AUTOMATED DESIGN OF DIGITAL FILTERS FOR 3-D SOUND LOCALIZATION IN EMBEDDED APPLICATIONS

Kosuke Tsujino†, Wataru Kobayashi††, Takao Onoye‡ and Yukihiro Nakamura†

† Dept. Communications and Computers Eng., Kyoto University, Japan
E-mail: tsujino@easter.kuee.kyoto-u.ac.jp
†† Arnis Sound Technologies, Co., Ltd., Japan
‡ Dept. Information Systems Eng., Osaka University, Japan

ABSTRACT

In this paper, an automated design method of digital filters for 3-D sound localization is proposed, which approximates head-related transfer function (HRTF). Target digital filter organization is dedicated for embedded applications so that ROM capacity and computational load are reduced with only slight degradation of 3-D sound effect. Proposed design method consists of two steps. In the first step, the least-squares method is adopted to acquire initial design result quickly, while the second step utilizes the gradient search algorithm for precise optimization. In an objective evaluation, the proposed two-step method is proved to be effective for improving design results. Results of subjective listening tests also show that the proposed method is comparable to manual design in the approximation of reference HRTF. In these subjective tests, our automated design achieved better results per ROM capacity than simple FIR filter did.

1. INTRODUCTION

The aim of 3-D sound localization is to give virtual impression of sound source elevation, azimuth and distance to listeners through 2-channel stereo sound. For this purpose, 3-D sound localization systems must simulate auditory cues which are utilized for human judgment of sound source position. Most of these cues are contained in a pair of acoustical transfer functions from the sound source position to the left/right ear of the listener. These transfer functions are called head-related transfer functions (HRTFs). A pair of digital filters which approximate left/right HRTFs are usually utilized in 3-D sound localization. Although other features such as head tracking and simulation of room transfer fucntion are also important for the presence of 3-D sound, we focus on the modelling of HRTFs in this paper.

There are tradeoffs between the approximation accuracy of HRTF and computational resources required in 3-D sound localization. The most precise approximation of an HRTF can be obtained by reproducing head-related impulse response (HRIR)[4], which is time-domain representation of HRTF, by high-order FIR filter. However, this method requires much computational cost and ROM capacity in processing systems. Reduction of computational cost and ROM capacity with minimum degradation of the precision of HRTF approximation is required in embedded applications such as mobile phones and portable audio players.

From this point of view, a novel digital filter organization for 3-D sound localization has been proposed by Kobayashi et al.[1], which reduces ROM capacity and computational cost considerably with only slight degradation of sound effect. Since conventional automated design method cannot be applied to this unique digital filter organization, filter coefficients for this digital filter organization is manually designed. However, this manual design for a complete set of HRTF measurement requires much time of a skilled engineer. In this paper, automated filter design method is proposed for this digital filter organization in order to facilitate quick analysis and tuning. As a matter of fact, the manual design process includes not only the approximation of each HRTF, but also overall adjustment of timbre among source positions for natural sound localization and sound movement. As a first step, simple approximation of each HRTF is automated in this paper without consideration of overall consistency.

2. DIGITAL FILTER DESIGN METHOD

Digital filters with complicated frequency response like those of HRTFs are generally designed by minimizing an error between desired and realized frequency responses. Logarithmic error measure is broadly accepted as a good measure in acoustical applications[2]. Logarithmic error $\epsilon_{\log}(\mathbf{x})$ is defined by the following equation, where $\mathbf{x}$ is the parameter vector, $\hat{H}(e^{j\omega})$ is desired response, $H(e^{j\omega}, \mathbf{x})$ is realized response and $W(e^{j\omega})$ is weighting function.

$$\epsilon_{\log}(\mathbf{x})^2 = \int_{-\pi}^{\pi} \left|W(e^{j\omega})\right|^2 \left|\log|\hat{H}(e^{j\omega})| - \log|H(e^{j\omega}, \mathbf{x})|\right|^2 d\omega \tag{1}$$

Unfortunately, minimizing $\epsilon_{\log}(\mathbf{x})$ leads to a nonlinear problem.

