Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **6**, November **2017**

Estimation of Individual Head-Related Impulse Responses from Impulse Responses Acquired in Ordinary Rooms based on the Spatial Principal Components Analysis

Shouichi Takane

Department of Electronics and Information Systems
Akita Prefectural University
84-4 Ebinokuchi, Tsuchiya, Yurihonjo, Akita, 015-0055 Japan
takane@akita-pu.ac.jp

Received February 2017; revised September 2017

Abstract. *This paper proposes a method to estimate people's Head-Related Impulse Responses (HRIRs) by using responses acquired in ordinary room based on Spatial Principal Component Analysis (SPCA). The average vector and the principal component matrix are obtained by adapting the SPCA to a set of HRIRs of multiple subjects in all directions. Part of the impulse response from the sound source to a subject's ear is used to estimate the coefficients of the weights of the principal components. The application of the proposed method, which uses a dataset containing the HRIRs of multiple subjects from all directions, yielded acceptable estimation accuracy for ipsilateral HRIRs. The proposed method was tested on impulse responses to a dummy head measured in a room.*
Keywords: Head-Related impulse response, Head-Related yransfer function, Spatial principal component analysis

1. Introduction.

Various perceptions can be obtained from a sound. The well-known fundamental attributes are loudness, pitch, and timbre [1]. They are related to perceiving what the sound is. These attributes can be reproduced by using conventional audio systems. Another important attribute is space, such as the positions of the sound sources and reverberation in a listening room. They are related to perceiving where the sound source is. Binaural hearing contributes to these spatial attributes. However, the reproduction of spatial attributes is insufficient, since these attributes are dynamically altered by the movements of the sound sources and listeners. To enrich sound in such contexts, technologies for the reproduction of its attributes are necessary.

In Virtual Reality (VR), devices for sound reproduction are called Virtual Auditory Displays (VADs). Some VADs are based on a synthesis of Head-Related Transfer Functions (HRTFs) or their corresponding inverse Fourier transforms, called Head-Related Impulse Responses (HRIRs). Their synthesis needs to be processed in real time by considering variation due to the movement of a listener and/or the sound sources. The HRTF is defined as the ratio of the transfer function from a sound source to a subject's ear to that from the same source to the center of the subject's head when he/she is absent in the frequency domain [2]. It is a function of the position of the sound source that varies among people owing to differences in the shapes of their heads and ears. A theory of VAD called ADVISE (Auditory Display based on VIrtual SpherE model) has been proposed and implemented at an elemental level [3, 4]. In ADVISE theory, a listener's own HRTFs in all directions are ideally required for the synthesis. The number of sound sources and listeners are assumed to be multiple in practical applications of VADs to synthesis in ordinary listening situations. Various implementations of VADs have been proposed based on the synthesis of binaural sound signals using HRTFs/HRIRs [5, 6, 7, 8, 9].

It is well known that the properties of HRTFs/HRIRs vary among people in both the objective and the subjective senses. Therefore a set of HRTFs/HRIRs for a certain person covering all directions needs to ideally be acquired. However, it is impractical to obtain this through measurement. One reason for this difficulty is that anechoic environment is required to obtain HRTFs/HRIRs for each person, leading to the situation where not everyone can acquire his/her own set of HRTFs/HRIRs. This is a significant limitation for VADs.

For an impulse response from a sound source to a person's ear in a room with reflection and noise, the direct component represents part of the corresponding HRIR. However, the shape of entire HRIRs is sometimes hard to obtain if some reflections or noise are involved. To estimate the reflected components in the impulse response, certain simulation methods, such as the image method and ray tracing method, are applicable [10]. If the reflected components can be accurately estimated by using these methods, the direct component can be obtained by subtracting the estimated components from the measured impulse response. However, accurate estimation requires accurate information concerning the configuration of the room, such as its size and the material of the walls. Three-dimensional (3D) scanners can be used to acquire the shape of the room, setting the boundary condition for the walls a challenging problem in simulating the acoustics. Approaches using data obtained from measurements are needed for this, where one such method applies the data to models of HRTFs/HRIRs. Spherical-mode expansion is one such model [11]. From a set of HRTFs of a certain subject, various modes are obtained in spherical coordinates with the origin at the center of the subject's head. This model is suitable for analyzing the spatial variation in the HRTF because it is based on physical coordinates. However, it is difficult to apply this model to the problem posed in this paper as it is constructed in the wave number domain, where the decomposition of a certain HRTF into its direct and reflected components is challenging if some assumption is introduced together. Takane estimated HRTFs/HRIRs from impulse responses acquired in an ordinary sound field based on the auto regression model [12], but the accuracy of the estimation obtained using this method was not acceptable except small number of sound source directions examined. Spatial Principal Component Analysis (SPCA), sometimes called spatial feature extraction, has been used to analyze and compactly represent the spatial variation of HRTFs/HRIRs. Its theoretical basis is the PCA or the Singular Value Decomposition (SVD) of matrices generated by using HRTFs/HRIRs. In such methods, the spatial variation in the HRTFs is modeled by using a small number of principal components or eigenvectors [13, 14, 15, 16, 17]. Such methods are called the SPCA of the HRTF/HRIR hereinafter, as in Xie [16]. An attraction of the SPCA is that the spatial variation of HRTFs/HRIRs is reconstructed by using a linear combination of fixed principal components. Moreover, the number of principal components required to obtain a certain reconstruction accuracy is relatively small, meaning that the spatial variation is sufficiently approximated by using a relatively small number of principal components. This property is expected to be applicable to the estimation of HRIRs by using a part of the impulse responses acquired by measurements in an ordinary sound field, if the weight coefficients of some major principal components can be properly estimated. These features are common to those of spherical-mode expansion, but the SPCA differs from it in that it can be applied to a set of time-domain impulse responses, *i. e.*, a set of HRIRs.

This paper proposes a method to estimate HRTFs/HRIRs based on SPCA, and the effectiveness of this method is examined through a simulation by using two kinds of HRTF datasets. A simple demonstration of the proposed method is conducted by using impulse responses acquired in a room with reflection. Since processing in the proposed method is carried out in the time domain, only the term "HRIR" is used hereinafter.

2. **Outline of SPCA of HRIRs.** This section outlines the procedure for the SPCA of HRIRs, according to research by Wu *et al.* [17].

