A DUAL-MODE ARCHITECTURE FOR HEADPHONES DELIVERING SURROUND SOUND: LOW-ORDER IIR FILTER MODELS APPROACH

S.N. Yao, Student Member, IEEE, T. Collins, and P. Jančovič
School of Electronic, Electrical and Computer Engineering, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK
SXY043@bham.ac.uk

ABSTRACT

In this paper we propose a method for conventional headphones to reproduce 5.1 and 7.1-channel surround sounds that are increasingly used in music and movies. When converting multi-channel audio to stereo, we adopted shuffler filter structures and least-squares approximation of FIR (Finite Impulse Response) by IIR (Infinite Impulse Response) filters in order to lower the computational cost and memory usage. By taking advantage of shuffler filters, the architecture is easy to be extended, thus compatible with discrete 5.1 and 7.1-channel sources. The algorithm in the least-squares sense enables the designed low-order IIR filters to approach the long-length HRTFs (Head Related Transfer Functions) with small approximation errors. As a result, the proposed architecture achieves not merely low cost but high performance, allowing headphones to deliver similar surround sound as heard in home theatres.

1 INTRODUCTION

With the advent of modern techniques, there are increasing audio and video discs encoded with discrete 5.1 or even 7.1-channel formats. Though it is easy to reproduce multi-channel audio in home theatre systems, consumers have difficulty in playing music or movies with 5.1 or 7.1-channel soundtracks when using mobile electronic appliances, such as smartphones, MP3/4 players, and portable DVD players. Therefore, it is necessary to design a system delivering spatial sound from a set of headphones.

The previous studies [1-4] have proposed several kinds of architectures based on HRTFs to downmix 5 or 5.1-channel inputs into 2-channel outputs. Designing the architecture of the HRTF-based downmixing method, Bai [1] implemented HRTFs released from MIT media lab without reducing any computational burden and memory usage. Fujinami [2] arranged 10 FIR-based equalizers to localize 5-channel sound sources, each of them consisting of two FIR filters. Because of taking reverberation into consideration, the size of his FIR filters is the largest. You [3] used a dynamic programming algorithm to calculate the optimal length of each FIR filter. This approach requires more than 200 coefficients to render 5-channel audio. Unlike the others, Kawano [4] used IIR filters to greatly reduce the processing time. However, the drawback of his architecture is that very low-order IIR filters were used which hardly approach the performance of the ordinary FIR filters. In this paper, we propose an architecture which allows consumers to enjoy not only 5.1 but 7.1-channel surround experience over conventional headphones by using few low-order IIR filters. This architecture can be implemented with low memory and computational cost. Also, the IIR filters present comparable performance to that of previous FIR filter models.

2 SOUND IMAGE LOCALIZATION FOR HEADPHONES

When hearing a sound, a listener can distinguish the location of the source by perceiving the differences in magnitude and phase of the sound between the ears. This is illustrated in Fig. 1.

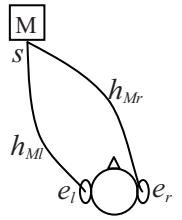

Figure 1. A pair of HRTFs in the transaural listening.

While s is the sound coming from the loudspeaker M, e_l and e_r are the sounds reaching the listeners' left and right eardrums, respectively. So, e_l and e_r can be represented as

$$e_l = s * h_{Ml} \quad (1)$$

$$e_r = s * h_{Mr} \quad (2)$$

In (1) and (2), "$*$" symbolizes convolution operator. h_{Ml} and h_{Mr} denotes a pair of HRTFs containing the spectral characteristic of sound affected by the head, pinna, shoulder, and torso. We can extend the situation from mono to multi-channel sound as shown in Fig. 2, so e_l and e_r will become (3) and (4), where h_{Xy} is the HRTF between the loudspeaker X (C, L, R, LS, RS, LB, or RB) and the listener's ear y (l or r), and s_X denotes the original audio signal produced by the loudspeaker X.