Several methods have been proposed for the design of digital filters based on the logarithmic error measure. Blommer et al.[2] utilized the gradient search algorithm for nonlinear optimization. On the other hand, Kobayashi et al.[3] and Kulkarni et al.[4] have proposed design methods for IIR filters by iteration of linear problems. The gradient search algorithm is applicable to various filter organizations, while the other methods depend on the specific form of the problem. Therefore, the gradient search algorithm is adopted in

This work was partly supported by the Japan Society for the Promotion of Science (JSPS) the 21st Century COE Program (Grant No. 14213201).

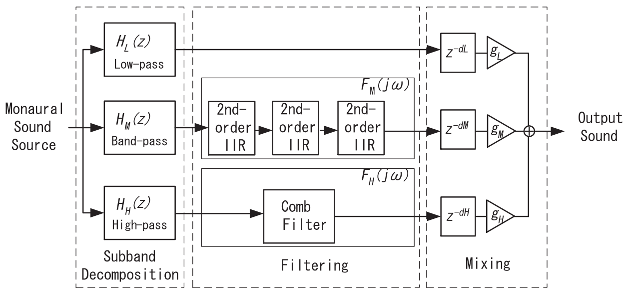

Fig. 1. Digital filter organization for 3-D sound localization

the proposed method.

In the gradient search algorithm, the optimization result generally depends on the initial value of the parameter vector. If the initial value is chosen in the neighborhood of optimum solution, the result of gradient search is expected to be better than the result from an arbitrary initial value. Therefore, the proposed design method consists of the following two steps: 1) rough design of the initial parameters, and 2) gradient search from the pre-designed initial parameters. In the rest of this section, target digital filter organization and each step in the proposed method are described.

2.1. 3-D Sound Localization for Embedded Applications

The digital filter organization for 3-D sound localization in embedded applications[1] is briefly introduced in this subsection. As shown in Fig. 1, this filter organization consists of three stages: subband decomposition, filtering, and mixing. In the subband decomposition stage, input sound is splitted into three subbands using band-pass filters. Next, in the filtering stage, each band-limited signal is processed with different digital filters, considering the difference of physical characteristics and psychological importance of HRTF among these subbands. Finally, three subbands are mixed after the adjustment of gain and delay, which are also important psychological cues for human spatial hearing.

Each of band-pass filters $H_L(z)$, $H_M(z)$, $H_H(z)$ is currently implemented by FIR filter, whose cutoff frequency is derived from the analysis of HRTFs[5]. $F_M(z)$, the digital filter for the processing of the intermediate subband, comprises serial connection of N 2nd-order IIR filters and $F_H(z)$, the filter in high subband, is a comb filter as illustrated in Fig. 2. Transfer functions of $F_M(z)$ and $F_H(z)$ are as follows:

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

$$F_H(z) = \frac{G_d + (G_e - G_f)z^{-D}}{1 - G_f z^{-D}} \tag{3}$$

Since gain factors g_M and g_H are multiplied to the output of $F_M(z)$ and $F_H(z)$, assuming $b_{0n} = 1$ and $G_d = 1$ does not lose the generality of the model. In the rest of this paper, we assume $b_{0n} = 1$ and $G_d = 1$.

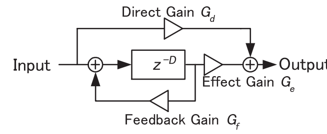

Fig. 2. Comb filter

2.2. Design of initial parameters

Let us introduce weighting function $W(e^{j\omega})$ in order to approximate frequency dependence of human hearing. In this paper, $W(e^{j\omega})$ has a value of 1 between 100Hz and 16kHz and has a value of 0.05 outside that range in order to approximate human audible range.

Initial values of gain factors g_L, g_M, g_H and delay factors d_L, d_M, d_H are set to the weighted average of gain and phase delay of the desired HRTF in the pass-band of corresponding band-pass filters, as described in Eq. (4) and Eq. (5). In Eq. (4) and Eq. (5), $\mathrm{pd}(\hat{H}(e^{j\omega}))$ denotes the phase delay of the desired response $\hat{H}(e^{j\omega})$ and suffix $\mathcal{X}$ is one of L, M, H.

