{\varphi}_3\sin\tilde{\varphi}_2 \\ &+ \sin(\tilde{\theta}_2-\tilde{\theta}_1)\cos\tilde{\varphi}_2\cos\tilde{\varphi}_1\sin\tilde{\varphi}_3\end{aligned}$$

$$\begin{aligned}D_1 =& \sin(\tilde{\theta}_3-\tilde{\theta}_2)\cos\tilde{\varphi}_3\cos\tilde{\varphi}_2\sin\tilde{\varphi} \\ &+ \{\sin(\theta-\tilde{\theta}_3)\cos\tilde{\varphi}_3\sin\tilde{\varphi}_2 - \sin(\theta-\tilde{\theta}_2)\cos\tilde{\varphi}_2\sin\tilde{\varphi}_3\}\cos\varphi \\ D_2 =& \sin(\tilde{\theta}_1-\tilde{\theta}_3)\cos\tilde{\varphi}_1\cos\tilde{\varphi}_3\sin\tilde{\varphi} \\ &+ \{\sin(\theta-\tilde{\theta}_1)\cos\tilde{\varphi}_1\sin\tilde{\varphi}_3 - \sin(\theta-\tilde{\theta}_3)\cos\tilde{\varphi}_3\sin\tilde{\varphi}_1\}\cos\varphi \\ D_1 =& \sin(\tilde{\theta}_2-\tilde{\theta}_1)\cos\tilde{\varphi}_2\cos\tilde{\varphi}_1\sin\tilde{\varphi} \\ &+ \{\sin(\theta-\tilde{\theta}_2)\cos\tilde{\varphi}_2\sin\tilde{\varphi}_1 - \sin(\theta-\tilde{\theta}_1)\cos\tilde{\varphi}_1\sin\tilde{\varphi}_2\}\cos\varphi\end{aligned}$$

Solution (11) guarantees coincidence of the particle velocities of the fields. Now let us introduce a condition for the coincidence of sound pressure. In the current case, the Fourier transform of the sound pressure at the receiving point due to a loudspeaker at (σ θ ϕ) is

$$p(\omega) = G\frac{e^{-ik\sigma}}{\sigma}s(\omega) \tag{12}$$

and that due to three loudspeakers at $(\tilde{\sigma}_j\ \tilde{\theta}_j\ \tilde{\varphi}_j)$, ($j$=1,..,3) is

$$\tilde{p}(\omega) = G\left(\sum_{j=1}^{3}\frac{e^{-ik\tilde{\sigma}_j}}{\tilde{\sigma}_j}w_j\right)s(\omega). \tag{13}$$

From Eqs. (12) and (13), the coincidence of sound pressures is obtained when

$$\frac{e^{-ik\tilde{\sigma}_1}}{\tilde{\sigma}_1}w_1 + \frac{e^{-ik\tilde{\sigma}_2}}{\tilde{\sigma}_2}w_2 + \frac{e^{-ik\tilde{\sigma}_3}}{\tilde{\sigma}_3}w_3 = \frac{e^{-ik\sigma}}{\sigma} \tag{14}$$

is true. The solution $\hat{\boldsymbol{w}}_u$ does not satisfy condition (14) because substituting (11) into (14) yields an incorrect equation:

$$\frac{\tilde{D}_1+\tilde{D}_2+\tilde{D}_3}{\tilde{D}} = 1.$$

On the other hand, a new vector

$$\hat{\boldsymbol{w}} = \begin{pmatrix} \frac{\tilde{\sigma}_1}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_1)}\frac{\tilde{D}_1}{D} \\ \frac{\tilde{\sigma}_2}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_2)}\frac{\tilde{D}_2}{D} \\ \frac{\tilde{\sigma}_3}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_3)}\frac{\tilde{D}_3}{D} \end{pmatrix} \tag{15}$$

satisfies condition (14), where

$$D = \tilde{D}_1 + \tilde{D}_2 + \tilde{D}_3. \tag{16}$$

Although solution (15) does not satisfiy condition (10), it guarantees coincidence of the directions of particle velocities. Therefore, solution (15) gives the same sound pressure and sound direction in both fields, and we chose it as the local solution.

