PAPER

A Low-Complexity Down-Mixing Structure on Quadraphonic Headsets for Surround Audio

Tai-Ming CHANG[†a)], *Student Member*, Yi-Ming SHIU[†b)], *Nonmember*, *and* Pao-Chi CHANG[†c)], *Member*

SUMMARY This work presents a four-channel headset achieving a 5.1-channel-like hearing experience using a low-complexity head-related transfer function (HRTF) model and a simplified reverberator. The proposed down-mixing architecture enhances the sound localization capability of a headset using the HRTF and by simulating multiple sound reflections in a room using Moorer's reverberator. Since the HRTF has large memory and computation requirements, the common-acoustical-pole and zero (CAPZ) model can be used to reshape the lower-order HRTF model. From a power consumption viewpoint, the CAPZ model reduces computation complexity by approximately 40%. The subjective listening tests in this study shows that the proposed four-channel headset performs much better than stereo headphones. On the other hand, the four-channel headset that can be implemented by off-the-shelf components preserves the privacy with low cost.

key words: *surround audio, head-related transfer function, virtual loudspeaker, reverberation*

1. Introduction

With the rapid development of information technology and internet access in recent years, digital video and audio applications have become commonplace in human life. Traditional mono or stereo audio playback is often un-able to satisfy the increasing demands of human perception for multimedia. Placing multiple channels (such as 5.1 channels) at different positions can create different three-dimensional (3D) spatial effects.

The market is currently capable of providing high quality multichannel audio systems, but loudspeaker systems are relatively expensive. However, space limitations often mean it is not suitable to place a multichannel system in a crowded space, as this may interfere with neighboring people. A conventional earphone set provides privacy, but not surround sound. Therefore, it is desirable to produce a headset that provides privacy and the ability to experience the rich multichannel audio of DVD or Blu-ray multimedia at much lower cost than a complete loudspeaker system.

People can easily experience the sound source changes produced by sophisticated multichannel systems in theaters. The rich spatial information in these signals creates a virtual sound field for the listener. However, the same signals in a headphone, without any spatial cues may blur the sound location and yield unnatural perception. This phenomenon is called the "in head" effect. To overcome this problem, several studies have attempted to enhance the acoustic field to construct approximate out-of-head sound field [1]–[3].

The human auditory system (HAS) uses the physical distance of two ears and the difference of sound arriving time to localize sound sources. In this transmission path, the sound characteristic may be affected by obstructions and time difference to ears. Head-related Transfer Function (HRTF) is often used to synthesize spatial sound. The HRTF represents all the information of the sound transmission from a given source position to the ear drum. Those spatial cues of the Inter-aural Intensity Differences (IIDs) and Inter-aural Time Differences (ITDs) thoroughly describe the filtering effect of the head, torso, and pinna [4], [5]. Since the major part of HRTF that is included in the loudspeaker path will not present in the earphone path, adding HRTF in the headphone system can certainly enhance HAS perception.

The "virtual loudspeaker" is an interesting issue in headset systems. A prolific sound field typically places stereo or 5.1-channel loudspeakers in a space. The virtual loudspeaker maps the loudspeaker position in a real environment to headset devices. One approach uses the HRTF to filter the signals from the loudspeaker location in a degree according to the propagation of sound from the speaker to the listener's two ears [4], [6]. In general, the HRTF can be realized by a Finite Impulse Response (FIR) model with long coefficients to accurately present the spatial characteristic. To reduce the number of coefficients, an Infinite Impulse Response (IIR) filter has been used to modeling the HRTF [7], [8]. The IIR just needs fewer order of pole/zero that can efficiently generate the response of FIR. However, a different set of poles and zeros is still needed corresponding to each different sound direction. Thus, the common-acoustical-pole and zero (CAPZ) model has been proposed for modeling head-related transfer function [9]. This approach is based on the acoustic resonance of the pinna, and reconstructs HRTF coefficients using a set of common-pole and variable zeros for every sound direction. This approach greatly reduces storage requirement and computation capability.