$$g_{\mathcal{X}} = \exp\left(\frac{\int_0^\pi \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| \log|\hat{H}(e^{j\omega})|d\omega}{\int_0^\pi \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| d\omega} \right) \tag{4}$$

$$d_{\mathcal{X}} = \frac{\int_0^\pi \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| \mathrm{pd}(\hat{H}(e^{j\omega}))d\omega}{\int_0^\pi \left|W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega})\right| d\omega} \tag{5}$$

In the design of $F_M(e^{j\omega})$ and $F_H(e^{j\omega})$, a simplified error measure defined by the following equation is utilized.

$$\sigma_{\mathcal{X}}{}^2 = \int_0^\pi \left|W_{\mathcal{X}}(e^{j\omega})\right|^2 \left| \frac{\hat{H}(e^{j\omega})}{g_{\mathcal{X}}} A_{\mathcal{X}}(e^{j\omega}) - B_{\mathcal{X}}(e^{j\omega}) \right|^2 d\omega$$
$$\text{where} \quad F_{\mathcal{X}}(z) = \frac{B_{\mathcal{X}}(z)}{A_{\mathcal{X}}(z)},\ W_{\mathcal{X}}(e^{j\omega}) = W(e^{j\omega})H_{\mathcal{X}}(e^{j\omega}) \tag{6}$$

Since input of $F_{\mathcal{X}}(z)$ is band-limited by $H_{\mathcal{X}}(z)$, error outside the pass-band of $H_{\mathcal{X}}(z)$ is ignored by multiplying $H_{\mathcal{X}}(e^{j\omega})$ to weighting function $W(e^{j\omega})$. As already shown in Eq. (2) and Eq. (3), $A_{\mathcal{X}}(z)$ and $B_{\mathcal{X}}(z)$ are linear expression of unknown variables. Therefore, the least-squares method can be utilized to get $a_{1n}, a_{2n}, b_{1n}, b_{2n}, G_e$, and G_f which minimize $\sigma_{\mathcal{X}}$.

Delay factor D of the comb filter $F_H(z)$ must be decided within the integer values according to the pattern of peaks and dips in the high-frequency component of the desired HRTF. In the proposed method, design of G_e and G_f are tried for each integer value of D from 2 to 25, and then the set of D, G_e, and G_d which gives minimum value of σ_H is adopted in the initial values.

2.3. Optimization by the Gradient Search Algorithm

First, parameter vector $\mathbf{x}$ used in the gradient search is to be defined. There are some ways for the parametrization of Eq. (2) and Eq. (3). If filter coefficients $a_{1n}, a_{2n}, b_{1n}, b_{2n}, G_e$, and G_f are chosen as optimization parameters, complicated form of stability condition makes the design difficult. Therefore, Eq. (2) and Eq. (3) are converted into Eq. (7) and Eq. (8), 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 zeros, respectively. In Eq. (8), variables A and B also control the absolute value of poles and zeros. This conversion allows simple expression of stability condition by defining the maximum value of the absolute value of the poles.

$$F_M(z) = \prod_{n=0}^{N-1} \frac{(1 - r_{\beta n} e^{j\omega\theta_{\beta n}} z^{-1})(1 - r_{\beta n} e^{-j\omega\theta_{\beta n}} z^{-1})}{(1 - r_{\alpha n} e^{j\omega\theta_{\alpha n}} z^{-1})(1 - r_{\alpha n} e^{-j\omega\theta_{\alpha n}} z^{-1})} \tag{7}$$

$$F_H(z) = \frac{1 - Bz^{-D}}{1 - Az^{-D}} \tag{8}$$

Since the delay factor D of the comb filter is integer, which cannot be treated as continuous value, D is not included into the parameter vector. Thus as for D, the initial value is utilized without change. Although delay factors $d_{\mathcal{X}}$ are also integer originally, they can be treated as continuous values using fractional delay digital filters[6]. Therefore, they are included into parameter vector and rounded to the nearest integer after optimization if necessary. As a result of above discussion, parameter vector $\mathbf{x}$ is defined as follows:

$$\mathbf{x} = (g_L, g_M, g_H, d_L, d_M, d_H, \{r_{\alpha n}, \theta_{\alpha n}\}, \{r_{\beta n}, \theta_{\beta n}\}, A, B) \tag{9}$$