The obtained solution depends not on the input signal, but rather on the positions of the loudspeakers and receiving point in both fields. One of the obtained signals is

$$q_l(\omega) = \frac{\tilde{\sigma}_l}{\sigma}e^{-ik(\sigma-\tilde{\sigma}_l)}\frac{\tilde{D}_l}{D}s(\omega)$$

in the frequency domain. The time-domain representation is

$$q_l(t) = \frac{\tilde{\sigma}_l}{\sigma}\cdot\frac{\tilde{D}_l}{D}s\left(t-\frac{\sigma-\tilde{\sigma}}{c}\right). \tag{17}$$

From Equation (17), it is clear that the conversion is not frequency dependent and satisfies requirement (i).

Global solution

A global solution W ($m\times n$ matrix) can be obtained by putting each element of local solutions into its appropriate portion. Let the loudspeaker of the j-th channel in the original space be included in the area bounded by three loudspeakers of the $l_1(j)$-, $l_2(j)$- and $l_3(j)$-th channels in the reproduction field. The elements of the local solution in this case are denoted as

$$\hat{w}_{l_1(j)j},\ \hat{w}_{l_2(j)j},\ \hat{w}_{l_3(j)j}.$$

Using this notation, we define w_{lj} as

$$w_{lj} = \begin{cases} \hat{w}_{l_1(j)j}; & l = l_1(j), \\ \hat{w}_{l_2(j)j}; & l = l_2(j), \\ \hat{w}_{l_3(j)j}; & l = l_3(j), \\ 0; & l \in \Gamma_j, \end{cases}$$

where $\Gamma_j = \{1,2,\cdots m\} - \{l_1(j), l_2(j), l_3(j)\}$. Accordingly, the matrix $W = \{w_{lj}\}$ becomes the global solution because the sound pressure and particle velocity are linear functions of the input signal.

Instead of the sound pressure and particle velocity, we could use the sound intensity [14] as the physical property of sound and obtain an alternative solution of (15). Although doing so is valid for the local solution, the global solution cannot be obtained by simple addition of local solutions, because the sound intensity is a quadratic variable of the original signal. That is the reason we selected the sound pressure and the direction of the particle velocity as the physical properties of sound.

SUBJECTIVE EVALUATION

Two subjective experiments were carried out in order to evaluate the proposed method. In the first experiment, a 22.2 sound signal without the two low frequency effect (LFE) channels was converted into a 10-channel signal. The converted sound was compared with the original sound by subjective evaluation. In the second experiment, a 22-channel signal was converted into 8- and 6-channel ones and the converted sound was compared with the original. Figure 3 shows the loudspeaker arrangement of the 22-channel used in the experiments. All experiments were carried out in a sound proof room where the reverberation time at 500Hz was 0.18 s. The distance between the listening point and each loudspeaker was 2 m. The loudspeaker triplets were manually decided in both experiments.

Eight sound stimuli were taken from 22.2 multi-channel programs exhibited at the World Expo 2005 held in Aichi, Japan and the NAB 2006 and 2007 shows held in Las Vegas, the United States. Each stimulus was from 10 seconds to 12 seconds long. They included musical sound, sounds in a sport, birds singing, and the sound of a light breeze.

Subjective evaluation method

The subjective evaluations used the "double-blind triple-stimulus with hidden reference" method [15], which can be used for subjective assessment of small impairments in a multi-channel sound system. Figure 4 shows the procedure. In Fig. 4, stimulus "R" indicates the reference sound and stimuli "A" and "B" the sounds for evaluation. The subject was asked to assess the impairment on "A" and "B" compared to "R", according to a continuous five-grade impairment scale shown in Table 1. One of the stimuli, "A" or "B", was the same sound as "R". The stimulus the same as the reference is referred the "hidden reference", and the other stimulus the "object". In the experiment, the reference was the original 22-channel sound and the object the converted sound. The impairment was assessed from two points of view, sound localization and sound envelopment. After the experiment, the "differential grade" was calculated for each object by subtracting the grade given to the hidden reference from that to the object, and therefore, it should be a non-positive value.