978-1-61284-842-6/11/$26.00 ©2011 IEEE

$$e_l = s_C * h_{Cl} + s_L * h_{Ll} + s_R * h_{Rl} + s_{LS} * h_{LSl} + s_{RS} * h_{RSl} + s_{LB} * h_{LBl} + s_{RB} * h_{RBl} \quad (3)$$

$$e_r = s_C * h_{Cr} + s_L * h_{Lr} + s_R * h_{Rr} + s_{LS} * h_{LSr} + s_{RS} * h_{RSr} + s_{LB} * h_{LBr} + s_{RB} * h_{RBr} \quad (4)$$

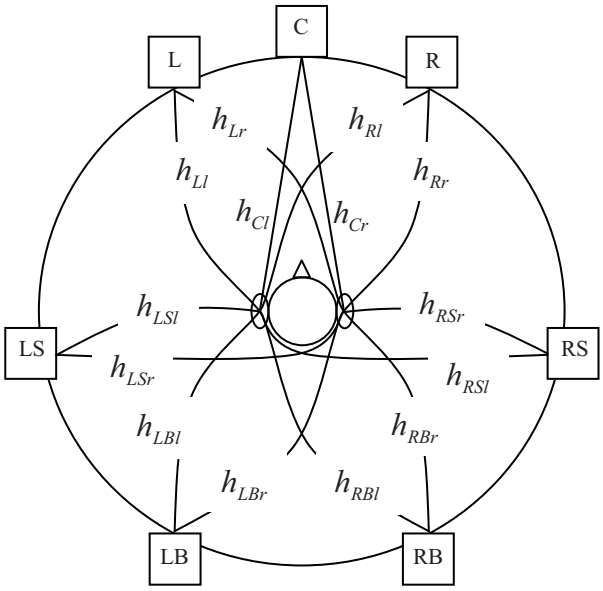

Figure 2. A multi-channel surround system.

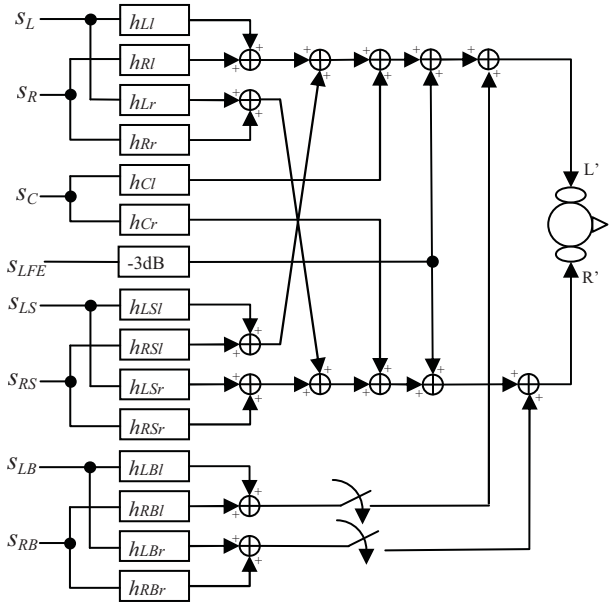

Figure 3. The architecture for a headphone delivering 5.1 or 7.1 surround sound.

Fig. 3 shows the general block diagram of a conventional headphone rendering 7.1-channel audio. The sound images simulate 7 loudspeakers clockwise located at 0°, 30°, 110°, 150°, 210°, 250°, and 330°, the layout which is the best setup for a multi-channel surround system proposed in ITU-R BS.775-1 [5] and by Dolby Laboratories [6]. LFE (Low-Frequency Effects) is the signal from a subwoofer called ".1" channel, the speaker which should be placed in a corner, but according to the idea from [4] and [5], we can simply use a gain, -3dB, to substitute for the characteristic of the desired transfer function between the source and a listener.

3 THE PROPOSED ARCHITECTURE FOR HEADPHONES

Because of symmetric loudspeaker geometries, the HRTF from θ° to a centre listener's left ear is similar to that from $(360 - \theta)^\circ$ to the right ear. Thus, h_{Cl} is very similar to h_{Cr}; h_{Rl} to h_{Lr}; h_{Rr} to h_{Ll}; h_{RSl} to h_{LSr}; h_{RSr} to h_{LSl}; h_{RBl} to h_{LBr}; h_{RBr} to h_{LBl}. By adopting symmetric solution, the number of coefficients of HRTFs could be decreased by 50%. Take the pair of loudspeakers, L and R, as an illustration. While Fig. 4(a) shows the original implementation, Fig. 4(b) represents the modified architecture based on a symmetrical set of HRTFs. From these two figures, we can notice that

$$h_{Ll} = h_{Rr} = h_i \quad (5)$$

$$h_{Lr} = h_{Rl} = h_c \quad (6)$$

where h_i is called the ipsilateral term; h_c, the contralateral term. While symmetrical sets reduce the data storage by sharing coefficients of the filters, the number of the filters maintains the same. Thereby, shuffler filter solution [7] is employed to simplify the symmetric system further. Fig. 5 shows the shuffler filter structure whose outputs are the same as those of the architecture shown in Fig. 4(b). It can be seen that the use of shuffler structure only requires two individual filters. The 7.1 architecture presented in Fig. 3 is using the shuffler structure as shown in Fig. 6. Each shuffler structure contains two different filters, and as such a 7.1-channel surround system required only 7 filters.