The HRTF is evaluated from an anechoic chamber without considering the information of sound reflection from walls. In real world, it is necessary to consider multiple reflections of sound waves to simulate a complete acoustic environment [10], [11]. When the panning technique of

Manuscript received September 18, 2012.
Manuscript revised February 18, 2013.
[†]The authors are with the Department of Communication Engineering, National Central University, Zhongli, Taiwan.
a) E-mail: tmchang@vaplab.ce.ncu.edu.tw
b) E-mail: ymshiu@vaplab.ce.ncu.edu.tw
c) E-mail: pcchang@ce.ncu.edu.tw
DOI: 10.1587/transfun.E96.A.1526

virtual loudspeaker creates a room system in headphones, it requests spatial information of the sound travel path to simulate the room effect. Wang et al. simulated a stereo sound field enhancement system [12]. Their system can achieve approximately normal stereo sound propagation in a space. However, it still cannot simulate a home cinema, such as a 5.1 multichannel environment. Hence, the current study proposes a down-mixing 5.1-channel structure to a four-channel structure using CAPZ technique to reduce implementation cost and a simplified reverberator to simulate the sound reflection of actual sound field.

The remainder of the paper is organized as follows. Section 2 describes the proposed multichannel down-mixing structure with HRTF and reverberator. Section 3 presents the implementation results. Finally, Sect. 4 draws conclusions.

2. Proposed Quadraphonic Headset Scheme

Following the ITU multichannel configuration [13] in Fig. 1, three loudspeakers were placed in front of the listener and the other two at the back, where θ_1 represents the angle between the Center channel and the Right channel. Similarly, θ_2, θ_3, and θ_4 represent the angles from Center (C) channel to Rear-right (RR), to Rear-left (RL), and to Left (L) channels, respectively. Headset enhancement attempts to re-create this multichannel effect by simulating the real surround sound effect. Figure 2 shows the six-to-four down-mixing structure, where X_L, X_R, X_C, X_{LFE}, X_{RL}, and X_{RR} indicate the input channels of the Left, Right, Center, Low frequency-effect (LFE), Rear left and Rear right, respectively.

The proposed structure performs the down-mixing based on the principle of energy conservation. Assume the ear phone elements are identical. To avoid the overload the earphone elements, signal of LFE is added only to the rear two channels. The signal of center channel is attenuated by 3 dB and added to the left and right channels. The signal of the low frequency-effect channel is also reduced by 3 dB and added to rear-left and rear-right channels. The left, right, rear-left and rear-right signals are filtered by respective locations of HRTF to present the virtual loudspeaker locations and passed through the reverberator to produce artificial room reverberation. This model can be expressed as (1)

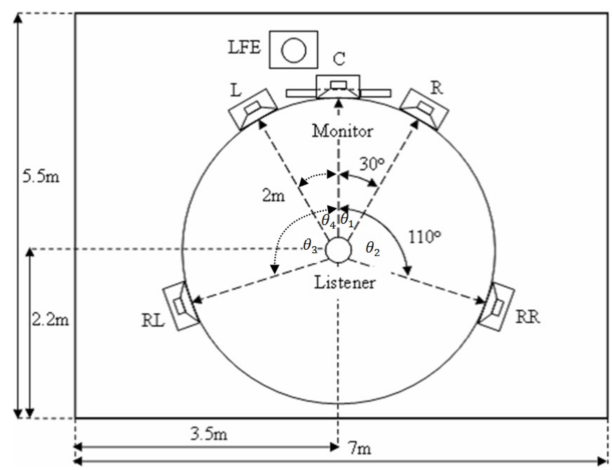

Fig. 1 Diagram of surround sound system.

