Acoust. Sci. & Tech. **28**, 4 (2007)

PAPER

Automatic filter design for 3-D sound movement in embedded applications

Kosuke Tsujino[1,*], Wataru Kobayashi[2], Takao Onoye[3] and Yukihiro Nakamura[1]

[1]*Department of Communications and Computers Engineering, Kyoto University, Japan*
[2]*Arnis Sound Technologies, Co., Ltd., Japan*
[3]*Department of Information Systems Engineering, Osaka University, Japan*

(*Received 5 April 2006, Accepted for publication 16 October 2006*)

Abstract: In this paper, we propose an automatic method of designing digital filters for three-dimensional (3-D) sound movement that is dedicated to embedded applications. By this method, read-only memory (ROM) capacity and computational load are reduced with only slight degradation of the 3-D sound effect. For practical applications of 3-D sound, the continuous movement of a virtual sound image is indispensable. To achieve 3-D sound movement, the interpolation of filter coefficients is often required; however, the frequency response of an intermediate filter obtained by interpolation is severely distorted using conventional automatic design methods. The proposed method reduces this distortion by evaluating the response of the intermediate filters in the optimization process. The results of objective evaluation and subjective listening tests show that the proposed method improves the perceptual quality of 3-D sound movement, even with runtime interpolation of filter coefficients.

Keywords: 3-D sound, Head-related transfer function, Interpolation, Gradient search algorithm

PACS number: 43.66.Pn, 43.66.Qp [doi:10.1250/ast.28.219]

1. INTRODUCTION

Three-dimensional (3-D) sound localization aims to give a virtual impression of sound source elevation, azimuth, and distance to listeners through 2-channel stereo sound. To achieve this, 3-D sound localization systems must simulate auditory cues utilized for the human judgment of sound source positions. Most of these cues are contained in a pair of acoustical transfer functions from the sound source position to the left/right ears of a listener [1–3]. These transfer functions are called head-related transfer functions (HRTFs). A pair of digital filters that approximates left/right HRTFs are usually utilized in 3-D sound localization.

Tradeoffs exist between the approximation accuracy of an HRTF and the computational resources required in 3-D sound localization. The most precise HRTF approximation is obtained by reproducing a head-related impulse response (HRIR) [4], a time-domain representation of an HRTF, using a high-order FIR filter. However, this method requires a high computational load and memory capacity in processing systems. The reduction of computational load and memory usage with minimum degradation of the approximated HRTF is required in such embedded applications as mobile phones and portable audio players. The utilization of IIR filters is effective for reducing the computational cost of sound localization [4,5]. From this point of view, novel digital filter architecture for 3-D sound localization has been proposed by Kobayashi *et al.* [6] that utilizes subband decomposition and IIR filters. Since this filter architecture successfully reduces memory usage and computational load with only a slight degradation of the sound effect, we adopt it in the present paper.

3-D sound localization is valuable for various application fields including entertainment, music, telecommunications, and information displays. Since 3-D sound is often interactively presented according to user manipulation in these fields, the movement of a virtual sound source created by 3-D sound localization is indispensable. Virtual sound movement can be realized by gradually changing the frequency response of the digital filter on the basis of the desired sound movement. To realize arbitrary sound movement in virtual space, filter coefficients for sound localization must be prepared for all directions around the listener. However, enormous ROM space is required to hold all of these filter coefficients. Therefore, in practical processing systems, filter coefficients for representative positions on a coarse grid are designed in advance and held in the ROM. To ensure gradual changes in frequency response, interpolation of filter coefficients among adjacent representative positions is usually performed at runtime [7–10].

*e-mail: tsujino@easter.kuee.kyoto-u.ac.jp

When FIR filters are utilized for sound localization, interpolation of the FIR coefficients and the delay factor result in smooth movement of the sound image if position-dependent delay is removed from the FIR coefficients [11–14]. For the IIR filters, the interpolation of MA coefficients yields as good interpolation results as the interpolation of FIR coefficients if the AR coefficients of the IIR filter are common for all sound source positions [15]. Although filter order can be further reduced by making AR coefficients position-dependent, interpolation results are severely distorted in this case when conventional automatic design methods [4,5,16,17] are used. To cope with this difficulty, in this paper a novel digital filter design method for 3-D sound movement in embedded applications is proposed that minimizes the distortion resulting from filter interpolation.

The rest of this paper is organized as follows. In Sect. 2, we explain several signal processing techniques that are used as a basis of the proposed method. The proposed method is described in Sect. 3 and evaluated in Sect. 4. Finally, the paper is concluded in Sect. 5.

2. 3-D SOUND FOR EMBEDDED APPLICATIONS

In this section, we explain several signal processing techniques that are used as the basis of the proposed filter design method. First, the filter architecture for 3-D sound localization in embedded applications is explained. An automatic filter design method for this filter architecture previously published by the authors is also shown. A sound movement method using this filter architecture is then described, followed by a discussion of the problems caused by the interpolation of the filter coefficients.

2.1. 3-D Sound Localization for Embedded Applications

The digital filter architecture for 3-D sound localization in embedded applications [6] is briefly introduced in this subsection. As shown in Fig. 1, this filter architecture consists of three stages: subband decomposition, filtering, and mixing. In the subband decomposition stage, the input

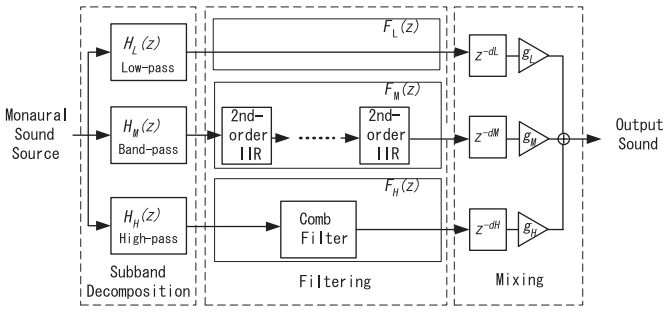

Fig. 1 Digital filter architecture for 3-D sound localization.

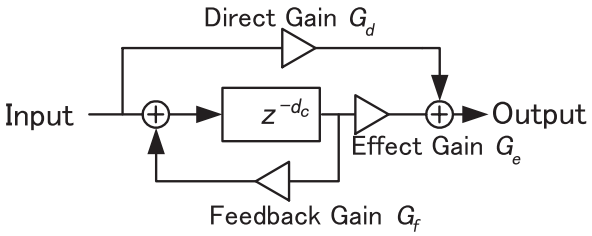

Fig. 2 Comb filter.

sound is decomposed into three subbands using bandpass filters. Next, in the filtering stage, each band-limited signal is processed using different digital filters based on the physical characteristics and psychological importance of the HRTF in each subband. Finally, the three subbands are mixed after adjustments for gain and delay, which are also important psychological cues for human spatial hearing.