1. The average vector of M HRIRs, $\mathbf{h}_m$ $(m = 1, \cdots, M)$, denoted by $\mathbf{h}_{\mathrm{av}}$ (its length denoted by N), is obtained as follows:
$$\mathbf{h}_{\mathrm{av}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{h}_m. \tag{1}$$
M is (the number of subjects) × (the number of positions of the sound source acquired per subject).
2. The covariance matrix for the set(s) of HRIRs is obtained by computing the following equation:
$$\mathbf{R} = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{h}_m - \mathbf{h}_{\mathrm{av}}) \cdot (\mathbf{h}_m - \mathbf{h}_{\mathrm{av}})^{\mathrm{T}}. \tag{2}$$
3. SPCA is carried out by using the matrix $\mathbf{R}$, which means that a set of eigenvalues and their corresponding eigenvectors, $\lambda_m \mathrm{A} \mathbf{q}_m$ $(m = 1, \cdots, M)$, are computed as follows:
$$\mathbf{R} \cdot \mathbf{q}_m = \lambda_m \cdot \mathbf{q}_m. \tag{3}$$
The eigenvalues are sorted in ascending order, *i.e.*, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_M$, and the eigenvectors $\mathbf{q}_k$ are orthonormal to one another. The eigenvectors are also called the principal components (PCs).
4. A weight vector, $\mathbf{w}_m$, to obtain the m-th HRIR, $\mathbf{h}_m$, is computed from the principal component matrix, $\mathbf{Q}$, which has $\mathbf{q}_m (m = 1, \cdots, M)$ in its column vectors as follows:
$$\mathbf{w}_m = \mathbf{Q}^{\mathrm{T}}(\mathbf{h}_m - \mathbf{h}_{\mathrm{av}}). \tag{4}$$
5. By using Eq. (4), the m-th HRIR, $\mathbf{h}_m$, can be reconstructed. Moreover, this equation is approximated such that the corresponding eigenvalues of the first L principal components are the largest. Expressing by $(\mathbf{Q})_L$ and $(\mathbf{w}_m)_L$ the principal component matrix using the first L principal components and the weight vector with components corresponding to the first L principal components, respectively, the approximately reconstructed HRIR, $(\mathbf{h}_m)_L$ is represented as follows:
$$(\mathbf{h}_m)_L \approx (\mathbf{Q})_L \cdot (\mathbf{w}_m)_L + \mathbf{h}_{\mathrm{av}}. \tag{5}$$

3. **Estimating HRIRs from impulse response acquired in ordinary room.** The weight vector $\mathbf{w}_m$ to obtain $\mathbf{h}_m$ using matrix $\mathbf{Q}$ is naturally computed if all components of $\mathbf{h}_m$ are known. In this case, the length of $\mathbf{w}_m$ is equal to the number of principal components N. On the contrary, it is possible to obtain the weight vector to compute the HRIR of a certain subject uninvolved as a data item to construct the $\mathbf{Q}$ matrix. This means that the weight vector, denoted by $\mathbf{w}$, for this HRIR, denoted by $\mathbf{h}$, is obtained by the following equation:
$$\mathbf{w} = \mathbf{Q}^{\mathrm{T}} \cdot (\mathbf{h} - \mathbf{h}_{\mathrm{av}}). \tag{6}$$
There is no guarantee that the HRIR $\mathbf{h}$ is properly reconstructed from the weight vector $\mathbf{w}$ computed by using the above equation. However, an acceptable approximation is expected by using Eq. (6) if the principal component matrix $\mathbf{Q}$ is composed of the covariance matrix

involving the variation in HRIRs due to the variety of subjects and location of the sound sources.

Moreover, a vector of a certain impulse response acquired in a room, denoted by $\mathbf{g}$, involves reflections where the first K samples are regarded as direct components, $K < N$, and the entire vector $\mathbf{g}$ cannot be used to compute the weight vector. Now, vectors containing the first K components of vectors $\mathbf{g}$ and $\mathbf{h}_{\mathrm{av}}$ are, respectively, denoted by $(\mathbf{g})_K$ and $(\mathbf{h}_{\mathrm{av}})_K$; it is also possible to obtain a vector containing K weight coefficients, denoted by $\overline{(\mathbf{w})_K}$ as follows:

$$\overline{(\mathbf{w})_K} = (\mathbf{Q})_K^{\mathrm{T}} \left\{ (\mathbf{g})_K - (\mathbf{h}_{\mathrm{av}})_K \right\}. \tag{7}$$

A line drawn as $\overline{(\mathbf{w})_K}$ shows that the vector is obtained as an estimation. By using the weight obtained from Eq. (7), estimates of $\mathbf{g}$ may be obtained as its reconstruction, denoted by $\overline{(\mathbf{g})_K}$, by using the following equation:

$$\overline{(\mathbf{g})_K} \approx (\mathbf{Q})_K \cdot \overline{(\mathbf{w})_K} + \mathbf{h}_{\mathrm{av}}. \tag{8}$$

Since vector $\overline{(\mathbf{g})_K}$ is obtained from weight vector $\overline{(\mathbf{w})_K}$, $\overline{(\mathbf{g})_K}$ is also regarded as an estimation. Using the principal components with larger corresponding eigenvalues, it is expected that $\overline{(\mathbf{g})_K}$ obtained from Eq. (8) is close to $\mathbf{g}$.

In the above statements, the following two assumptions hold:

1. The HRIRs of a certain subject from a certain sound source can be obtained by using the average vector $\mathbf{h}_{\mathrm{av}}$ and the principal component matrix $\mathbf{Q}$, both of which are constructed by using the HRIRs of subjects and sound source locations. In other words, anyone's HRIRs may be represented by using $\mathbf{h}_{\mathrm{av}}$ and $\mathbf{Q}$.
2. By using $\mathbf{h}_{\mathrm{av}}$, $\mathbf{Q}$ and a part of the impulse response of a subject's HRIR from a sound source seemingly corresponding to a direct component, $(\mathbf{g})_K$, the weight vector, $\overline{(\mathbf{w})_K}$, can be properly estimated by using Eq. (7), and $\overline{(\mathbf{g})_K}$ is estimated using $\overline{(\mathbf{w})_K}$, and is close to $\mathbf{g}$.

The first assumption may be statistically valid if $\mathbf{h}_{\mathrm{av}}$ and $\mathbf{Q}$ are constructed by using adequate amounts of data with many subjects covering all directions. The second is nominal and ambiguous. In the next section, the effectiveness and limitations of this method are examined through its application to the estimation of the HRIRs of a HATS (Head-And-Torso Simulator).

4. **Elementary investigation: Estimation of HRIRs of a HATS.** If a person's HRIRs are obtained in all directions, the average vector and the principal component matrix can be constructed for it, and the weight vector for each HRIR can be obtained by using them. Therefore, no HRIR has anything that needs to be estimated by using the proposed method. In this section, the HRIRs of a HATS are estimated by using the average vector and the principal component matrix of its own HRIRs to test the proposed method.

4.1. **Conditions of estimation.** An HRTF database downloaded from the Media Lab. at MIT was used [20]. This database consisted of a set of HRIRs of KEMAR HATS at a sampling frequency of 44.1 kHz on 512 points. The initial delay in each response was extracted, and the first 256 sample points of the response were taken as data for the analysis, windowing the response with latter half of a 512-point Blackman–Harris window function by adjusting its peak to that of the HRIR. All the 1,420 HRIRs (710 directions of the sound source, of the left and right ears) were used to compute the average vector and the principal component matrix.