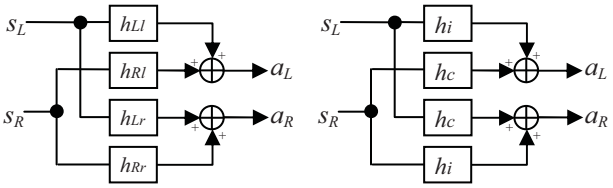

Figure 4. Symmetric solution. (a) A part of the block diagram from Fig. 3. (b) The modified architecture based on a symmetrical set of HRTFs.

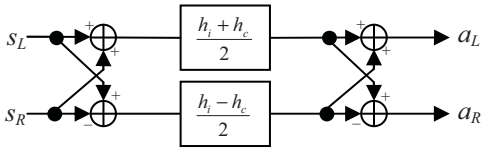

Figure 5. The shuffler filter structure.

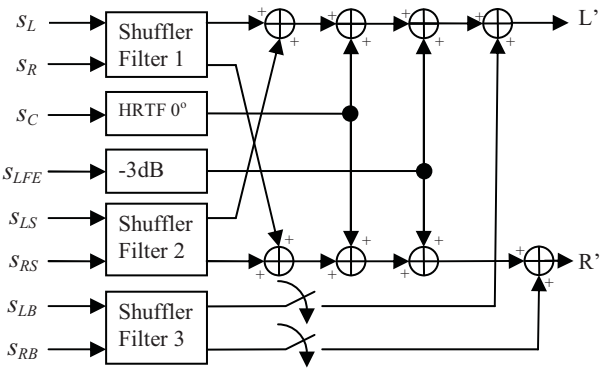

Figure 6. The proposed architecture for a headphone delivering 5.1 or 7.1 surround sound.

4 LOW-ORDER IIR FILTER MODELS OF THE SHUFFLER FILTERS

Using FIR filters to implement HRTFs is straightforward. However, not only does the long length of the FIR filter cause high computational cost, but the memory size required for the coefficients of filters is also substantial. So, we design tenth-order IIR filters on the basis of least-squares approximation for the shuffler filters. The much more detailed description of least-squares approximation of FIR by IIR filters can be found in [8], and thus only the main idea is given in this paper. $F(z)$ and $H(z)$ are the transfer functions of an FIR and IIR digital filter, respectively. By the algorithm, we can determine the numerator and denominator coefficients, minimizing the l_2-norm of

$$\Delta(z) = F(z) - H(z) \tag{7}$$

The l_2-norm of $\Delta(z)$ is defined by

$$\left\|\Delta(z)\right\|_2 = \left[\frac{1}{2\pi j}\oint_{|z|=1} \Delta(z)\Delta^*(z)\frac{dz}{z}\right]^{\frac{1}{2}} \tag{8}$$

where Δ^* is the complex conjugate of Δ. When determining the denominator coefficients of the corresponding IIR filter, we have insufficient information, therefore having to do the iterative procedure in the least-squares sense until the approximation error converging. Deciding the denominator, then we could solve the numerator coefficients by (7).

The approximation errors of these kinds of filters are similar to those of the filters using balanced model truncation [9], but owing to the simplicity of the algorithm, the computational complexity of designing the interesting IIR filters is considerably smaller.

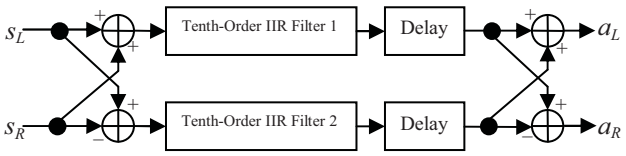

Figure 8. The proposed shuffler filter structure using IIR filters.