All bandpass filters, $H_{\mathrm{L}}(z)$, $H_{\mathrm{M}}(z)$, and $H_{\mathrm{H}}(z)$, are implemented using an FIR filter, whose cutoff frequency is derived from HRTF analysis [18]. The boundary frequency between the low and middle subbands is about 1 kHz and is about 6–8 kHz between the middle and high subbands. For digital filters in the filtering process, since the HRTF is almost flat in the low subband, $F_{\mathrm{L}}(z)$ is always set to 1. $F_{\mathrm{M}}(z)$, the digital filter for the processing of the middle subband, comprises a cascade connection of N second-order IIR filters, and $F_{\mathrm{H}}(z)$, the filter in the high subband, is a comb filter, as shown in Fig. 2. The transfer functions of $F_{\mathrm{M}}(z)$ and $F_{\mathrm{H}}(z)$ are as follows:

$$F_{\mathrm{M}}(z) = \prod_{n=0}^{N-1} \frac{b_{0n} + b_{1n}z^{-1} + b_{2n}z^{-2}}{1 + a_{1n}z^{-1} + a_{2n}z^{-2}} \tag{1}$$

$$F_{\mathrm{H}}(z) = \frac{G_{\mathrm{d}} + (G_{\mathrm{e}} - G_{\mathrm{f}})z^{-d_{\mathrm{c}}}}{1 - G_{\mathrm{f}}z^{-d_{\mathrm{c}}}}. \tag{2}$$

Since the outputs of $F_{\mathrm{M}}(z)$ and $F_{\mathrm{H}}(z)$ are multiplied by gain factors g_{M} and g_{H}, respectively, assuming that $b_{0n} = 1$ and $G_{\mathrm{d}} = 1$ does not lead to loss of the generality of the model. In the rest of this paper, we set $b_{0n} = 1$ and $G_{\mathrm{d}} = 1$.

In a typical case, stereo output using this filter architecture requires about 7.7×10^6 multiply-and-accumulate (MAC) operations per second with a 44.1 kHz sampling rate, which can easily be processed on embedded DSPs. The read-only memory (ROM) capacity for the filter coefficient database for all directions is about 5.4 k words. The realtime operation of sound localization using this filter architecture with commercially available DSPs is also confirmed [6,7].

2.2. Digital Filter Design Method for Sound Localization

Digital filters with such complicated frequency responses as HRTFs are generally designed by minimizing

errors between desired and realized frequency responses. In this paper, an HRTF for the left ear with a sound source at elevation ϕ and azimuth θ is denoted by $\hat{\boldsymbol{H}}(e^{j\omega}, \phi, \theta)$. The logarithmic error is broadly accepted as a good error measure in acoustical applications [17]. The logarithmic error $\epsilon^2(\boldsymbol{x}, \phi, \theta)$ is defined by the following equation, where $\boldsymbol{x}$ is the parameter vector, $H(e^{j\omega}, \boldsymbol{x})$ is the realized response, and $W(e^{j\omega})$ is a weighting function.

$$\epsilon^2(\boldsymbol{x}, \phi, \theta) = \frac{\displaystyle\int_{-\pi}^{\pi} \left|W(e^{j\omega})\right|^2 \left|\log|\hat{\boldsymbol{H}}(e^{j\omega}, \phi, \theta)| - \log|H(e^{j\omega}, \boldsymbol{x})|\right|^2 d\omega}{\displaystyle\int_{-\pi}^{\pi} \left|W(e^{j\omega})\right|^2 d\omega} \quad (3)$$

A gradient search algorithm, as used by Blommer and Wakefield [17], is used to solve this nonlinear problem, which can solve it within a practical computation time.

2.2.1. Derivation of initial parameters

In gradient search algorithms, the optimization result depends on the initial value of the parameter vector. To consistently obtain good design results, the initial parameters for the gradient search are designed in advance using the following method.

Let us introduce a weighting function $W(e^{j\omega})$ defined by the following equation to approximate the human audible range, where f_s is sampling frequency.

$$W(e^{j\omega}) = \begin{cases} 1 & \left(100 \le \dfrac{f_s \omega}{2\pi} \le 16000\right) \\ 0.05 & (\text{otherwise}). \end{cases} \quad (4)$$

Initial values of gain factors g_L, g_M, and g_H, and delay factors d_L, d_M, and d_H are set to the weighted average of gain and phase delay of the desired HRTF in the passband of corresponding bandpass filters, as given by Eqs. (5) and (6). In Eqs. (5) and (6), $\mathrm{pd}(\hat{\boldsymbol{H}}(e^{j\omega}))$ denotes the phase delay of the desired response $\hat{\boldsymbol{H}}(e^{j\omega})$, and suffix $\mathcal{X}$ is one of L, M, or H:

$$g_{\mathcal{X}} = \exp\left(\frac{\displaystyle\int_0^{\pi} \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| \log|\hat{\boldsymbol{H}}(e^{j\omega})| d\omega}{\displaystyle\int_0^{\pi} \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| d\omega}\right) \quad (5)$$

$$d_{\mathcal{X}} = \frac{\displaystyle\int_0^{\pi} \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| \mathrm{pd}(\hat{\boldsymbol{H}}(e^{j\omega})) d\omega}{\displaystyle\int_0^{\pi} \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| d\omega}. \quad (6)$$

In the design of $F_M(z)$ and $F_H(z)$, a simplified error measure defined by the following equation is utilized, where $A_{\mathcal{X}}(z)$ and $B_{\mathcal{X}}(z)$ are the denominator and the numerator of $F_{\mathcal{X}}(z)$, respectively (i.e., $F_{\mathcal{X}}(z) = \frac{B_{\mathcal{X}}(z)}{A_{\mathcal{X}}(z)}$).

$$\sigma_{\mathcal{X}}{}^2 = \int_{-\pi}^{\pi} \left|W(e^{j\omega})\right|^2 \left|\hat{\boldsymbol{H}}_{\mathcal{X}}(e^{j\omega})A_{\mathcal{X}}(e^{j\omega}) - B_{\mathcal{X}}(e^{j\omega})\right|^2 d\omega$$

$$\text{where} \quad \hat{\boldsymbol{H}}_{\mathcal{X}}(e^{j\omega}) = \frac{\hat{\boldsymbol{H}}(e^{j\omega})}{g_{\mathcal{X}}} \quad (7)$$

Since the input of $F_{\mathcal{X}}(z)$ is band-limited by $H_{\mathcal{X}}(z)$, error outside the passband of $H_{\mathcal{X}}(z)$ is ignored by multiplying $H_{\mathcal{X}}(e^{j\omega})$ to the weighting function $W(e^{j\omega})$. As already shown in Eqs. (1) and (2), $A_{\mathcal{X}}(z)$ and $B_{\mathcal{X}}(z)$ are linear expressions of unknown variables. Therefore, the least-squares method can be utilized to obtain a_{1n}, a_{2n}, b_{1n}, b_{2n}, G_e, and G_f that minimize $\sigma_{\mathcal{X}}{}^2$. d_c is an integer between 2 and 25 so that $\sigma_{\mathcal{X}}{}^2$ is its minimum.