The number of sample points used for the estimation of the HRIRs (value of parameter K in Eq. (7)) was set to 30. Therefore, the estimation was carried out for HRIRs after the K-th sample. The azimuths of the sound source, 0°, 90°, 180°, and 270° corresponded to the front, to the right, to the back, and to the left of the subject, respectively.

4.2. **Results and discussion.**

4.2.1. *Cumulative proportion of variance (CPV).* In PCA, the cumulative proportion of variance (CPV) is used to express the variance in data consisting of the first L principal components, defined as follows:

$$R^2(L) = \frac{\sum_{k=1}^{L} \lambda_k}{\sum_{k=1}^{N} \lambda_k}, \tag{9}$$

where N denotes the total number of principal components, equal to the length of the HRIR vector. The change of the CPV with the number of principal components is shown in Fig. 1. The figure shows that the CPV monotonically increased and converged to 1.0 with the increase in the number of components. The least number of components to cover the four values of the CPV —0.90, 0.95, 0.99, and 0.999— is indicated in Table 1. The table shows that the first 39 principal components covered 99.9% of the variance in data. This corresponded to approximately 15% (39/256) of the total number of principal components.

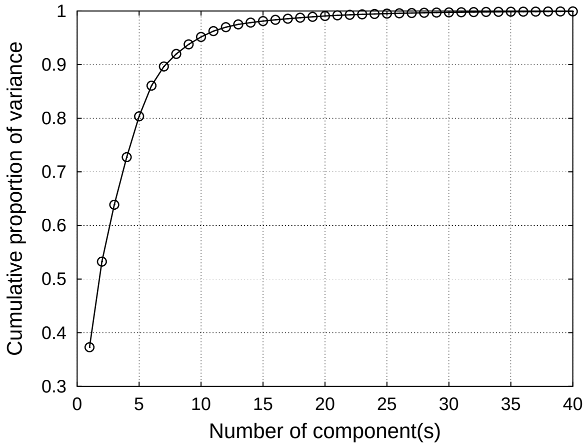


Fig. 1. Change in cumulative proportion of variance (CPV) with the number of component(s)

Table 1. Smallest numbers of principal components required to cover various CPV values

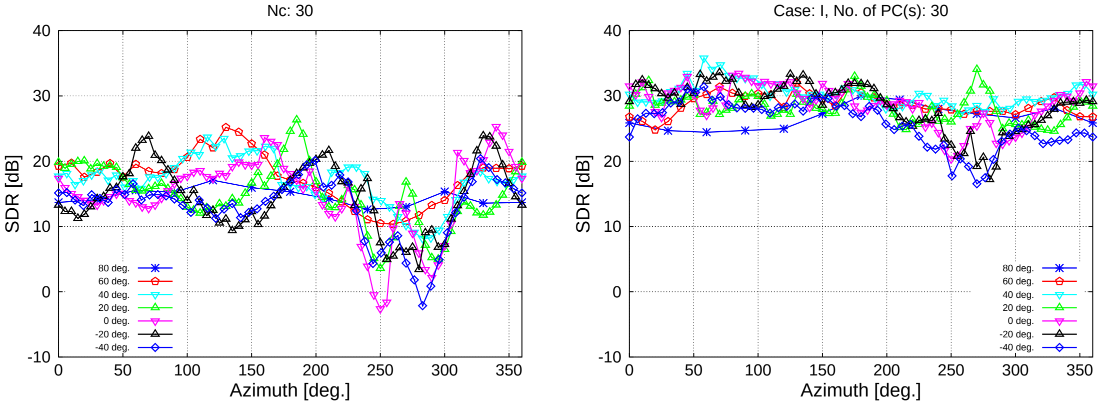
CPV	0.9	0.95	0.99	0.999
No. of components	8	10	20	39

4.2.2. *Estimation accuracy in time and frequency domains.* To evaluate estimation accuracies in time and frequency domains, SDR (Signal-to-Distortion Ratio) and SD (Spectral Distortion) were computed. Each is defined as follows:

$$\mathrm{SDR}\left[\mathbf{h}, \hat{\mathbf{h}}\right] = 10 \log_{10} \frac{||\mathbf{h}||}{\left|\left|\mathbf{h} - \hat{\mathbf{h}}\right|\right|} \quad [\mathrm{dB}], \tag{10}$$

$$\mathrm{SD}\left[\mathbf{H}, \hat{\mathbf{H}}\right] = \sqrt{\frac{1}{N_f} \sum_{k=1}^{N_f} 10 \log_{10} \left| \frac{\hat{H}(k)}{H(k)} \right|} \quad [\mathrm{dB}], \tag{11}$$

where $\mathbf{h}$ and $\hat{\mathbf{h}}$ are the original and estimated HRIRs, respectively, and $\mathbf{H}$ and $\hat{\mathbf{H}}$ are the original and estimated HRTFs, respectively. The original and estimated HRTF were obtained by using Fourier transform. $H(k)$ and $\hat{H}(k)$ in Eq. (11), respectively, denote the k-th component of $\mathbf{H}$ and $\hat{\mathbf{H}}$. High accuracy is obtained when the value of SDR is high and the value of SD is close to 0. In this section, SDR $\left[\mathbf{h}_m, \overline{(\mathbf{h}_m)_K}\right]$ and SD $\left[\mathbf{H}_m, \overline{(\mathbf{H}_m)_K}\right]$ are calculated for a certain value of K.



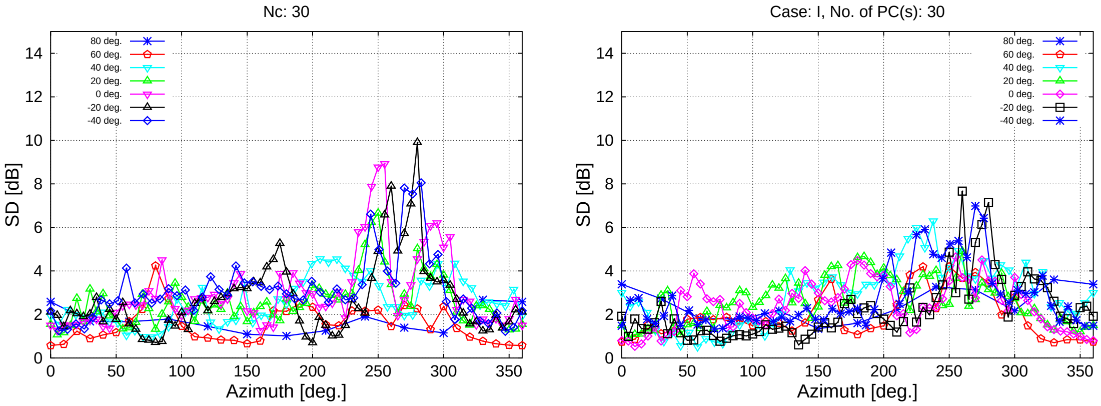
(a) Estimated results ($K = 30$)

(b) SDR of reconstructed HRIRs (No. of components: 30)

FIG. 2. Change in the SDR of the estimated right-ear HRIRs with the sound source azimuths at varying elevation angles.