The logarithmic error $\epsilon_{\log}(\mathbf{x})^2$ defined in Eq. (1) is minimized with regard to $\mathbf{x}$ using the gradient search algorithm. As already mentioned, 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 value of the absolute value of poles and zeros.

$$0 \le d_{\mathcal{X}} \tag{10}$$

$$0 \le r_{\alpha n}, r_{\beta n} \le r_{\max} \tag{11}$$

$$0 \le A, B \le r_{\max} \tag{12}$$

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

In the above optimization, the phase response of the design result is not taken into account. Error in phase response distorts interaural time difference (ITD), which is also an important cue for sound localization. In order to decrease this error, delay factors $d_{\mathcal{X}}$ are adjusted after optimization. In this paper, delay of a digital filter is defined by the time between the input of single impulse to the filter and the maximum output from the filter after that input. An identical offset d_{o} is added to $d_{\mathcal{X}}$ so that the delay of the designed filter is equal to that of the reference HRIR. The amplitude response of the designed filter is not affected by this adjustment, since relation among d_L, d_M, d_H is not changed by adding identical offset.

By repeating this design procedure for a set of HRTFs of various source positions, a set of digital filter coefficients for 3-D sound localization can be derived. Another idea for further improvement is described here. If optimization result is already derived from another source position in the neighborhood, that result tends to give better initial value for the gradient search than the pre-designed initial value. Therefore, actual design is performed sequentially in an order described in Fig. 3. In positions other than azimuth = 0° and elevation = 0°, average of design result $\mathbf{x}$ for adjacent positions are reused as the initial values for $r_{\alpha n}$, $\theta_{\alpha n}$, $r_{\beta n}$, $\theta_{\beta n}$, $g_{\mathcal{X}}$, and $d_{\mathcal{X}}$. The parameters of the comb filter A, B, and D are excluded from this reuse since delay factor D of comb filter is not optimized in the gradient search. In order to fit the characteristics of each HRTF, initial values of A, B, and D are designed independently for all positions. Both of calculation time and design result are improved with this sequential design.

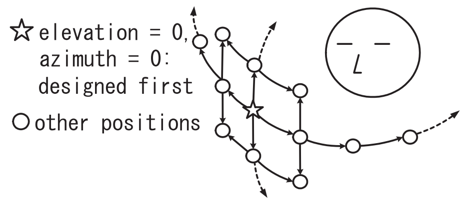

Fig. 3. Design order

3. EVALUATION

Actual design of digital filters was performed with the proposed method. HRTFs measured with Head Acoustics HMS dummy head[5] for 205 positions in 1m distance were used for design. Design parameters were as follows: the number of 2nd-order IIR filters $N = 3$, the maximum of absolute value of poles and zeros $r_{\max} = 0.99$, and the sampling rate $F_s = 44.1$kHz. Optimization by SQP method is terminated when error function changes by less than 10^{-5}. Calculation for 205 positions took about 60 minutes using Matlab 6.5 on an IBM-compatible PC with Windows Server 2003 OS, Xeon 3.2GHz dual processors, and 2GB SDRAM. In the rest of this section, objective evaluation and subjective evaluation of the designed filters are described.

3.1. Objective Evaluation

Average error of design results from reference HRTFs of the following two methods is illustrated in Fig. 4.

- two-step automated design (proposed method)
- optimization by SQP method from identical initial parameters

The initial parameters utilized in the latter method is chosen so that the filter shows flat frequency response. The proposed method using two-step design yielded better results than the gradient search from the identical initial parameters did. The average of the error function $\epsilon_{\log}(\mathbf{x})^2$ after optimization was 6.06 with the proposed method, while the average result from the identical initial parameters was 16.06. From these results, proposed two-step design method is proved to be effective for improving design results.