2.2.2. Optimization by gradient search algorithm

First, the parameter vector $\boldsymbol{x}$ used in the gradient search is defined. To allow a simple expression for the stability condition, Eqs. (1) and (2) are converted into Eqs. (8) and (9), where $r_{\alpha n}$ and $\theta_{\alpha n}$ are the absolute value and the argument of the poles and $r_{\beta n}$ and $\theta_{\beta n}$ are those of the zeros, respectively. In Eq. (9), variables a_c and b_c also control the absolute value of the poles and zeros.

$$F_M(z) = \prod_{n=0}^{N-1} \frac{(1 - r_{\beta n} e^{j\omega\theta_{\beta n}} z^{-1})(1 - r_{\beta n} e^{-j\omega\theta_{\beta n}} z^{-1})}{(1 - r_{\alpha n} e^{j\omega\theta_{\alpha n}} z^{-1})(1 - r_{\alpha n} e^{-j\omega\theta_{\alpha n}} z^{-1})} \quad (8)$$

$$F_H(z) = \frac{1 - b_c z^{-d_c}}{1 - a_c z^{-d_c}} \quad (9)$$

The parameter vector $\boldsymbol{x}$ is defined as follows:

$$\boldsymbol{x} = (g_L, g_M, g_H, d_L, d_M, d_H, \{r_{\alpha n}, \theta_{\alpha n}\}, \{r_{\beta n}, \theta_{\beta n}\}, a_c, b_c). \quad (10)$$

The logarithmic error ϵ^2 defined in Eq. (3) is minimized with respect to $\boldsymbol{x}$ using a gradient search algorithm. As already mentioned, the stability condition of IIR filters must be fulfilled in the design. Assuming that IIR filters are of minimum-phase, stability constraints can be expressed by the following equations, where $r_{\max}$ ($r_{\max} < 1$) is the maximum of the absolute value of the poles and zeros.

$$0 \le d_{\mathcal{X}} \quad (11)$$

$$0 \le r_{\alpha n}, r_{\beta n} \le r_{\max} \quad (12)$$

$$0 \le a_c, b_c \le r_{\max} \quad (13)$$

The sequential quadratic programming (SQP) method, a standard gradient search algorithm for constrained optimization, is used in the gradient search.

2.3. Sound Movement in Embedded Applications

IIR filters are utilized in the middle subband of the filter architecture adopted in this paper. Therefore, interpolation of IIR filters is required for sound movement based on this filter architecture. Unlike for FIR filters, the stability of interpolation results must be considered when the coefficients of IIR filters are interpolated. With second-order IIR filters, intermediate filters obtained by linear interpolation from two stable filter coefficient sets are guaranteed to be stable [19]. To guarantee the stability of the

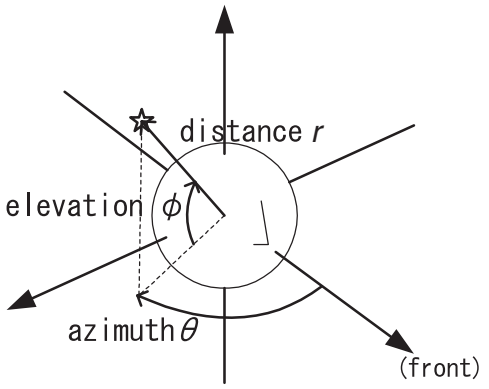

Fig. 3 Coordinate system.

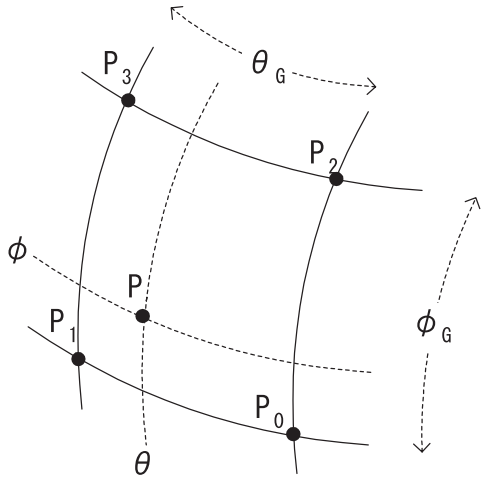

Fig. 4 Linear interpolation of parameters.

intermediate filters, $F_M(z)$, the filters for the middle subband, comprise a cascade connection of second-order IIR filters. Linear interpolation is performed for each second-order section in $F_M(z)$.

Other parameters in the filter architecture of Fig. 1 can be handled simply. Since the frequency selectivity of human hearing is relatively poor at high frequencies [20], output sound is not noticeably degraded when the comb filter in the high subband is not interpolated. Therefore, the coefficient set for the nearest representative position is used without interpolation for the comb filter. For the gain and delay factors for each subband, linear interpolation among adjacent representative positions gives good interpolation results.

As a result of the above discussion, sound movement is assumed to be realized by the following method. A spherical coordinate system, shown in Fig. 3, is used in the explanation below. Parameters g_L, g_M, g_H, d_L, d_M, and d_H and coefficients a_{in} and b_{in} for each second-order IIR filter in the middle subband are linearly interpolated. Interpolation is performed on an equidistant sphere from the listener. The representative positions are placed on a grid with fixed angle steps, θ_G and ϕ_G, relative to azimuth and elevation, respectively. Figure 4 shows position P with azimuth θ and elevation ϕ surrounded by four representative positions P_0, P_1, P_2, and P_3. Parameter p at P can be calculated by linear interpolation among the four representative positions, as shown in the following equation, where p_0, p_1, p_2, and p_3, respectively, are parameters at P_0, P_1, P_2, and P_3,

$$c_\theta = (\theta \bmod \theta_G)/\theta_G, \text{ and } c_\phi = (\phi \bmod \phi_G)/\phi_G.$$
$$p = (1-c_\theta)(1-c_\phi)p_0 + c_\theta(1-c_\phi)p_1 + (1-c_\theta)c_\phi p_2 + c_\theta c_\phi p_3 \quad (14)$$

2.4. Problems with Interpolation of Filter Coefficients

Although the intermediate filter obtained by the above interpolation method is guaranteed to be stable, its frequency response is greatly affected by the design method of the IIR filters. The numerator and denominator of each second-order section in the IIR filter correspond to a dip and a peak in the HRTF, respectively. Even if the total transfer functions of the IIR filters are identical, the responses of the intermediate filters may be affected by the assignment of the second-order sections to the peaks/dips on the adjacent representative positions.

All peaks and dips in HRTFs are caused by different physical phenomena such as reflection at a pinna or a shoulder [21]. If a second-order section is assigned to physically different peaks or dips between adjacent representative positions, the frequency response of the intermediate filter changes abruptly with the sound movement, resulting in serious errors in interpolation results. One of the worst cases of the interpolation of filter coefficients is shown in Fig. 5(a), while the linear interpolation of target HRTFs in the time domain is shown in Fig. 5(b). Solid lines indicate the responses of the target HRTFs and the designed filters at the representative positions, while the dotted lines indicate the responses of the intermediate HRTFs and the filters between these representative positions. Although Fig. 5(b) shows gradually changing responses between target HRTFs, Fig. 5(a) shows abruptly changing responses with jumps of peak/dip positions. This is due to the inconsistent assignment of second-order sections to peaks/dips in HRTFs, shown by circles, triangles, and stars in Fig. 5(a).

If such abrupt changes in frequency response occur with sound movement, the 3-D sound effect may be collapsed since such changes are inconsistent with the physical characteristics of HRTFs. In addition, abrupt changes in frequency response degrade the timbre of the output sound. To overcome these problems, a second-order section must be consistently assigned to a peak or a dip with the same physical meaning among sound source positions.