For example, the SDR and SD of the HRTFs/HRIRs were computed for the case $K = 30$ and are shown in Figs. 2 and 3, respectively. In these figures, panels (a) show the results of estimation and (b) the results of reconstruction using the first K principal components. Therefore, the difference in panels (a) and (b) may be owing to the difference between the estimated and the original weight vectors. The obvious tendency in Fig. 2(b) and Fig. 3(b) is one whereby the SDR of the HRIRs and the SD of the HRTFs were both relatively worse (low SDR and high SD) in the contralateral direction than those in the ipsilateral direction. A known property caused by the SPCA [16, 18] occurs owing to relatively low energy in the HRIRs in the contralateral direction. By observing each line drawn for each elevation angle, it was found that the accuracy in the time domain was relatively low when the elevation angle was large. This might also have been the case because of the relatively low energy of the HRIRs. On the contrary, it was found that accuracy in the frequency domain was relatively low when the elevation angle was small. This might have been owing to the complexity of the frequency spectra of the HRTFs in this direction due to the influence of the shapes of the pinna and shoulder of subjects.

Comparing Fig. 2 (a) and (b), the SDR values in the former decreased by approximately 10 dB from those in the latter in most directions of the sound source. On the contrary, the difference in the SD values shown in Fig. 3 (a) and (b) was not very large. The change in estimation accuracy with elevation angle shown in Fig. 2(a) and Fig. 3(a) was roughly the same as that in Fig. 2(b) and Fig. 3(b).



(a) Estimated results ($K = 30$)

(b) SD of reconstructed HRIRs (No. of components: 30)

FIG. 3. Change in the SDs of the estimated right-ear HRIRs with source azimuths at varying elevation angles.

The estimated HRIR and HRTF are compared with their corresponding originals for many values of K in Figs. 4 and 5. Figure 4 is an example with relatively high accuracy and Fig. 5 is one with relatively low accuracy. Note that the waveforms were normalized with the absolute value of the maximum magnitude of the original, and the frequency amplitude characteristics are shown at relative levels with the maximum of the original HRTF as 0 dB. Observing Fig. 4, almost all of the difference was in the high-frequency region when the value of K was small, and almost no difference was observable when $K = 30$. On the other hand, the estimated waveform was visibly different from the original as shown in Fig. 5. From these results, the method of estimation seemed effective for the HRIRs when the sound source was in the ipsilateral direction.

5. **Estimation of HRIRs based on SPCA of dataset of HRIRs of multiple subjects.** In this section, the proposed method is examined by using the dataset of the HRIRs of multiple subjects in all directions. A dataset of HRTFs (HRIRs) from the Research Institute of Electrical Communication (RIEC), Tohoku University, was used [21]. It is called the “RIEC dataset” hereinafter.

5.1. **Conditions of estimation.** The RIEC dataset contains data for the HRIRs of 105 subjects at a sampling frequency of 48 kHz. The sound source direction varies from $-30°$ to $90°$ in $10°$ intervals for elevation, and from $0°$ to 355 ° in $5°$ intervals for the azimuth, except one azimuth when the elevation was $90°$. Totally 1,730 HRIRs (865 directions $\times$ 2 ears) were used per subject. Each HRIR was windowed by using the latter half of the Blackman–Harris window of 512 points. The windowing procedure was identical to that in **4.1**.

In the primary stage, to examine the effectiveness of the proposed method, the following procedures were carried out:

1. Subjects in the RIEC dataset were divided into two groups of almost equal number of subjects, *i.e.*, 001-053 and 054-105.

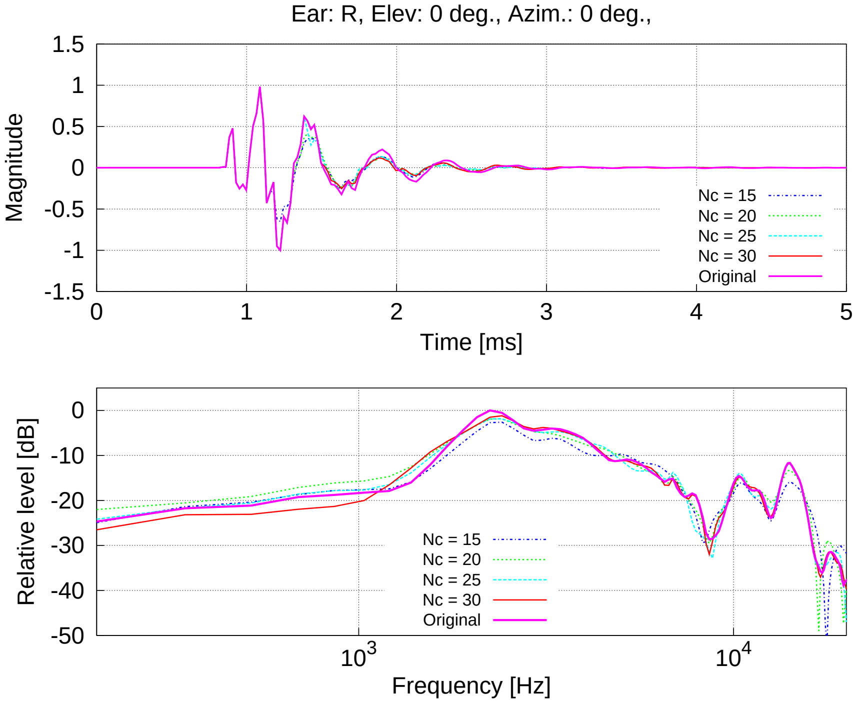


FIG. 4. Comparison of estimated HRIRs and HRTFs with the original (Elevation: 0°, Azimuth: 0°, $K = 15, 20, 25, 30$)