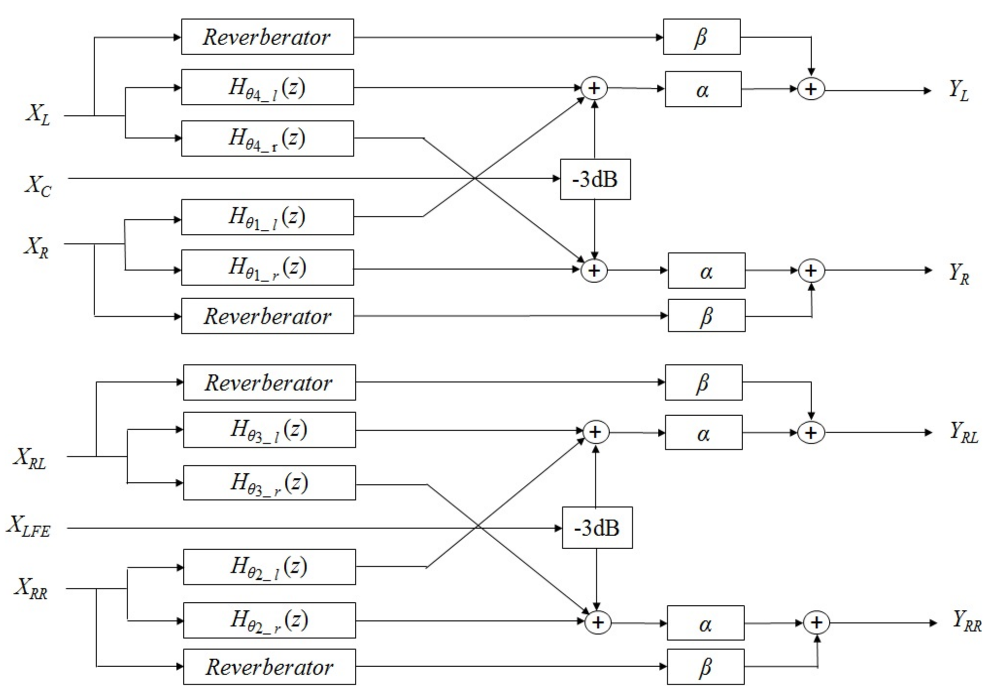

Fig. 2 Proposed multichannel down-mixing system.

$$\begin{cases} Y_L(z) = \alpha[X_L(z)H_{\theta_4\_l}(z) + X_R(z)H_{\theta_1\_l}(z) \\ \qquad +0.707X_c(z)] + \beta(X_L(z)Revb(z)) \\ Y_R(z) = \alpha[X_R(z)H_{\theta_1\_r}(z) + X_L(z)H_{\theta_4\_r}(z) \\ \qquad +0.707X_c(z)] + \beta(X_R(z)Revb(z)) \\ Y_{RL}(z) = \alpha[X_{RL}(z)H_{\theta_3\_l}(z) + X_{RR}(z) * H_{\theta_2\_l}(z) \\ \qquad +0.707X_{LFE}(z)] + \beta(X_{RL}(z)Revb(z)) \\ Y_{RR}(z) = \alpha[X_{RR}(z)H_{\theta_2\_r}(z) + X_{RL}(z)H_{\theta_3\_r}(z) \\ \qquad +0.707X_{LFE}(z)] + \beta(X_{RR}(z)Revb(z)) \end{cases} \tag{1}$$

where $Y_L(z)$, $Y_R(z)$, $Y_{RL}(z)$, $Y_{RR}(z)$ indicate the output channels of the Left, Right, Rear left and Rear right, respectively. The $H(z)$ represents the z-transfer function of HRTF, and suffix θ denotes the angle of virtual loudspeaker location. *Revb* is the room reverberator, and α, β represent the gains. Basically, α and β represent how much the HRTF effect and reverberation effect are included in the output, respectively. Proper values of these two parameters can be obtained from experiments.

An anechoic chamber is generally used as the optimal environment for measuring HRTF data because it absorbs the influence of reflected signals. The HRTF is typically implemented by a finite impulse response filter (FIR) that has long filter coefficients to simulate various source directions. This work uses a common-acoustic-pole and zero model that reduces the number of filter coefficients, and a simplified artificial reverberator that simulates the reverberation of a space. Two major components, the HTRFs and the reverberator, are discussed as follows.