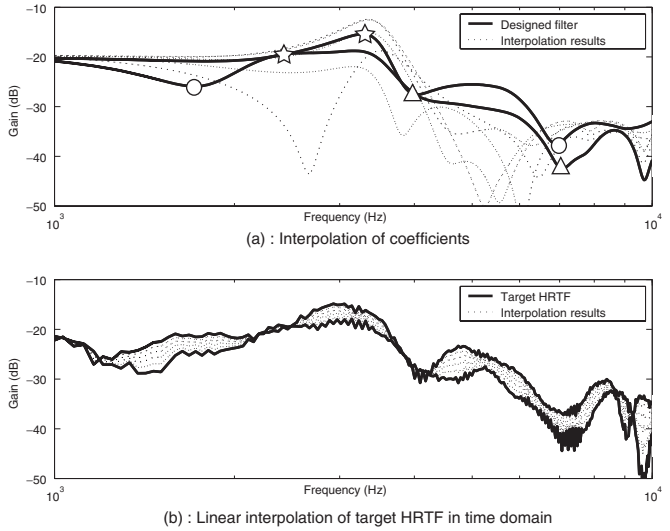

Fig. 5 Worst case of coefficient interpolation (between $\theta = 30^\circ$ and 45° at $\phi = 0^\circ$).

3. DIGITAL FILTER DESIGN METHOD FOR SOUND MOVEMENT

Although the perceptual effectiveness of the design method described in the previous section has been confirmed using the localization of stationary sound images [22], degradation in localization and timbre occurs with sound movement by this method. As already mentioned, this degradation is mainly caused by inconsistent correspondence between second-order sections in $F_M(z)$ and peaks/dips in HRTFs. This inconsistency cannot be eliminated by simply reordering the second-order sections, since the selection of peaks/dips from an HRTF is not unique. In low-order modeling of HRTFs, we have to select some peaks/dips and discard others, considering the consistent correspondence of the selected peaks/dips among the adjacent representative positions.

In this section, we describe our proposed design method, which causes the intermediate filters to linearly interpolate the HRTFs while eliminating the above inconsistency. This inconsistency cannot be eliminated by independently designing a filter for each representative position. However, it is difficult to simultaneously design a filter for all the representative positions, since that results in nonlinear optimization of thousands of variables. Instead, the filter is designed for each representative position, considering consistency between filter coefficients for adjacent representative positions. By the proposed method, a set of consistent design results for all representative positions can be obtained within a reasonable computation time.

The proposed method introduces the following two features to the basic design method described in the previous section:

(1) Design results for the adjacent representative positions are reused to obtain the initial parameters for optimization.
(2) Responses of intermediate filters are evaluated in the optimization process.

The initial parameters obtained using Eq. (7) are often inconsistent with the design results for the adjacent representative positions. This causes inconsistency in the optimization results, which yields abrupt changes in the response of the intermediate filters, as shown in Fig. 5(a). Feature (1) is introduced so that the initial parameters are consistent with the design results for the adjacent representative positions. If consistency is preserved after nonlinear optimization, the responses of the intermediate filters change smoothly with the movement of sound sources.

Although feature (1) is effective for reducing abrupt changes in the intermediate filters, consistency in the design results is sometimes corrupted during nonlinear optimization. Feature (2) preserves this consistency during optimization by minimizing the average of the spectral distances between the responses of the intermediate filters and the linear interpolation of HRTFs. With this feature, intermediate filters linearly interpolate HRTFs, as shown in Fig. 5(b). Thus, both features (1) and (2) facilitate smooth changes in the intermediate filters.

These features are described in the rest of this section.

3.1. Derivation of Initial Parameters

For the reuse of the design results as initial parameters, filter design is conducted in the order shown in Fig. 6. The design starts from a representative position with $\phi = 90^\circ$ (1). The design for other representative positions is conducted in descending order of elevation. In a horizontal circle with the same elevation, the filter coefficients for the ipsilateral side ($\theta = 270^\circ$) are designed first (2), and then the design proceeds toward the contralateral side (3, 3′).

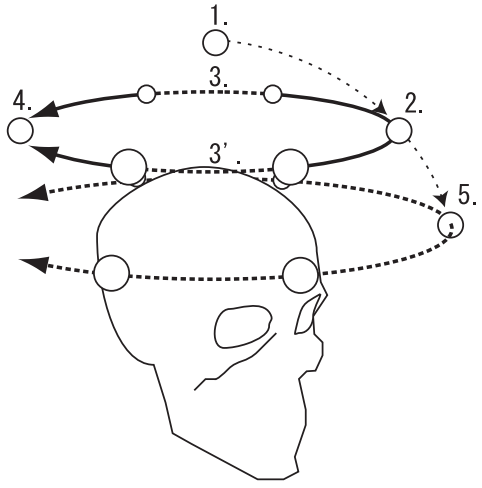

Fig. 6 Design order.

After filter coefficients for the contralateral side ($\theta = 90^\circ$) are designed (4), the design proceeds to the next elevation (5).

For a distinctive sound source position with elevation $\phi = 90^\circ$, the method described in the previous section is utilized to obtain initial parameters $\boldsymbol{x}^0_{\phi,\theta}$. For other sound source positions, the initial parameters are calculated using Eq. (15), where $\hat{\boldsymbol{x}}_{\phi,\theta}$ denotes the design result for a representative position with elevation ϕ and azimuth θ. As already mentioned, ϕ_G and θ_G denote angle steps between representative positions. Therefore, $\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta}$ and $\hat{\boldsymbol{x}}_{\phi,\theta\pm\theta_\mathrm{G}}$ denote design results for the representative positions above the target position and at its side, respectively. If $\phi < 90^\circ$ and $\theta = 270^\circ$, the filter for the representative position above the target position has already been designed, since filter design is conducted in the order shown in Fig. 6. Therefore, design result $\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta}$ is reused as $\boldsymbol{x}^0_{\phi,\theta}$. If $\phi < 90^\circ$ and $\theta \neq 270^\circ$, the design results for the representative positions above it and at its side have already been designed. In these cases, their average is used as $\boldsymbol{x}^0_{\phi,\theta}$, considering consistency with both representative positions.

$$\boldsymbol{x}^0_{\phi,\theta} = \begin{cases} \hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta} & (\theta = 270^\circ), \\ \frac{1}{2}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta} + \frac{1}{2}\hat{\boldsymbol{x}}_{\phi,\theta+\theta_\mathrm{G}} & (90^\circ < \theta < 270^\circ), \\ \frac{1}{2}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta} + \frac{1}{2}\hat{\boldsymbol{x}}_{\phi,\theta-\theta_\mathrm{G}} & (\theta < 90^\circ, 270^\circ < \theta), \\ \frac{1}{2}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta} + \frac{1}{4}\hat{\boldsymbol{x}}_{\phi,\theta+\theta_\mathrm{G}} + \frac{1}{4}\hat{\boldsymbol{x}}_{\phi,\theta-\theta_\mathrm{G}} & (\theta = 90^\circ) \end{cases} \tag{15}$$

Since the comb filter in the high subband is not interpolated, the initial values of its parameters, a_c, b_c, and d_c, which are defined in Eq. (9), are independently designed for each representative position.