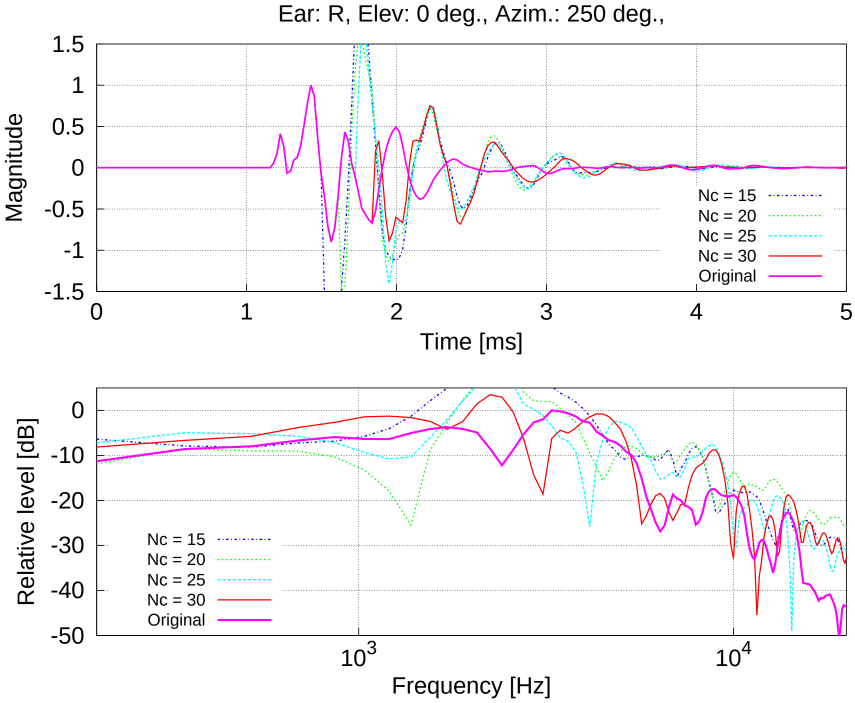


FIG. 5. Comparison of estimated HRIRs and HRTFs with the original (Elevation: 0°, Azimuth: 250°, $K = 15, 20, 25, 30$)

2. The average vector $\mathbf{h}_{\text{av}}$ and the principal component matrix $\mathbf{Q}$ were constructed by using data from the latter half of subjects (054-105).
3. Using $\mathbf{h}_{\text{av}}$ and $\mathbf{Q}$, the HRIRs of the former half of the subjects with various values of K were computed.

Definitions of azimuth and elevation were identical to those in **4.1**.

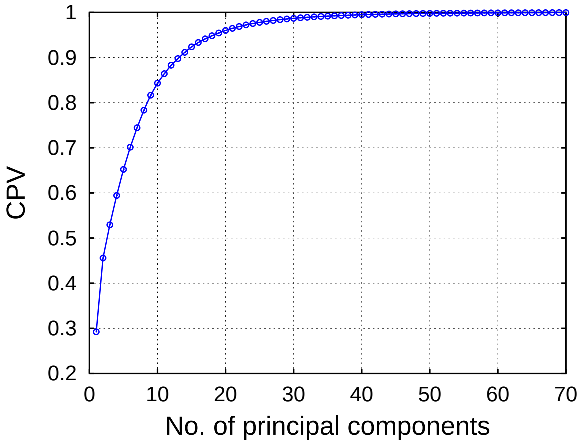


Fig. 6. Change in cumulative proportion of variance (CPV) with number of component(s)

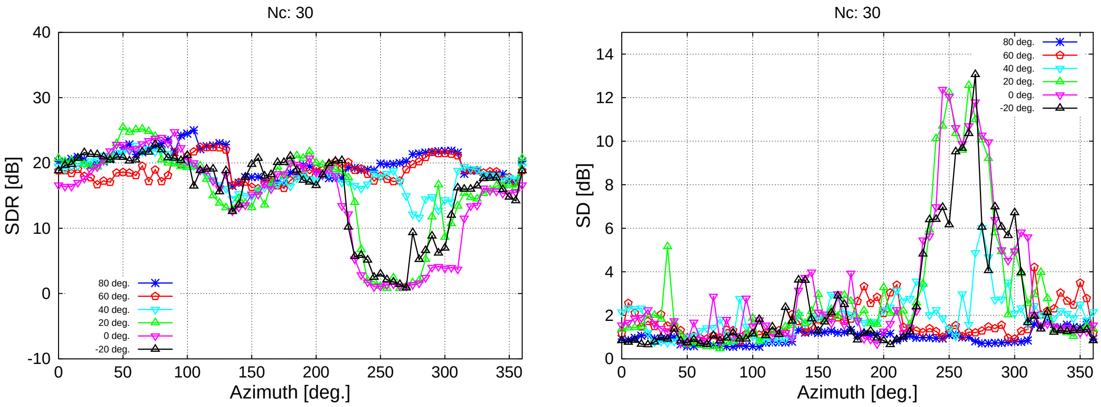
5.2. **Results and discussion.**

5.2.1. *CPV.* The change in CPV with the number of principal component(s) was plotted as shown in Fig. 6. For the same number of components, the CPV in Fig. 6 became greater than that in Fig. 1 because of the amount of data to construct the covariance matrix. The amount of data was approximately 62.5 times larger in size than that in the previous section. The smallest number of components used to cover the three values of the CPV —0.95, 0.99, and 0.999— is indicated in Table 2. From the table, it is clear that the first 61 principal components could cover 99.9% of the variance in data. This corresponded to approximately 24% (61/256) of the total number of principal components. This percentage increased from that stated in **4.2.1** owing to the increase in the amount of data for SPCA in this case.

Table 2. Smallest number of principal components required to cover various CPV values

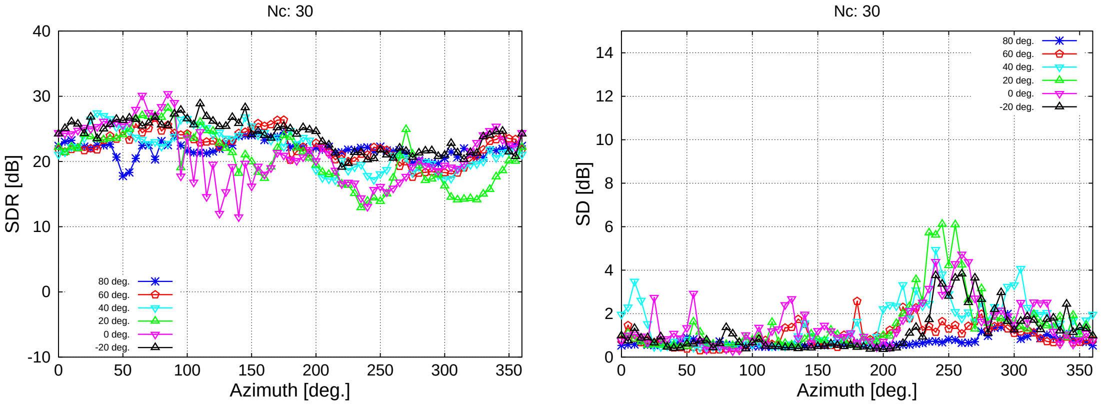
CPV	0.95	0.99	0.999
No. of components	19	33	61

5.2.2. *Estimation accuracy in terms of time and frequency domains.* The SDR and SD of the estimated HRIRs of the subjects not used to construct the average vector and the principal component matrix were computed. Examples of the results for two subjects are shown in Figs. 7 and 8 at $K = 30$. Both figures show the estimation accuracy of right-ear HRIRs. The results for subject 027 are shown in Fig. 7. Relatively high accuracy (higher SDR and lower SD) was achieved when the sound source was in the ipsilateral direction whereas accuracy was degraded when the source was in the contralateral direction. Although the same tendency was observed in Fig. 8, estimation accuracy was higher for the subject 048 than for subject 027.