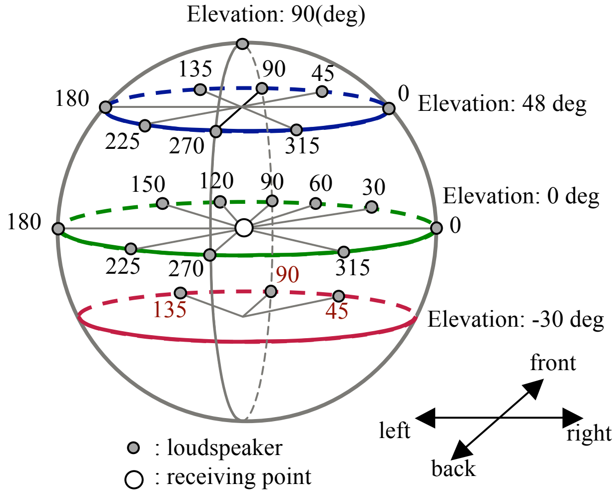


Figure 3. 22.2 multi-channel loudspeaker arrangement without LFE channels

Table 1. Scales used for the subjective evaluation

Impairment	Grade
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0

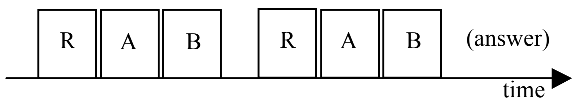


Figure 4. Procedure of "double-blind triple-stimulus with hidden reference" method

Experiment 1: Conversion from 22-channel sound to 10-channel sound

The three loudspeaker arrangements shown in Figure 5 were used to convert the 22-channel signal into a 10-channel signal. Layout 1 had 4 loudspeakers in the top layer, 5 loudspeakers in the middle layer, and 1 loudspeaker in the bottom layer. Layout 2 had 3, 6, and 1 loudspeakers in the top, middle, and bottom layer, and layout 3 had 3, 5, and 2 loudspeakers in the top, middle, and bottom layer. Subjects were 38 people in their twenties, thirties, and forties and who were experienced in playing musical instruments. They evaluated each conversion twice. Eight sound stimuli were used in the experiment.

The results are shown in Figure 6. P-1, P-2, and P-3 in Fig. 6 show the proposal method's results for layouts 1, 2, and 3, respectively. D-2 and D-3 show the results of the conventional down-mixing algorithm where each original signal was converted with the coefficients of 1.0, 0.5, and 0.0 satisfying the coincidence of sound pressures at the receiving point.

Layout 1 was every other loudspeaker of the original 22-channel, and hence, the conversion matrix of the conventional down-mixing was almost the same as that of the proposed method. For example, if we use the notation "(azimuthal angle, elevation angle)" for the channel location, both methods evenly distributed a channel located at (90, 0) in the original space to two channels located at (120, 0) and (60, 0) in the reproduction space. Because of this, the conventional algorithm was not evaluated for layout 1. The proposed method obtained a differential grade of more than -1.0 for both spatial impressions. Thus, even with 10 channels, the proposed method kept the spatial impressions of the original 22-channel sound. It gave a better result than the conventional one, and the difference between the two methods was significant at a level of 0.05.

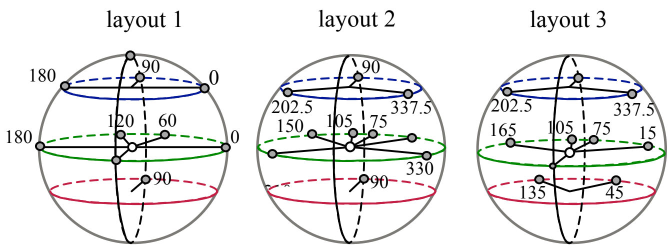
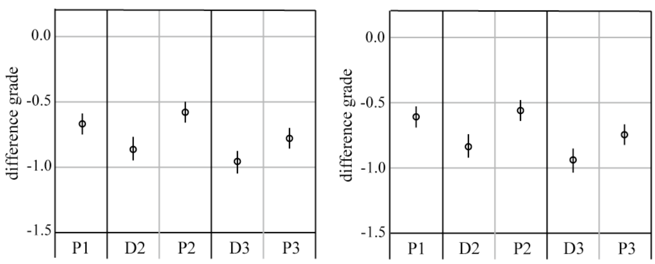


Figure 5. 10-channel loudspeaker arrangements in the reproduction space



(a) Sound envelopment (b) Sound localization

Figure 6. Results of experiment 1

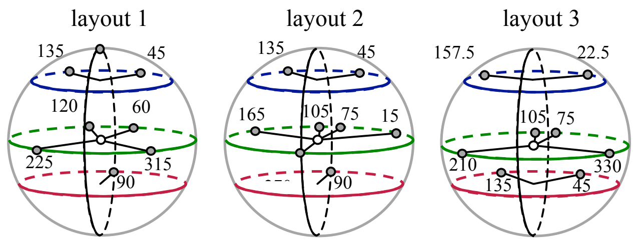
○ :mean scores, | :95% confidence limits