3.2. Modification of Evaluation Function

The second idea in the proposed method is to evaluate the response of the intermediate filters in the optimization. In the proposed method, the spectral distance between the response of the intermediate filter and the linear interpolation of the desired HRTFs is evaluated on grid points between the target representative position and the adjacent representative positions. This spectral distance is calculated using the logarithmic error measure defined in Eq. (3), assuming that $\hat{H}(e^{j\omega})$ is the linear interpolation of HRTFs and that $H(e^{j\omega})$ is the intermediate filter on that grid point. The weighted average of these spectral distances on the grid points is used as the evaluation function for the gradient search algorithm. This causes the intermediate filters on the grid points to linearly interpolate the desired HRTFs. If these grid points are densely distributed between the target representative position and the adjacent representative positions, the optimized filter yields a series of smoothly changing interpolation results, with good correspondence of IIR filters between adjacent representative positions.

The expression of the evaluation function depends on the target sound source position. In the rest of this subsection, the evaluation function is defined for several groups of sound source positions.

[$\phi = 90^\circ$]

The normal logarithmic error function ϵ^2 is used as the evaluation function, since no other filter coefficients have been designed at this time.

[$\phi \neq 90^\circ$, $\theta = 270^\circ$]

The coefficients for the representative position above the current position have already been designed. Therefore, the error function is evaluated on a one-dimensional grid between the target position (ϕ, θ) and the representative position $(\phi + \phi_\mathrm{G}, \theta)$, as shown in Fig. 7, so that the design result is consistent with the representative position above the current position. The number of grid points in one dimension is N_b.

In this case, the evaluation function is defined by Eq. (16), where $\boldsymbol{x}_{\phi,\theta}$ is the currently designed parameter vector, w_j $(0 \le j \le N_\mathrm{b} - 1)$ are constant weight values, and $\boldsymbol{x}_{0j}(\boldsymbol{x}_{\phi,\theta})$ is an intermediate filter on a grid point. ϵ^2_V is a weighted average of the error function ϵ^2 on grid points $j = 0 \sim N_\mathrm{b} - 1$ in Fig. 7.

$$\epsilon^2_V(\boldsymbol{x}_{\phi,\theta}, \phi, \theta) = \sum_{j=0}^{N_\mathrm{b}-1} w_j \epsilon^2\left(\boldsymbol{x}_{0j}(\boldsymbol{x}_{\phi,\theta}), \phi + \frac{j\phi_\mathrm{G}}{N_\mathrm{b}}, \theta\right) \tag{16}$$

An intermediate filter $\boldsymbol{x}_{ij}(\boldsymbol{x}_{\phi,\theta})$ on a grid point is defined by Eq. (17), where $-N_\mathrm{b} + 1 \le i \le N_\mathrm{b} - 1$, $0 \le j \le N_\mathrm{b} - 1$. $\boldsymbol{x}_{ij}(\boldsymbol{x}_{\phi,\theta})$ is calculated by linear interpolation among the current design target $\boldsymbol{x}_{\phi,\theta}$ and design results $\hat{\boldsymbol{x}}$ for the adjacent representative positions.

$$\boldsymbol{x}_{ij}(\boldsymbol{x}_{\phi,\theta}) = \left(1 - \frac{i}{N_\mathrm{b}}\right)\left(1 - \frac{j}{N_\mathrm{b}}\right)\boldsymbol{x}_{\phi,\theta} + \frac{i}{N_\mathrm{b}}\left(1 - \frac{j}{N_\mathrm{b}}\right)\hat{\boldsymbol{x}}_{\phi,\theta+\theta_\mathrm{G}}$$

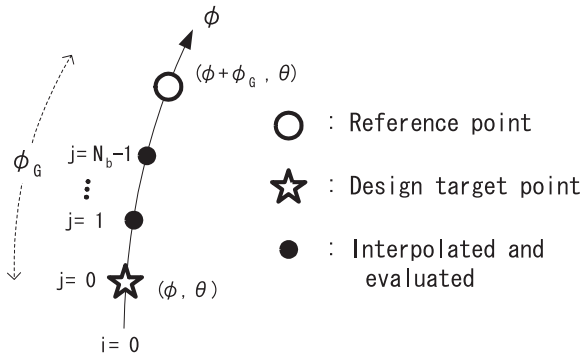

Fig. 7 One-dimensional grid.

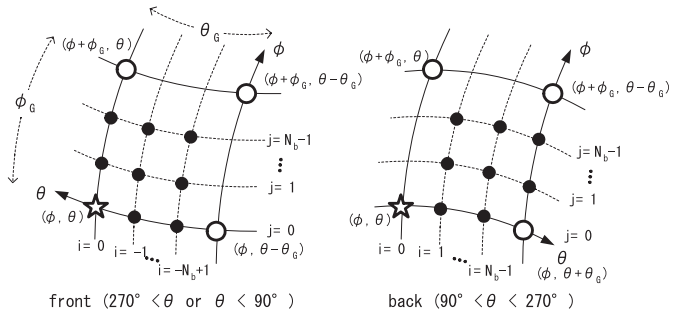

Fig. 8 Two-dimensional grid.

$$+\left(1-\frac{i}{N_\mathrm{b}}\right)\frac{j}{N_\mathrm{b}}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta}+\frac{i}{N_\mathrm{b}}\frac{j}{N_\mathrm{b}}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta+\theta_\mathrm{G}} \quad (i\geq 0)$$

$$\boldsymbol{x}_{ij}(\boldsymbol{x}_{\phi,\theta})=\left(1-\frac{i}{N_\mathrm{b}}\right)\left(1-\frac{j}{N_\mathrm{b}}\right)\boldsymbol{x}_{\phi,\theta}+\frac{i}{N_\mathrm{b}}\left(1-\frac{j}{N_\mathrm{b}}\right)\hat{\boldsymbol{x}}_{\phi,\theta-\theta_\mathrm{G}}$$
$$+\left(1-\frac{i}{N_\mathrm{b}}\right)\frac{j}{N_\mathrm{b}}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta}+\frac{i}{N_\mathrm{b}}\frac{j}{N_\mathrm{b}}\hat{\boldsymbol{x}}_{\phi+\phi_\mathrm{G},\theta-\theta_\mathrm{G}} \quad (i<0) \quad (17)$$

[$\phi\neq 90^\circ$, $\theta\neq 270^\circ$, $\theta\neq 90^\circ$]

The filter coefficients have already been designed for the representative position above the current position, at the side of the current position, and above the representative position at the side of the current position. In this case, linear interpolation among these representative positions and the target position must be considered in the optimization. Therefore, the error function ϵ^2 is evaluated on a two-dimensional grid with N_b intermediate points for each dimension, as shown in Fig. 8. The coefficients for the target position can be made consistent by minimizing the weighted average of ϵ^2 on the grid points.