(a) SDR (b) SD

FIG. 7. Change in the SDR and SD of right-ear HRIRs with source azimuths at various elevation angles ($K = 30$, subject: 027)



(a) SDR (b) SD

FIG. 8. Change in the SDR and SD of the estimated right-ear HRIRs with source azimuths at various elevation angles ($K = 30$, subject: 048)

Comparing the results in Figs. 7 and 8 with those in Figs. 2 and 3, the values of SDR and SD in the former pair are higher and lower, respectively, than those in the latter, except for the directions of the sound source at low accuracy. Since the HRTF database and the RIEC dataset stored HRIR data at different sampling frequencies, a direct comparison was not possible. However, this indicates that the proposed estimation method works better when the average vector and the principal component matrix are constructed using a large number of subjects and directions of the source of sound.

6. **Testing the proposed estimation method.** The proposed method was applied to estimate the HRIRs of a dummy head based on its impulse responses measured in a room.

6.1. **Conditions of measurement and estimation.** In Fig. 9 the setup for the HRIR measurement of the dummy head at a meeting room in the author's affiliation is shown. The SAMREC QII (manufactured by Southern Acoustics, Co. Ltd., Japan) was used as the dummy head, and its HRIRs were not included in the RIEC dataset. Moreover, the SAMREC QII did not have a torso. Electret condenser microphones were installed on both ears of the dummy head. The impulse responses from the sound source to the ears were

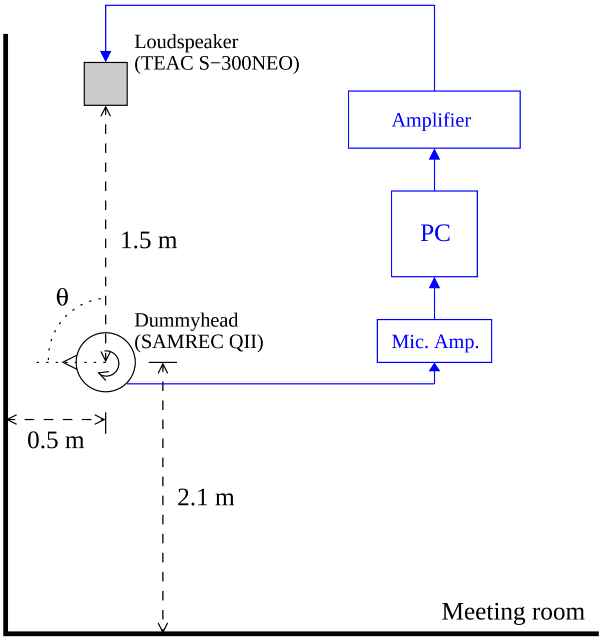


Fig. 9. Setup for HRIR measurement of dummy head in a room (from above). The blue boxes and lines represent the equipment and signal flows, respectively, of the measurement. The ear of the dummy head was 1.15 m above the floor and the center of the diaphragm of the loudspeaker was set at the same height.

measured along the horizontal plane. The azimuth was changed by rotating the dummy head. The measured azimuths were 0∼180 degrees at an interval of 30 degrees. The reverberation time and the A-weighted sound pressure of background noise in the room were approximately 0.8 s and 32.3 dB, respectively. For comparison, impulse responses of the dummy head were also measured at anechoic room in the author's affiliation. The distance between the loudspeaker and the dummy head, and the height of the ear relative to the floor, were kept identical to those in Fig. 9. An impulse response from the sound source to the center of the dummy head with its absence was measured, and its inverse filter was convoluted with all responses to eliminate the characteristics of the frequency of the loudspeaker and the microphones. Since the reproducible frequency range of the loudspeaker was limited, the inverse filter was computed with the target response set to a band-pass filter with lower and higher cutoff frequencies, respectively, set to 100 Hz and 18000 Hz. The sampling frequency of the measurement was 48 kHz.

As shown in Fig. 9, the dummy head was located near the side wall of the room to examine the effectiveness of the proposed method. An example of impulse response with the the azimuth source of the sound set to 150 degrees is shown in Fig. 10. In the upper panel of Fig. 10, the impulse response acquired in the meeting room (blue line) was compared with that in the anechoic room (green line). Their corresponding frequency spectra are plotted in the lower panel of Fig. 10. In this panel, the blue and green lines denote the BRTF (Binaural Room Transfer Function) and the HRTF, respectively. Note that these lines were computed according to the corresponding lines in the upper panel, where the frequency spectrum depicted by the blue line did not indicate the characteristics of the meeting room. At the upper panel, the green line is almost identical to the direct component of the blue line. The successive waveform represented by the blue line indicates the varying reflections, and the earliest one was from the side wall of the room. The time difference between the first and the second peaks in the waveform was approximately 0.8 ms, corresponding to the difference between the direct and the reflected paths. Since the number of sample points in 1 ms was 48 at a 48-kHz sampling frequency, this difference was equal to approximately 30 points. According to the results in **5**, the estimation yielded acceptable accuracy even in such cases. By using the impulse responses measured at the room with varying azimuths, the HRIRs were estimated at $K = 30$. As the principal component matrix $\mathbf{Q}$ and the average vector $\mathbf{h}_{\mathrm{av}}$, these obtained in **5** were used.

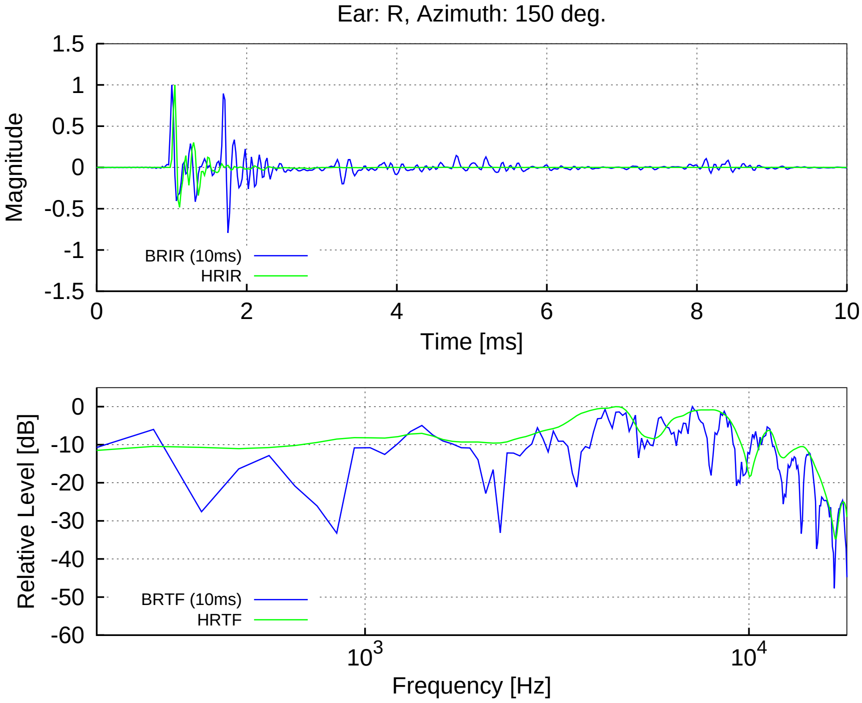
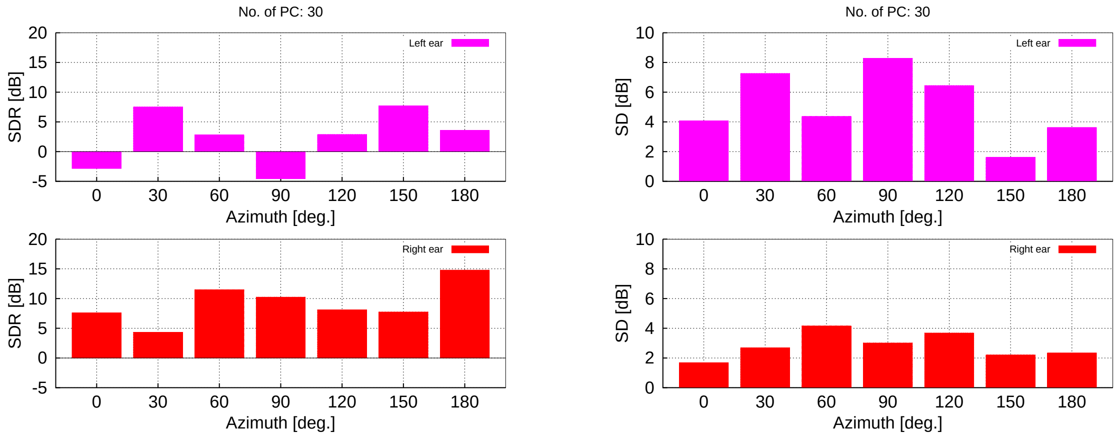