3.2. Subjective Evaluation

Listening tests were performed to evaluate subjective quality of the designed filters. In addition to the automatically designed filters, 32-tap and 96-tap FIR was tested for comparison. The manually designed filters were also added for reference. 32-tap FIR consumes almost equal ROM space to the

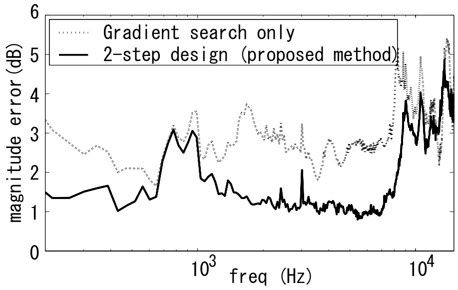

Fig. 4. Error of design results

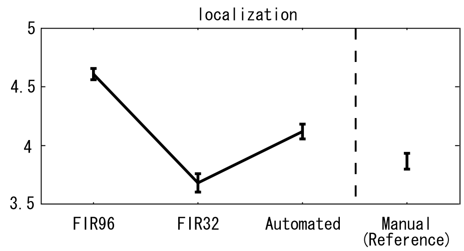

Fig. 5. Results of listening tests

target filter organization, while 96-tap FIR requires almost the same computational load as the target organization. The subjects were asked to grade localization impairment against the reference on a 1.0 (very different) to 5.0 (no difference) scale with 0.1 unit steps. A trial consisted of a stimulus presentation and an answer. In each trial, reference (A) and test target (B) processed with different methods are presented one after the other with 0.5s pause between samples(i.e. A/B//A/B). Pink noise of one-second duration with smoothed 50ms onset and offset was used in the tests. The reference was processed with 256-tap FIR, which almost perfectly approximates the characteristic of reference HRTF. The test target was processed one of the following four filters: the automatically designed filter, 32-tap FIR, 96-tap FIR, and the manually designed filter (for reference). Coefficients of 32-tap and 96-tap FIR were cut off from HRIR using rectangular window. One test block consisted of 60 trials with these four processing methods for 15 uniformly distributed sound source positions.

Six male volunteer subjects with ages between 20 and 25 participated in the experiment. Each subject answered two blocks of trials after a short practice block. Stimulus sound was presented to the subject through Sennheiser HD600 headphone. The subject adjusted the volume of sound during the practice block.

Fig. 5 shows 95% confidence interval of average score on localization for all subjects. The automatically designed filters were superior to 32-tap FIR. Therefore, in the similarity to the reference HRTF, the proposed method achieved better performance than the manual design and FIR filter with almost equal ROM consumption. Note that the automatically designed filter records even higher score than the referenced manually designed filters. As already mentioned, this is mainly because the manual design is accomplished with referring to many design factors including the approximation of HRTF, the adjustment of timbre for natural sound localization, and the smoothness of moving sound image.

4. CONCLUSION

An automated design method of digital filter for 3-D sound localization is proposed in this paper. The target digital filter organization is dedicated for embedded application with reduced ROM consumption and computational load. The proposed method consists of two steps. The first step is rough design based on the least-squares error measure, and the second step is optimization using the gradient search algorithm with the logarithmic error measure. Experimental results show that the filters designed by the proposed method approximate the desired HRTF better than FIR filter with equal ROM consumption.

The proposed design method does not take sound movement into account, although movement of virtual sound source is often required in actual application of 3-D sound[7, 8]. Further improvement of the proposed design method for sound movement is our future task.

5. REFERENCES

[1] W. Kobayashi, N. Sakamoto, T. Onoye, and I. Shirakawa, "3D acoustic image localization algorithm by embedded DSP," *IEICE Trans. Fundamentals, Vol. E84-A, No.6*, pp. 1423–1430, 2001.

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

[3] T. Kobayashi and S. Imai, "Design of IIR digital filters with arbitrary log magnitude function by WLS techniques," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 247–252, 1990.

[4] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 115, pp. 1714–1728, 2004.

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

[6] T. I. Laakso, V. Valmaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay," *IEEE Signal Processing Magazine*, pp. 30–60, Jan. 1996.

[7] K. Tsujino, W. Kobayashi, T. Izumi, T. Onoye, and Y. Nakamura, "Realtime filter redesign for interactive 3-D sound systems," in *Proc. 2004 IEEE Region 10 International Conference (TENCON2004)*, Paper No. 1024, 2004.

[8] K. Tsujino, K. Furuya, W. Kobayashi, T. Izumi, T. Onoye, and Y. Nakamura, "Design of realtime 3-D sound processing system," *IEICE Trans. on Inf. and Syst.*, vol. E88-D, pp. 954–962, 2005.