2.1 Common-Acoustic-Pole and Zero Model of HTRFs

To reduce computational complexity, this study employs an infinite impulse response filter (IIR) with lower orders to approximate the original FIR. This study also adopts Haneda's method, called common-acoustic-pole and zero model (CAPZ) [9]. The CAPZ concept model states that different degrees of HRTF involve a common resonance structure due to the pinna structure. This resonance can be approximated by a set of common poles.

Assuming that θ is the degree of sound source location, the CAPZ modeled from HRTF can be expressed as

$$H_c(\theta, z) = \frac{B(\theta, z)}{Ac(z)} = \frac{\sum_{i=0}^{Q} b_i(\theta) z^{-i}}{1 - \sum_{i=1}^{P} a_{ci} z^{-i}} \tag{2}$$

where Q and P are the order of zeros and poles respectively, a_{ci} denotes the common pole coefficients, and the b_i is the zero coefficient of each angle θ. $H_c(\theta, z)$ represents the transfer function of CAPZ model in the z-transform domain. Taking the inverse z-transform of (2), the impulse response of the CAPZ model can be described in the time domain as

$$h_c(\theta, n) = \sum_{i=1}^{P} a_{ci} h_c(\theta, n-i) + \sum_{i=0}^{Q} b_i(\theta)\delta(n-i) \tag{3}$$

where n is the index in time and $\delta(n)$ is the impulse function. Hence, the output error between the original HRTF $h(\theta, n)$ and the CAPZ model $h_c(\theta, n)$ is given by

$$\begin{aligned} E(\theta, n) &= h(\theta, n) - h_c(\theta, n) \\ &= h(\theta, n) - \sum_{i=1}^{P} a_{ci} h_c(\theta, n-i) - \sum_{i=0}^{Q} b_i(\theta)\delta(n-i) \end{aligned} \tag{4}$$

However, it is difficult to calculate the coefficients of a_{ci} and b_i by minimizing the mean-square of the output error [14]. Hence, known information can be used to redefine the error function as

$$\begin{aligned} E_{eq}(\theta, n) &= h(\theta, n) \\ &- \sum_{i=1}^{P} a_{ci} h(\theta, n-i) - \sum_{i=0}^{Q} b_i(\theta)\delta(n-i) \end{aligned} \tag{5}$$

The cost function is defined as the square sum for time index n and the source direction index m of the error function is

$$J_{eq} = \sum_{m=0}^{M-1} \sum_{n=0}^{N+P} E_{eq}^2(\theta_m, n) \tag{6}$$

where M is number of HRTF and N is the length of the original impulse response. The minimized cost function, which uses the least-squares method, can then be represented in matrix form as (7)

$$x = (A^T A)^{-1} A^T h_a \tag{7}$$

where superscript T is the transpose, x is a vector that contains estimated common poles and zeros, h_a indicates the coefficient matrix for the HRTF in each direction, and A stands for least-squares error to minimize the calculation of the matrix. The detailed operations are described in [10].

2.2 The Room Effect Simulator

In addition to the sound waves produced by the source itself, the reflected sound field contains reflections of sound waves resulting from walls or obstacles [15]. It consists of three components: direct sound, early reflection, and late reverberation (Fig. 3). Those sound elements can enhance the listening experience.

The reflection of the wall enriches the sound, making it more vivid. The total reflections can be represented by a simple model, proposed by Moorer [16]. This model consists of six parallel comb filters representing the sound reflections from the wall and cascades these filters with an all-pass filter to simulate diffuse reverberation. Figure 4 shows

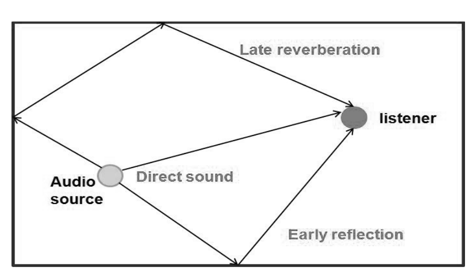

Fig. 3 Sound reflection path in a closed space.

the reverberator framework.