P1: Proposed, layout 1
D2: Conventional, layout 2 P2: Proposed, layout 2
D3: Conventional, layout 3 P3: Proposed, layout 3

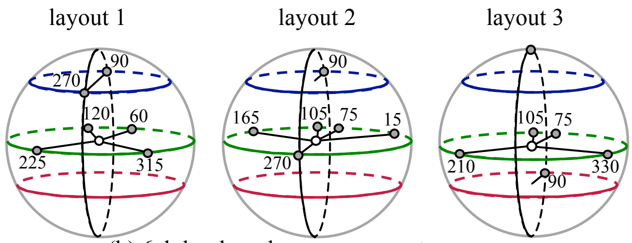
Experiment 2: Conversion from 22-channel sound to 8- and 6- channel sound

In the first experiment, the proposed method obtained a differential grade of more than -1.0 in converting 22-channel sound into 10-channel sound. The question arose as to how many channels in the reproduction space can keep a differential grade of more than -1.0 for both spatial impressions. Hence, we conducted the second experiment in which the original 22-channel signal was converted into 8- and 6-channel sound. We set up three loudspeaker arrangements for both conversions as shown in Figure 7. The subjects were 32 people in their twenties, thirties, and forties and who were experienced in playing musical instruments. They evaluated each conversion twice. As the loudspeaker arrangements were twice that of the first experiment, we picked up the representative four stimuli and used them for this experiment.

The other experimental conditions were the same in the first experiment.

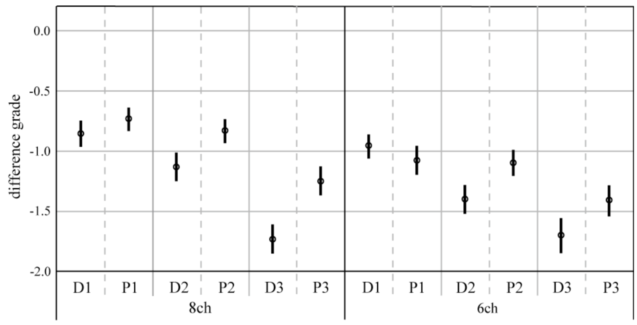


(a) 8ch loudspeaker arrangements

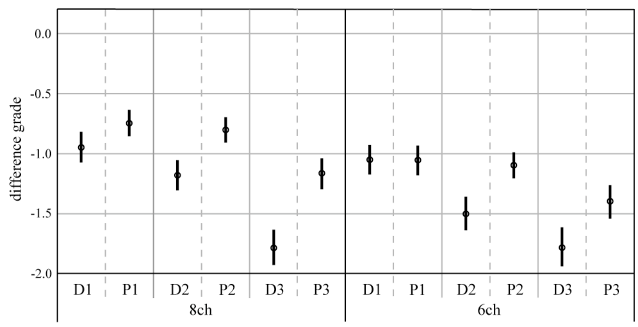


(b) 6ch loudspeaker arrangements

Figure 7. 8- and 6-ch loudspeaker arrangements in the reproduction space



(a) Sound envelopment



(b) Sound localization

Figure 8. Results of experiment 2

○ :mean scores, | :95% confidence limits

D1: Conventional, layout 1 P1: Proposed, layout 1
D2: Conventional, layout 2 P2: Proposed, layout 2
D3: Conventional, layout 3 P3: Proposed, layout 3

The results are shown in Figure 8. Regarding layout 1 for the 8- and 6-channel loudspeaker arrangements, the difference between the conventional down-mixing method and the proposed method is not significant at a level of 0.05. The reason is likely that both methods distributed the sound of the front channels of the original 22-channel in almost the same manner. For the other layouts, the proposed method yielded better results than conventional down-mixing did on both spatial impressions, and the difference was significant at a level of 0.05. The converted 8-channel sound in layouts 1 and 2 gave spatial impressions of more than -1.0 in difference grade (layout 3 had only a few loudspeakers in the frontal area). On the other hand, the converted 6-channel sound did not maintain a difference grade of more the -1.0.