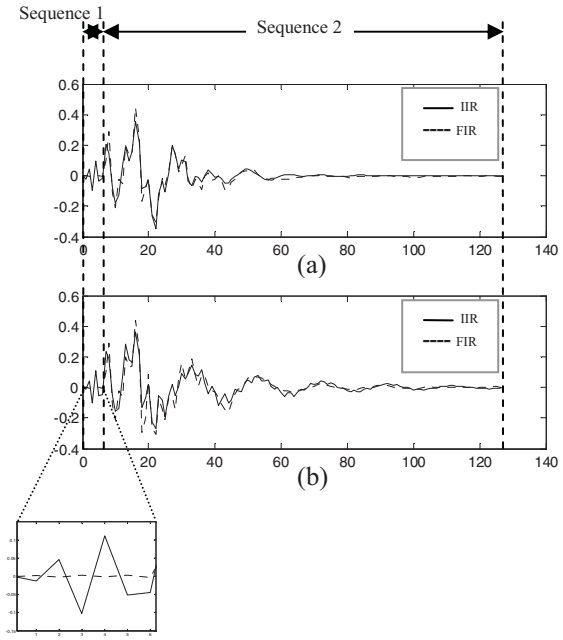

Figure 7. The time responses of two tenth-order IIR (solid) and 128-tap FIR (dashed) filters in Shuffler Filter 1. (a) The impulse response approximating $\frac{h_i + h_c}{2}$. (b) The impulse response approximating $\frac{h_i - h_c}{2}$.

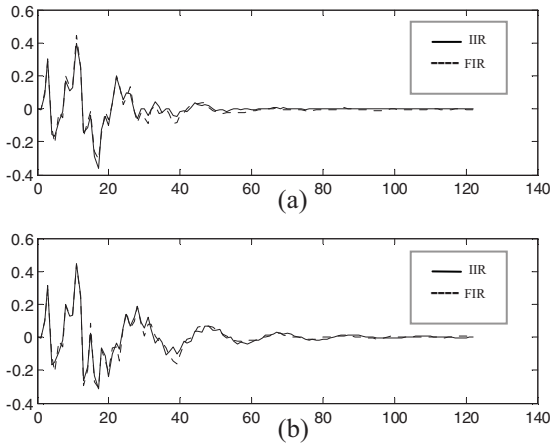

Figure 9. The time responses of two tenth-order IIR filters (solid) in Fig. 8 that approximate the FIR filters (dashed) without the initial time delays. (a) The response of Tenth-Order IIR Filter 1. (b) The response of Tenth-Order IIR Filter 2.

Fig. 7 shows impulse responses of the 128-tap FIR filter models and the corresponding tenth-order IIR filter for Shuffler Filter 1 in time domain. We could notice that during the period of the initial time delay, the IIR filters have difficulty in approaching the FIR versions. The oscillations will cause pre-echo phenomenon noticeable,

especially when playing percussion music. In order to prevent this undesired effect, the initial time delay, Sequence 1 in Fig. 7, is removed from the FIR filter beforehand, so the algorithm only targets the rest of the response, Sequence 2 in Fig. 7. Then, the time delay is compensated back by using a delay buffer after filtering as shown in Fig. 8.

Designing the IIR filters with truncation of the initial time delays, we observe that not merely are the pre-echo effects alleviated, but the approximation errors of the rest sequences are also smaller. This can be seen in Fig. 9 which shows the responses of the IIR and FIR filters with the delays removed.

5 EXPERIMENTAL AND COMPARISON RESULTS

We design tenth-order IIR filter models from HRTFs of 128-tap FIR filter models released by MIT media lab [10]. A variety of audio pieces including stringed, wind, percussion, and piano music were used in experimental evaluations which are performed in terms of MSE (Mean Square Error) between normalized FIR and IIR version in time domain. Experimental results for different types of filter structures, using shuffler filter structures or not and removing initial time delays or not, are depicted in Fig 10. While in Type 1 and 2, we design the tenth-order IIR filters approximating the differences and sums of the ipsilateral and contralateral terms, in Type 3 and 4, the desired tenth-order IIR filters approach the ipsilateral and contralateral terms. Because the algorithm in the least-squares sense can approximate the truncated ipsilateral and contralateral terms very well, the error result of Type 4 is smallest. Nevertheless, to determine the optimum design of architecture, the trade-off between computational cost and performance is considered for efficiency of hardware or firmware implementation. Though presenting the smallest approximation errors, Type 4 requires 14 filters as shown in Fig. 3 to implement a whole system. As a result, we choose the second best performance which requires only 7 tenth-order IIR filters.