For the representative positions in front of the listener ($270^\circ < \theta$ or $\theta < 90^\circ$), design proceeds in ascending order of azimuth θ starting from $\theta = 270^\circ$, as shown in Fig. 6. In contrast, those at the back of the listener ($90^\circ < \theta < 270^\circ$) are designed in descending order of θ. Therefore, slightly different evaluation functions, ϵ_F^2 and ϵ_B^2, are used for the positions in front and at the back, respectively. ϵ_F^2 and ϵ_B^2 are weighted averages of the error functions ϵ^2 on the two-dimensional grids shown in Fig. 8.

$$\epsilon_\mathrm{F}^2(\boldsymbol{x}_{\phi,\theta},\phi,\theta)=\sum_{i=-N_\mathrm{b}+1}^{0}\sum_{j=0}^{N_\mathrm{b}-1}w'_{ij}\epsilon^2\left(\boldsymbol{x}_{ij},\phi+\frac{j\phi_\mathrm{G}}{N_\mathrm{b}},\theta+\frac{i\theta_\mathrm{G}}{N_\mathrm{b}}\right) \quad (18)$$

$$\epsilon_\mathrm{B}^2(\boldsymbol{x}_{\phi,\theta},\phi,\theta)=\sum_{i=0}^{N_\mathrm{b}-1}\sum_{j=0}^{N_\mathrm{b}-1}w'_{ij}\epsilon^2\left(\boldsymbol{x}_{ij},\phi+\frac{j\phi_\mathrm{G}}{N_\mathrm{b}},\theta-\frac{i\theta_\mathrm{G}}{N_\mathrm{b}}\right) \quad (19)$$

w'_{ij} ($0\leq i\leq N_\mathrm{b}-1$, $0\leq j\leq N_\mathrm{b}-1$) in the above equations are weight values.

[$\phi\neq 90^\circ$, $\theta=90^\circ$]

As shown in Fig. 6, two design paths (3, 3′) starting from $\theta = 270^\circ$ are connected at the representative position with $\theta = 90^\circ$. Since the representative positions at both sides have already had their filters designed, $\frac{\epsilon_\mathrm{B}^2+\epsilon_\mathrm{F}^2}{2}$ is used as the evaluation function at this position.

The design result depends on the number of grid points N_b. If the grid is sufficiently fine, optimization results show smoothly changing frequency responses with sound movement. Since the calculation time for the whole design is almost proportional to N_b^2, N_b must not be set at too large a value. From the results of informal listening experiments, we found that $N_\mathrm{b} = 5$ is adequate for obtaining smooth sound movement.

4. EVALUATION

A set of digital filters was designed using the proposed method. For comparison, a design using the basic method described in Sect. 2 was also performed.

HRTFs measured with a Head Acoustics HMS dummy head were used for the design. The measurement positions of HRTFs were chosen so that ϕ and θ satisfied the following equations:

$$\phi = m\cdot 15^\circ(-3\leq m\leq 6) \quad (20)$$

$$\theta = n\cdot 15^\circ(0\leq n\leq 23) \quad (21)$$

The design parameters are as follows: the number of 2nd-order IIR filters, $N = 3$; the maximum absolute value of poles and zeros, $r_\mathrm{max} = 0.99$; and sampling rate, $F_\mathrm{s} = 44.1$ kHz; and the number of grid points, $N_\mathrm{b} = 5$. The order of the FIR highpass/lowpass filters $F_\mathrm{L}(z)$ and $F_\mathrm{H}(z)$ is 32 and that of the bandpass filter $F_\mathrm{M}(z)$ is 64. Weight values w_j and w'_{ij} are set to the following values:

$$w_j=\begin{cases}0.2 & (j=0)\\ \dfrac{1-w_0}{N_\mathrm{b}-1} & (\text{otherwise})\end{cases} \quad (22)$$

$$w'_{ij}=\begin{cases}0.2 & (i=0,\ j=0)\\ \dfrac{1-w'_{00}}{N_\mathrm{b}^2-1} & (\text{otherwise}).\end{cases} \quad (23)$$

Optimization by the SQP method is terminated when the evaluation function changes by less than 10^{-5}. Calculation for 205 positions took 60 minutes using the basic method and 770 minutes using the proposed method, using Matlab 6.5 on an IBM-compatible PC with Windows Server 2003 OS, Xeon 3.2 GHz dual processors, and 2 GB SDRAM. The proposed method requires more computation time than the basic method due to error function calculation at the intermediate positions. In the rest of this section, objective evaluation and subjective listening experiments on the designed filters are described.

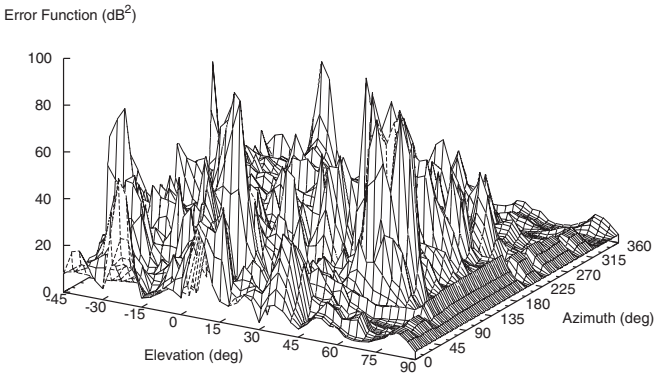

Fig. 9 ϵ^2 using basic method for all directions.

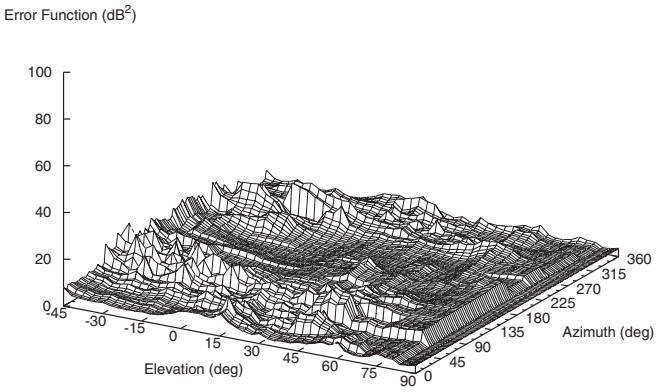

Fig. 10 ϵ^2 using proposed method for all directions.

4.1. Objective Evaluation

The logarithmic error function ϵ^2 is evaluated on a grid with an interval of 1° using both the basic and proposed methods. The method described in Sect. 2.3. is used for the interpolation of filter coefficients. The average of ϵ^2 for all the representative positions is $6.903\,\mathrm{dB}^2$ using the proposed method, which is slightly greater than the average value of $6.600\,\mathrm{dB}^2$ using the basic design method. The average of ϵ^2 on a grid with an interval of 1° for all directions is $5.389\,\mathrm{dB}^2$ using the proposed method, whereas the average using the basic design method is $12.727\,\mathrm{dB}^2$.

The average error on the representative positions is slightly greater using the proposed method than using the basic method due to the constraint on the intermediate filters. As already mentioned, the proposed method considers correspondence between second-order sections and peaks/dips in HRTFs, but this is not considered the basic method. The proposed method gives precedence to correspondence over errors on the representative positions, resulting in an increase in average error on the representative positions. The proposed method successfully decreases the error function of the intermediate filters at the cost of a slight increase in error at the representative positions.