DISCUSSION

The proposed method guarantees coincidence of pressures and directions of sound only at the receiving point. It can be extended into a method minimizing the square errors of those sound properties over a receiving area (a infinite set of receiving points) [16]. In such case, however, the obtained solution will inevitably be a frequency-dependent function and would not satisfy requirement (i) mentioned in Section I. Although the proposed method ensures coincidence of the physical properties only at a receiving point, the subjective evaluation results showed that it is effective in the nearby area because the both ears would not be on the receiving point at the same time. We shall study the size of the effective listening area reproduced by the proposed method.

The down-mixing method proposed in [2] has coefficients based on the conservation of sound energy. On the other hand, the conventional down-mixing used in this paper had coefficients based on the conservation of sound pressure, so as to match the proposed method. The local solution of the proposed method can be naturally made to conserve energy by modifying Eq. (14). In such case, however, the global solution cannot be obtained by adding the local solutions for the same reason as in the coincidence of sound intensity.

Notwithstanding the above remarks, the down-mixing method described in this paper performed very well. There would be no reason for these results other than it guarantees coincidence of the sound pressures. The principal difference between the conventional and proposed methods is the reproducibility of the sound direction. A subjective evaluation of sound materials having a clear sound direction would widen the difference between two methods.

In the experiments, the loudspeaker triplets were manually decided. An algorithm automatically dividing the reproduction space into the subspaces with the triplets of the loudspeaker positions has not been completed yet. We shall study this issue in due course.

CONCLUSION

This paper proposed a new method for converting multichannel sound signals while maintaining the pressure and direction of sound at the receiving point in the reproduced sound field. We found that the conventional down-mixing method would be effective if the sound pressure were conserved. Even in such cases, the proposed method performed better than the conventional method because it can reproduce the sound direction. Subjective evaluations revealed that 8-channel sound converted with the proposed method gave almost the same spatial impressions as the original 22-channel sound at a receiving point.

REFERENCES

1 K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando, “A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV),” *SMPTE Motion Imaging J.* pp. 40-49, Apr. (2008)

2 ITU-R Rec. BS.775-2, “Multi-channel stereophonic sound system with or without accompanying picture,” ITU, Geneva, Switzerland (2006)

3 S. K. Zielinski, F. Rumsey, and S. Bech, “Effects of down-mix algorithms on quality of surround sound,” *J. Audio Eng. Soc*. **51**, no. 9, pp.780-798, Sep. (2003)

4 F. Rumsey, “Spatial Audio,” Focal Press (2001)

5 M. Camras, “Approach to recreating sound field,” *J. Acoust. Soc. Am*. **43**, no. 6, pp.1425-1431, June (1968)

6 M. A. Gerson, “Periphony: With-height sound reproduction,” *J. Audio Eng. Soc.* **21**, no.1, pp.2-10, Jan./Feb. (1973)

7 A. J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” *J. Acoust. Soc. Am*. **93**, no.5, pp.2764-2778, May (1993)

8 S. Ise, “A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems,” *Acustica*, **85**, pp.78-87 (1999)

9 R. Nicol, and M. Emerit, “3D-sound reproduction over an extensive listening area: a hybrid method derived from holophony and ambisonic,” *AES 16th International Conference* (1999)

10 D. B. Ward, and T. D. Abhayapala, “Reproduction of a plane-wave sound field using an array of loudspeakers,” *IEEE Trans. Speech and Audio Proc*. **9**, no. 6, pp.687-707, Sep. (2001)

11 F. M. Fazi, P. A. Nelson, J. E. N. Christensen, and J. Seo, “Surround system based on three dimensional sound field reconstruction,” *AES 125th Convention Paper*, Oct. (2008)

12 G. Martin, W. Woszczyk, J. Corey, and R. Quesnel, “Controlling Phantom Image Focus in a Multichannel Reproduction System,” *AES 107th Convention Paper*, Sep. (1999)

13 A. D. Pierce, “Acoustics, An Introduction to Its Physical Principles and Applications,” *Acoustical Society of America* (1989)

14 F. J. Fahy, “Sound Intensity, 2nd Edition,” *E & FN Spon* (1995)

15 ITU-R Rec. BS.1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” ITU, Geneva, Switzerland (1997)

16 A. Ando, “Adaptation of multichannel sound reproduction to restricted speaker arrangement,” *Proc. 19th Internatinal Congress on Acoustics* (*ICA2007*), ELE-01-002 (2007)