Fig. 10. Upper panel: A comparison of the impulse response of the right ear when the azimuth was 150 degrees with the corresponding HRIR in 10 ms. Lower panel: the corresponding frequency spectra. Impulse responses were aligned so that both peaks occurred at 1 ms. The frequency spectra were computed by using the time domain waveforms in 10 ms and their largest levels were set to 0 dB.

6.2. **Results and discussion.** The SDR and the SD were calculated from the original and estimated HRIRs, respectively, and are shown in Fig. 11. In Fig. 11(a) and (b), characteristics are plotted by using bars for both the left-ear and the right-ear HRIRs, respectively, in magenta and red. The left and right ears correspond to the contralateral and the ipsilateral sides, respectively. With these panels, the SDR values were higher and the SD values lower for the right-ear HRIRs, and the opposite characteristics were observed for the left-ear HRIRs. This means that estimation accuracy for the ipsilateral-side HRIRs was relatively better than that for contralateral-side HRIRs. This was common in cases reported in the previous sections. The SDR and the SD values are lower and higher, respectively, than those for the cases in **5**, meaning that estimation accuracy was relatively worse than that in the previous section. Two examples of the original and estimated HRIRs/HRTFs are shown in Fig. 12 and Fig. 13, respectively. Relatively better accuracy was observed in Fig. 12. The original and the estimated results agreed well in Fig. 12. In Fig. 13, the frequency spectra of the original and the estimated HRTFs roughly agreed,



(a) SDR (b) SD

FIG. 11. Change in SDR and SD of the estimated HRIRs/HRTFs with source azimuths ($K = 30$).

but the details, especially in high-frequency regions, was somewhat different. Almost all data in the RIEC dataset consisted of HRIRs of human subjects, whereas the HRIRs used in this section were those of the dummy head without a torso. Subjective evaluations of the proposed method are essential to determine the required accuracy. This will form a subjects for future work.

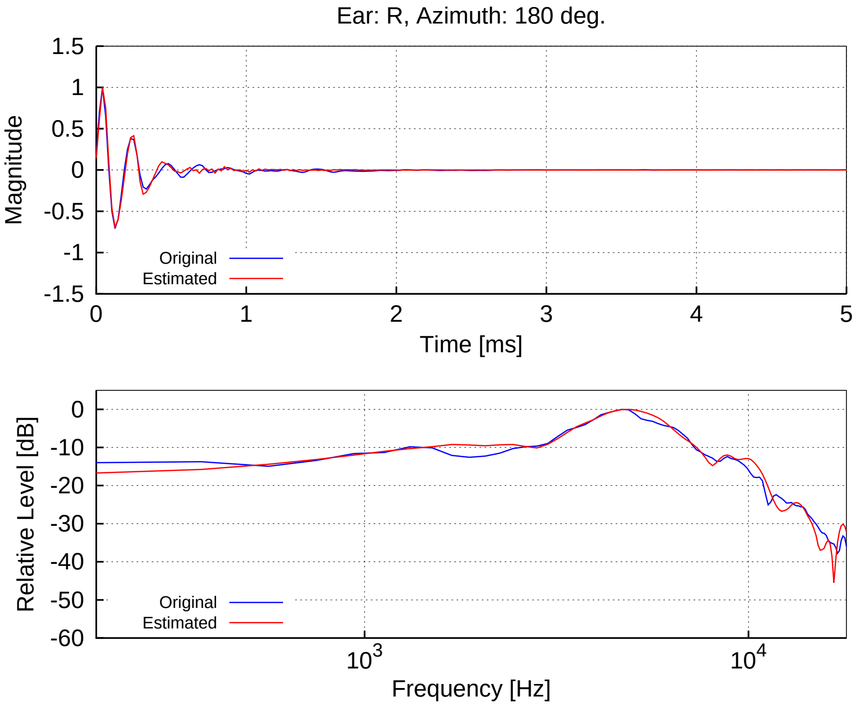


FIG. 12. Comparison of original and estimated HRIRs (upper panel), and the original and estimated HRTFs (lower panel) (azimuth: 180 degrees, right ear, $K = 30$). Note that the HRIRs are normalized, their initial delays are extracted, and the HRTFs are plotted at their relative levels.