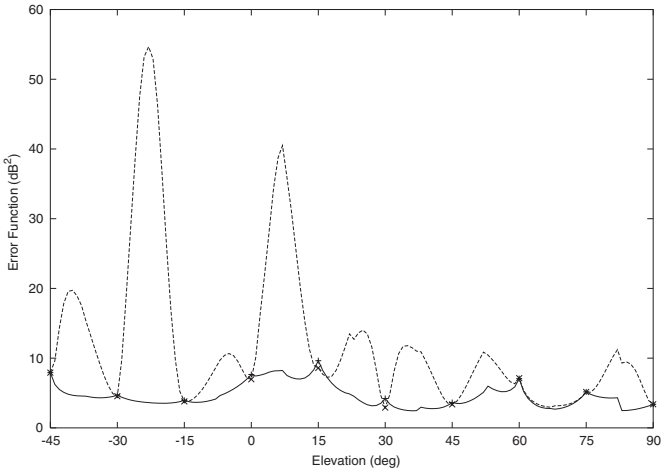

Fig. 11 Comparison of ϵ^2 for both methods ($\theta = 0$).

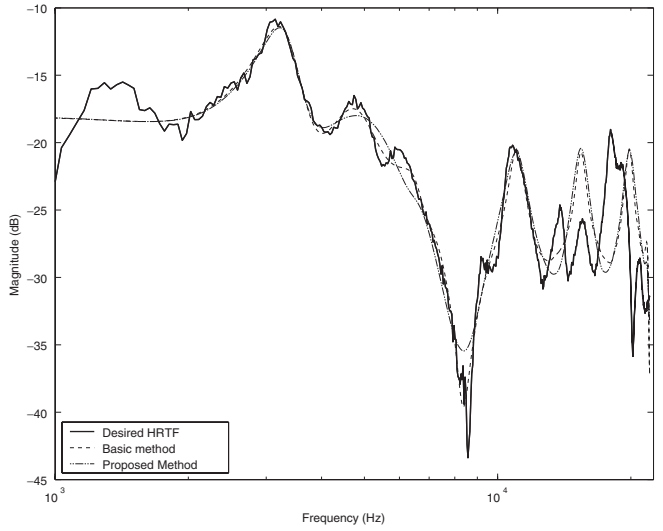

Fig. 12 Response of designed filter at $\phi = 30$, $\theta = 0$ for both methods.

The values of the error function ϵ^2 for all directions using the basic and proposed methods are shown in Figs. 9 and 10. The error function ϵ^2 using both methods for azimuth $\theta = 0^\circ$ is compared in Fig. 11. Although ϵ^2 at the representative positions, indicated by $\times$ and $+$ in Fig. 11, is low for both methods, ϵ^2 for the intermediate filters is very large using the basic method. In contrast, ϵ^2 using the proposed method is relatively constant and low, showing that the proposed method improves interpolation results.

Examples of frequency responses for both methods are also shown in Figs. 12 and 13. The filter responses at the representative position are slightly better using the basic method. In contrast, the spectral distance between the responses of the intermediate filter and the linear interpolation of HRTFs is larger using the basic method than using the proposed method.

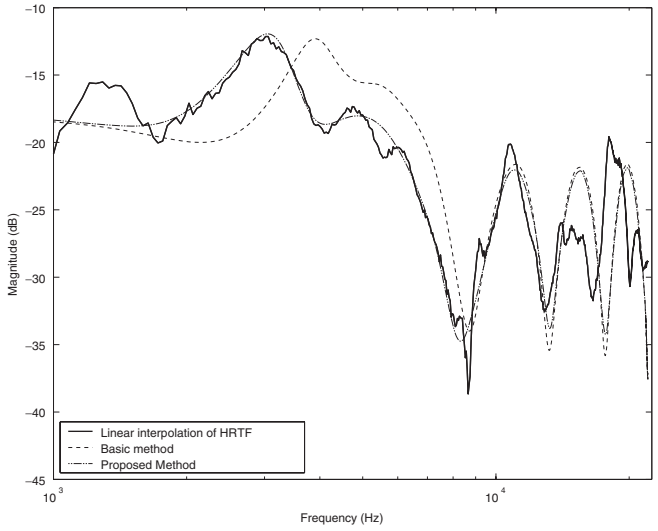

Fig. 13 Response of intermediate filter at $\phi = 37.5$, $\theta = 0$ for both methods.

Table 1 Sound source positions.

ϕ	θ	ϕ	θ	ϕ	θ
90	0	60	45	60	225
30	30	30	120	30	210
30	300	0	0	0	90
0	180	0	270	−30	60
−30	150	−30	240	−30	330

Table 2 Curves of sound sources.

from		through		to	
ϕ	θ	ϕ	θ	ϕ	θ
30	0	0	90	−30	180
−30	0	0	90	30	180
0	270	0	0	0	90
0	270	—	90	0	90
0	0	90	0	0	180

4.2. Subjective Evaluation

In addition to objective evaluation based on comparisons of error functions, two subjective experiments were conducted to evaluate the perceptual quality of the design results. The first experiment deals with the localization of stationary sound images and the second with sound movement.

We followed the experimental method used by Huopaniemi *et al.* [23], which is based on ITU-R BS 1116 [24]. An experimental trial consisted of a stimulus presentation and an response. In each trial, reference (A) and test target (B) processed using different methods are presented one after the other with 0.5 s pauses between samples (i.e., A/B//A/B). Subjects were asked to grade impairment in the localization and timbre of the target against the reference on a 1.0 (very different) to 5.0 (no difference) scale with 0.1 unit steps.

In experiments on the localization of stationary sound images, reference (A) is processed using a 256 tap FIR, which almost perfectly approximates the reference HRTF. Stimulus (B) is processed using a filter designed by either the basic design or the proposed method. Pink noise of 1 s duration with smoothed 50 ms onset and offset is used in this experiment. One experimental block consisted of 30 randomly ordered trials using the two design methods for 15 sound source positions, as shown in Table 1. These sound source positions were chosen from the representative positions so that their azimuth and elevation are uniformly distributed in all directions.

In experiments on sound movement, reference (A) is processed using linear interpolation of FIR filter coefficients and the delay factor, which gives almost ideal sound movement [11,12]. Stimulus (B) is processed using a filter designed by either the basic design method or the proposed method, using the interpolation method described in Sect. 2.3. In either case, filter coefficients are updated on the basis of the current sound source position once every 64 audio samples. In both cases, the computational load for the interpolation of filter parameters based on Eq. (14) is much smaller than that for the filter operation. The computational load for the calculation of reference sound (A) is 24.0×10^6 MAC per second, and the load is 7.8×10^6 MAC per second for stimulus (B).

Pink noise of 3 s duration with smoothed 50 ms onset and offset is used. Sound sources are moved on five representative curves, shown in Table 2, keeping a 1 m distance from the listener. These representative curves are chosen so that they cover various sound source directions and include all front/back, upper/lower, and left/right sound source movements. The angular velocity of the sound movement is 90°/s. One test block consists of 10 randomly ordered trials using the two design methods for five curves.

Eleven volunteer subjects, 20 to 25 years old, participated in the experiment. Each subject responded to two blocks of trials after a short practice block for each experiment. The stimulus sound was presented to them through Sennheiser HD600 headphones. Subjects adjusted the volume during the practice blocks.