Moorer's reverberator consists of comb filters and an all-pass filter. A comb filter gives a simple delay and feedback with an exponential decay gain g. Thus, the comb filters represent the early reflection sound wave between walls. The function of comb filter is described as follows:

$$H_{comb}(z) = \frac{z^{-m}}{1 - gz^{-m}} \tag{8}$$

The output of comb filters is imported to an all-pass filter to enhance the echo density. The all-pass filter structure resembles a comb filter with a forward path around the delay. It has a configuration with the same pole and zero position in the pole-zero diagram, which results in a flat frequency response. The all-pass filter can be expressed as

$$H_{all}(z) = \frac{z^{-m} - g}{1 - gz^{-m}} \tag{9}$$

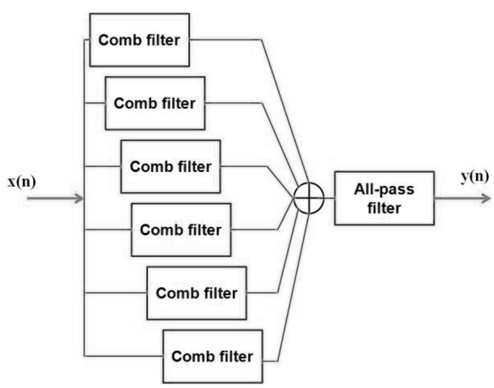

Fig. 4 Diagram of Moorer's reverberator.

Finally, the down-mixing structure was modified using the CAPZ technique, as Fig. 5 shows. The HRTF is replaced by the CAPZ model, which reconstructs the impulse response by sharing an all-pole filter with each all-zero filter. Therefore, (1) can be simplified as

$$\begin{cases} Y_L(z) = \alpha[(X_L(z)B_{\theta_4\_l}(z) + X_R(z)B_{\theta_1\_l}(z))\frac{1}{A(z)} \\ \qquad +0.707X_c(z)] + \beta(X_L(z)Revb(z)) \\ Y_R(z) = \alpha[(X_R(z)B_{\theta_1\_r}(z) + X_L(z)B_{\theta_4\_r}(z))\frac{1}{A(z)} \\ \qquad +0.707X_c(z)] + \beta(X_R(z)Revb(z)) \\ Y_{RL}(z) = \alpha[(X_{RL}B_{\theta_3\_l}(z) + X_{RR}B_{\theta_2\_l}(z))\frac{1}{A(z)} \\ \qquad +0.707X_{LFE}(z)] + \beta(X_{RL}(z)Revb(z)) \\ Y_{RR}(z) = \alpha[(X_{RR}(z)B_{\theta_2\_r}(z) + X_{RL}(z)B_{\theta_3\_r}(z))\frac{1}{A(z)} \\ \qquad +0.707X_{LFE}(z)] + \beta(X_{RR}(z)Revb(z)) \end{cases} \tag{10}$$

3. Simulaion and Experience Results

Figure 6 illustrates the implemented prototype of the four-channel headset system. Other than the computer used to generate audio signals, this system only requires two conventional two-channel sound cards and two stereo earphones. This system can provide excellent surround audio. All devices are inexpensive, off-the-shelf components.

3.1 Performance of Common-Acoustic-Pole and Zero Model

The orders of common poles and zeros in the CAPZ model can be determined by the cost function J_{out}, which represents the average error of impulse responses between CAPZ model and actual HRTF. The cost function is defined as