The proposed method did not work well when the source of sound was on the contralateral side. Although details of the frequency spectrum of the contralateral-side HRTF

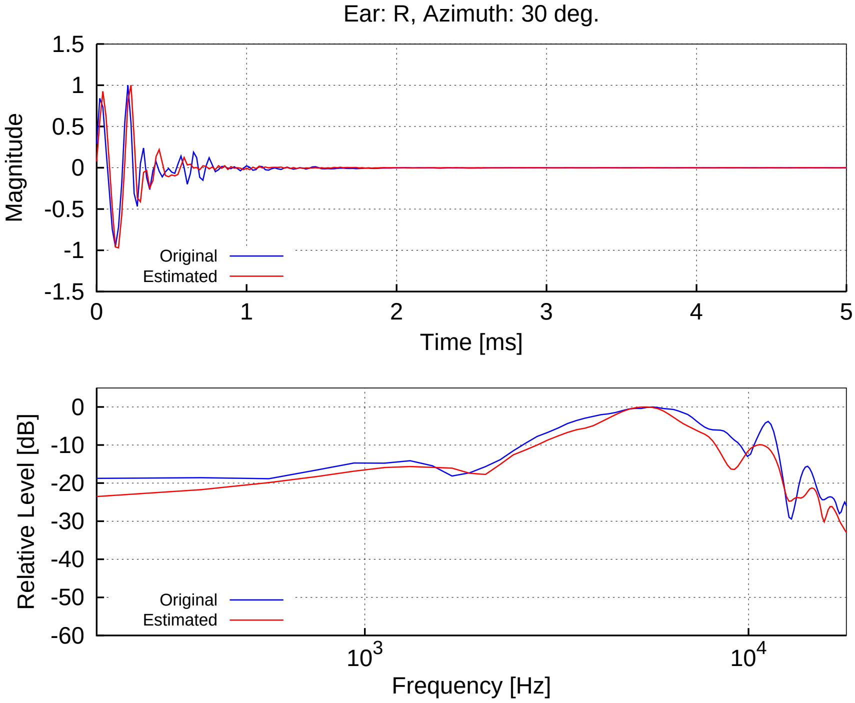


FIG. 13. Comparison of original and estimated HRIRs (upper panel), and the original and estimated HRTFs (lower panel) (azimuth: 30 degrees, right ear, $K = 30$). Note that the HRIRs are normalized, their initial delays are extracted, and the HRTFs are plotted at their relative levels.

might not be important for sound localization [22], the proposed method needs to be improved or other methods should be investigated. In this sense, the proposed method is not universally applicable to the estimation of HRIRs. The most practical way to acquire HRIRs is to set the subject apart from anything that might reflect sound. This is possible in a meeting room but not in a living room with many objects. The experiment dealt with the case where the walls of the room were close to the dummy head. The results showed that part of the direct component of the impulse response measured in the room might be sufficient for the acquisition of HRIR when the sound source was on the ipsilateral side. This might partially ease the limitation on the acquisition of HRIRs, especially with regard to the need for an anechoic environment.

7. **Concluding remarks and future work.** This paper proposed a method to estimate HRIRs from impulse responses acquired in a room with reflection and noise based on the SPCA of the HRIRs. It was shown that the proposed method works well when the estimated HRIR is used to construct the average vector and the principal component matrix. Moreover, estimation using the RIEC dataset varied with case, but an acceptable accuracy was achieved in terms of both time and frequency when the sound source was in the ipsilateral direction. Tests of the proposed method yielded roughly the same results.

A more detailed investigation into the differences in estimation accuracies, the amount of data required for adequate accuracy, a subjective evaluation of the estimated results, and an assessment of the estimation when the sound source is in contralateral side, will form part of future work.

Acknowledgment. Part of this study was supported by JSPS KAKENHI (JP25330202, JP17K00242). The author thanks Mr. Keisuke Sakamoto and Mr. Takeki Miura for cooperation on the measurement of impulse responses.

REFERENCES

[1] T. D. Rossing, E. R. Moore, P. A. Wheeler, *The Science of Sound, third edition*, Addison Wesley, 2002.

[2] J. Blauert, *Spatial hearing, revised edition*, MIT Press, 1999.

[3] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, A new theory for high definition virtual acoustic display named ADVISE, *Acoust. Sci. & Tech.*, vol. 24, no. 5, pp. 276-283, 2003.

[4] S. Takane, S. Takahashi, Y. Suzuki, T. Miyajima, Elementary real-time implementation of a virtual acoustic display based on ADVISE, *Acoust. Sci. & Tech.*, vol. 24, no. 5, pp. 304-310, 2003.

[5] M. Otani and T. Hirahara, A dynamic auditory display: Its design, performance, and problems in HRTF switching, *Proc. Japan-China Joint Conf. Acoust.*, SS-1-3, 2007.

[6] S. Yairi, Y. Iwaya, Y. Suzuki, Estimation of detection threshold of system latency of virtual auditory display, *Applied Acoustics*, vol. 68, no. 8, pp. 851-863, 2007.

[7] J. D. Miller, Slab: A software-based real-time virtual acoustic environment rendering system, *Proc. 2001 ICAD*, pp. 279-280, 2001.

[8] D. R. Begault, *3D sound for Virtual Reality and multimedia*, AP Professional, 1994.

[9] K. Watanabe, Y. Oikawa, S. Sato, S. Takane, K. Abe, Development and performance evaluation of virtual auditory display system to synthesize sound from multiple sound sources using graphics processing unit, *Proc. 21th International Congress on Acoustics*, 2pEAba12 (7 pages in CD-ROM), 2013.

[10] H. Kuttruff, *Room Acoustics, third edition*, Elsevier Applied Science, 1991.

[11] W. Zhang, T. D. Abhayapala, R. A. Kennedy and R. Duraiswami, Insights into head-related transfer function: Spatial dimensionality and continuous representation, *J. Acoust. Soc. Am.*, vol. 127, no. 4, pp. 2347-2357, 2010.

[12] S. Takane, Estimation of whole waveform of head-related impulse responses based on auto regressive model for their acquisition without anechoic environment, *Principles and Applications of Spatial Hearing, edited by Y. Suzuki et al.*, pp. 216-225, World Scientific, 2011.

[13] W. L. Martens, Principal component analysis and resynthesis of spectral cues to perceived direction, *Proc. ICMC*, pp. 274-281, 1987.

[14] D. J. Kistler and F. L. Wightmann, A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction, *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1637-1647, 1992.

[15] J. Chen, B. D. Van Veen, K. E. Hecox, A spatial feature extraction and regularization model for the head-related transfer functions, *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 439-452, 1995.

[16] B.-S. Xie, Recovery of individual head-related transfer functions from a small set of measurements, *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 282-294, 2012.

[17] Z. Wu, F. H. Y. Chan, F. K. Lam, J. C. K. Chan, A time domain binaural model based on spatial feature extraction for head-related transfer functions, *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2211-2218, 1995.

[18] S. Takane, Effect of domain selection for compact representation of spatial variation of head-related transfer function in all directions based on spatial principal components analysis, *Applied Acoustics*, vol. 101, pp. 64-77, 2016.

[19] B.-S. Xie, *Head-related transfer function and virtual auditory display, second edition*, J. Ross Pub., 2013.

[20] W. G. Gardner and K. D. Martin, HRTF measurements of a KEMAR, *J. Acoust. Soc. Am.*, vol. 97, pp. 3907-3908, 1995.

[21] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane and S. Sato, Dataset of head-related transfer functions measured with a circular loudspeaker array, *Acoust. Sci. & Tech.*, vol. 35, no. 3, pp. 159-165, 2014.

[22] K. Watanabe, R. Kodama, S. Sato, S. Takane and K. Abe, Influence of flattening contralateral head-related transfer functions upon sound localization performance, *Acoust. Sci.& Tech.*, vol. 32, no. 3, pp. 121-124, 2011.