Figures 14 and 15 show the 95% confidence intervals of the average scores for all subjects in the above experiments. In the localization of stationary sound images, the proposed method showed a small but significant degradation compared with the basic design method. This result is in agreement with the results of the objective error analysis in the previous subsection, which showed that average error of the representative positions using the

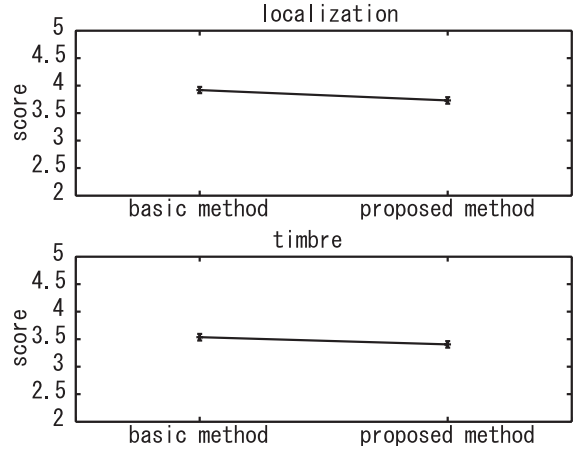

Fig. 14 Results of subjective experiments (stationary sound).

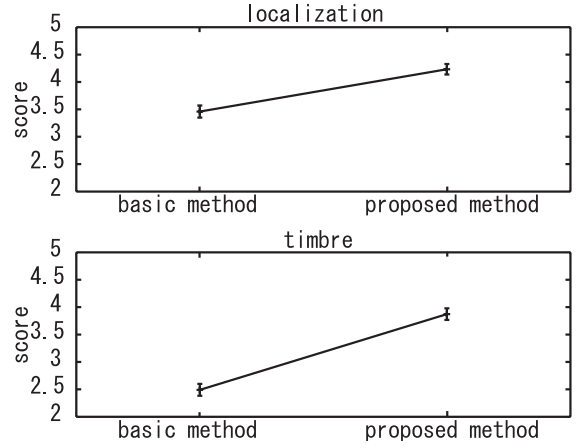

Fig. 15 Results of subjective experiments (sound movement).

proposed method is slightly greater than that of the basic method. In sound movement, however, the proposed method was considerably superior to the basic design method in both localization and timbre. These results demonstrate that the errors of the intermediate filters are reduced using the proposed method. The results of objective and subjective evaluations show that the proposed digital filter design method is effective for improving the subjective quality of sound movement, although the localization of stationary sound sources is slightly degraded.

5. CONCLUSION

An automated method of designing digital filters for 3-D sound movement was described in this paper. In sound movement using existing IIR-based filter architecture, the frequency response of the filter is severely distorted by the runtime interpolation of filter coefficients. The proposed method reduced this distortion by referring to the neighboring filter coefficients in the filter design. Results of objective evaluation and subjective experiments showed that the proposed method improved the subjective quality of 3-D sound movement at the cost of a slight degradation in the localization of stationary sound sources. Since the optimization method utilized in the proposed method does not rely on a specific filter architecture, the proposed method can be easily extended to other filter architectures.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Hiroyuki Ochi and Hiroshi Tsutsui for their advice. The help from Mr. Kazuhiko Furuya, Nobuyuki Iwanaga, Tomoya Matsumura, and Motoki Hirose was also greatly appreciated. This work was partly supported by the Japan Society for the Promotion of Science (JSPS) 21st Century COE Program (Grant No. 14213201).

REFERENCES

[1] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic Press, New York, 1994).
[2] J. Blauert, *Spatial Hearing*, revised edition (The MIT Press, New York, 1997).
[3] F. L. Wightman and D. J. Kisler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.*, **85**, 858–867 (1989).
[4] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," *J. Acoust. Soc. Am.*, **115**, 1714–1728 (2004).
[5] J. Mackenzie, J. Huopaniemi and V. Valimaki, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Process. Lett.*, **4**, 39–41 (1997).
[6] W. Kobayashi, N. Sakamoto, T. Onoye and I. Shirakawa, "3D acoustic image localization algorithm by embedded DSP," *IEICE Trans. Fundam.*, **E84-A**, 1423–1430 (2001).
[7] K. Tsujino, K. Furuya, W. Kobayashi, T. Izumi, T. Onoye and Y. Nakamura, "Design of realtime 3-D sound processing system," *IEICE Trans. Inf. Syst.*, **E88-D**, 954–962 (2005).
[8] E. M. Wenzel, J. Miller and J. S. Abel, "A software-based system for interactive spatial sound synthesis," *Proc. 2000 Int. Conf. Auditory Display* (*ICAD 2000*), pp. 151–156 (2000).
[9] D. N. Zotkin, "Customizable auditory display," *Proc. 2002 Int. Conf. Auditory Display* (*ICAD02*), Vol. 2, pp. 2113–2116 (2002).
[10] L. Savioja, J. Huopaniemi, T. Lokki and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, **47**, 675–705 (1999).
[11] E. M. Wenzel and S. H. Foster, "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," *Proc. ASSP* (*IEEE*) *Workshop on Applications of Signal Processing to Audio and Acoustics* (1993).
[12] M. Matsumoto, M. Tohyama and H. Yanagawa, "A method of interpolating binaural impulse responses for moving sound images," *Acoust. Sci. & Tech.*, **24**, 284–292 (2003).
[13] T. Nishino, S. Kajita, K. Takeda and F. Itakura, "Interpolation of the head related transfer function on the horizontal plane," *J. Acoust. Soc. Jpn.* (*J*), **55**, 91–99 (1999).
[14] T. Nishino, S. Kajita, K. Takeda and F. Itakura, "Interpolation of the head related transfer functions of azimuth and elevation," *J. Acoust. Soc. Jpn.* (*J*), **57**, 685–692 (2001).
[15] K. Watanabe, S. Takane and Y. Suzuki, "A novel interpolation method of HRTFs based on the common-acoustical-pole and zero model," *Acta Acust. - Acustica*, **91**, 958–966 (2005).

[16] T. Kobayashi and S. Imai, "Design of IIR digital filters with arbitrary log magnitude function by WLS techniques," *IEEE Trans. Acoust. Speech Signal Process.*, **38**, 247–252 (1990).

[17] M. A. Blommer and G. H. Wakefield, "On the design of pole-zero approximations using a logarithmic error measure," *IEEE Trans. Signal Process.*, **42**, 3245–3248 (1994).

[18] T. Matsumura, N. Iwanaga, W. Kobayashi, T. Onoye and I. Shirakawa, "Feature extraction of head-related transfer function for 3D sound movement," *Proc. 2003 Int. Tech. Conf. Circuits/Systems, Computers and Communications* (*ITC-CSCC2003*), pp. 685–688 (2003).

[19] J. M. Jot, V. Larcher and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," *98th Conv. Audio Eng. Soc. Prepr.*, 3980 (1995).

[20] B. C. J. Moore, *An Introduction to the Psychology of Hearing* (Academic Press, New York, 1997).

[21] V. R. Algazi, R. O. Duda, R. P. Morrison and D. M. Thompson, "Structural composition and decomposition of HRTFs," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 103–106 (2001).

[22] K. Tsujino, W. Kobayashi, T. Onoye and Y. Nakamura, "Automated design of digital filters for 3-D sound localization in embedded applications," *Proc. IEEE Int. Conf. Audio, Speech, and Signal Processing*, V-349–V-352 (2006).

[23] J. Huopaniemi, N. Zacharov and M. Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," *105th Conv. Audio Eng. Soc. Prepr.*, 4805 (1998).

[24] ITU-R, "Recommendation BS.1116. methods for the subjective assessment of small impairments in audio systems including multichannel sound systems" (1994).