Table I indicates some comparison results of virtual surround structures, such as the filter type used in the architecture, the size requirement of the coefficient memory of each filter, the computational cost of each filter per sample of the signal, special sound effects, and the type of surround experience. As mentioned before, there exists trade-off between computational complexity and spatial quality. Even though our design is not the most compact one, the MSE values in Fig. 10 suggest the responses of the proposed IIR filters are close to those of the ordinary FIR filters. Moreover, unlike the others, the proposed architecture can process both 5.1 and 7.1-channel audio signals.

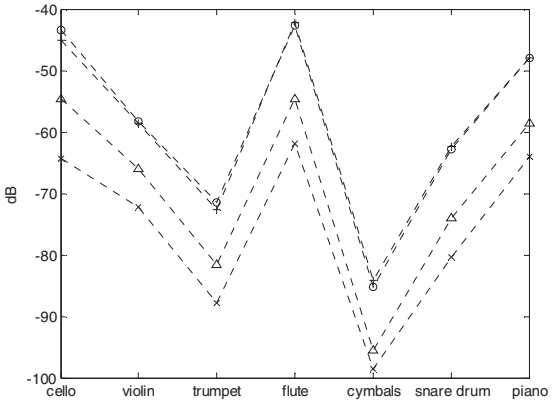

Figure 10. The MSE results in four kinds of situations.

6 CONCLUSIONS

In this paper, a dual-mode structure for virtual surround sound which is created on the basis of shuffler filters is proposed. This compact architecture is compatible with not merely discrete 5.1-channel but discrete 7.1-channel sources. The least-squares approximation is also suggested so as to enable low-order IIR filters for real-time implementation that memory use and computational

TABLE I
THE COMPARISON OF VIRTUAL SOUND STRUCTURES

	Filter Type	The Number of Coefficients	The Number of Multiplications	Sound Effect	Type of Surround Sound
Bai's [1]	128-tap FIR filter	128	128	None	5-channel
Fujinami's [2]	368-tap and 12-tap FIR filters	368 +12	368 +12	Reverberation	5-channel
You's [3]	44-tap(on average) FIR filter	44	44	None	5-channel
Kawano's [4]	Second-order IIR filter	5	5	None	5.1-channel
Proposed	Tenth-order IIR filter	21	21	None	5.1/7.1-channel

complexity can be considerably reduced. On account of the reduced memory space and processing time, the proposed system is suitable for low-cost design. In order to clarify the performance of the proposed algorithm, objective MSE measurements and subjective listening tests have been done.

7 REFERENCES

[1] Mingsian R. Bai and Geng-Yu Shih, "Upmixing and downmixing two-channel stereo audio for consumer electronics," *IEEE Trans. Consumer Electronics*, vol. 53, pp. 1011-1019, Aug. 2007.

[2] Y. Fujinami, "Improvement of sound image localization for headphone listening by wavelet analysis," *IEEE Trans. Consumer Electronics*, vol. 44, pp. 1183-1188, Aug. 1998.

[3] Shingchern D. You and Woei-Kae Chen, "Rendering five-channel audio on headsets," in *Proc. IEEE Intl. Symp. Consumer Electronics*, 2005, pp. 25-30.

[4] S. Kawano, M. Taira, M. Matsudaira, and Y. Abe, "Development of the virtual sound algorithm," *IEEE Trans. Consumer Electronics*, vol. 44, pp. 1189-1193, Aug. 1998.

[5] ITU-R BS.775-1, "Multi-channel stereophonic sound system with or without accompanying picture," International Telecommunications Union, Geneva, Switzerland, 1992–1994.

[6] Dolby Laboratories, Home Theater Speaker Guide, http://www.dolby.com/consumer/setup/speaker-setup-guide/index.html

[7] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, vol. 37, pp. 3–19, 1989.

[8] H. Brandenstein and R. Unbehauen, "Least-squares approximation of FIR by IIR digital filters," *IEEE Trans. Signal Processing*, vol. 46, no. 1, Jan. 1998.

[9] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 39-41, Feb. 1997.

[10] Bill Gardner and Keith Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," MIT Media Lab, May 1994.