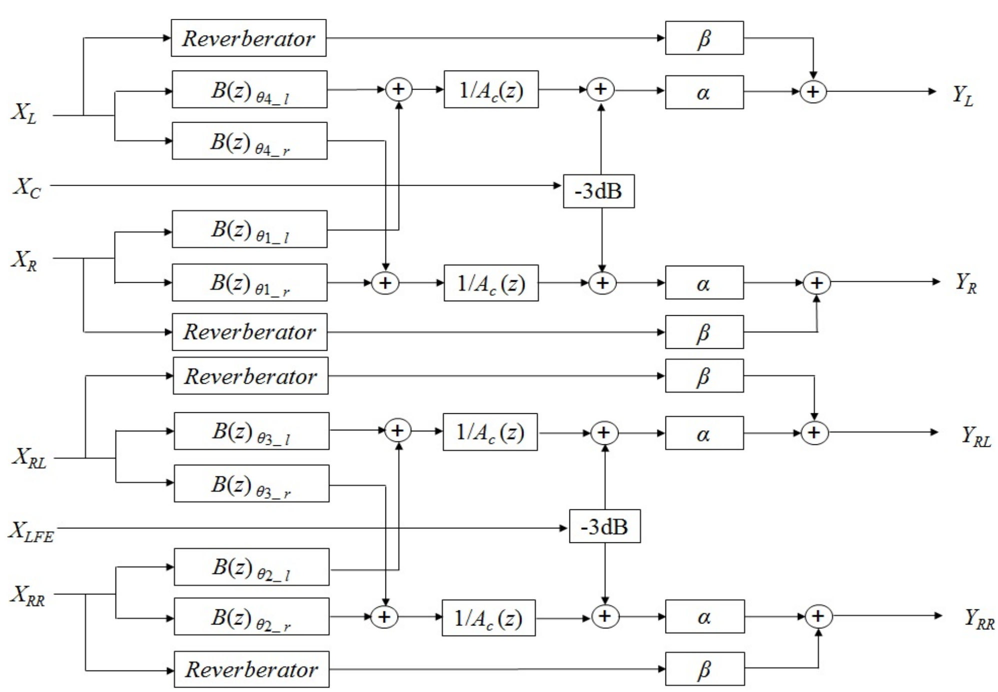

Fig. 5 Improved multi-channel down-mixing structure.

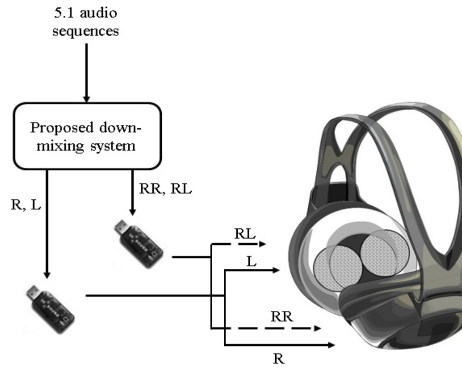

Fig. 6 Designed Four-channel headset.

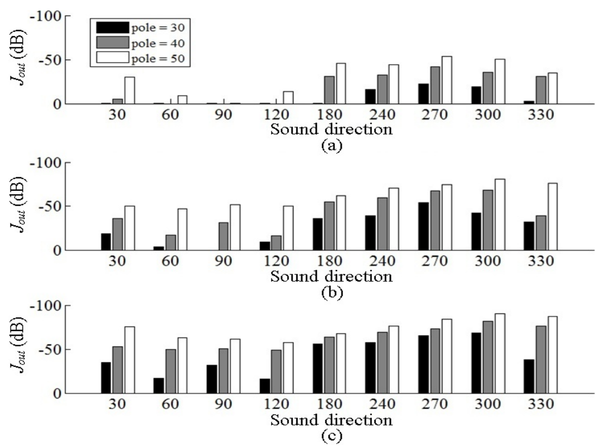

Fig. 7 The errors at the left ear side between common-acoustical-pole and zero model and actual HRTF at various degrees with (a) equal, (b) one and a half time, and (c) twice order of poles of zeros.

$$J_{out} = 10\log_{10}\left[\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{k=0}^{N} E^2(\theta_m, k)}{\sum_{k=0}^{N} H_{actual}^2(\theta_m, k)}\right] \tag{11}$$

where E is the output error as in (4) and H_{actual} presents the impulse response of actual HRTF.

The CAPZ coefficients were estimated from the original HRTF. The HRTF's length of the impulse responses were set to 200 and 36 source directions of HRTF at every 10° from 0° to 360° are chosen. Figure 7 shows the performance of the left ear with different order of pole and zero at various degrees. As expected, the higher order of poles it has, the less error it yields. In this experiment, equal order of poles and zeros has the worst performance in all directions. On the other hand, twice order of poles of zeros yields much better performance. More zeros could even do better. But the overall complexity might be over our budget. Figure 8 shows the cost function with different filter orders of poles, ranging from 20 to 60 with the interval 10, while the order of zeros is twice of the order of poles [9]. It can be observed that the value of cost function drops as the order increases, but becomes saturated when the order of the poles exceeds over 40. Considering the trade-off between complexity and efficiency, an order of 40 was chosen for implementation.

This study evaluates the advantage of the CAPZ model of HRTF by comparing it with the traditional all-zero model of HRTF. For a sound source located at 30 degrees to the listener, the CAPZ uses 40 common poles and 80 zeros to reconstruct the impulse response of the original HRTF to the right ear and the left year as shown in Fig. 9 and Fig. 10, re-

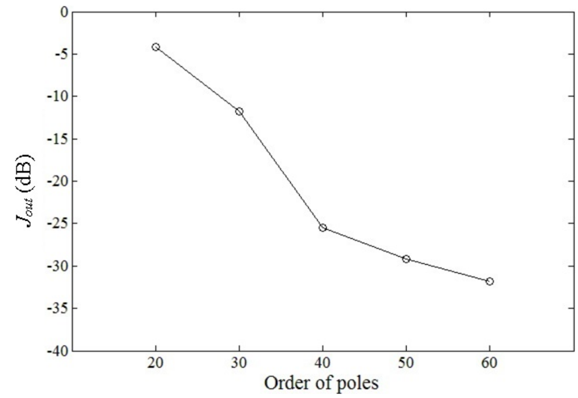

Fig. 8 Cost function vs. order of poles.

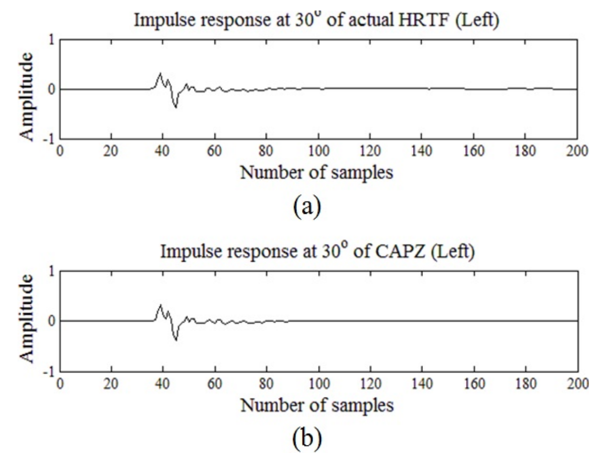

Fig. 9 Impulse response of left ear at 30°: (a) actual HRTF, (b) common-acoustic-pole and zero modeled HRTF.

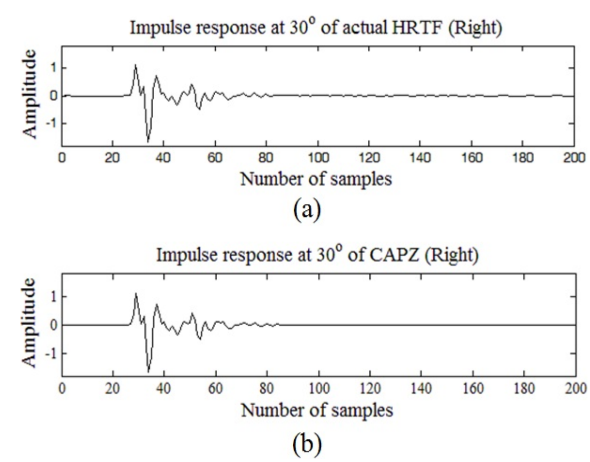

Fig. 10 Impulse response of right ear at 30°: (a) actual HRTF, (b) common-acoustic-pole and zero modeled HRTF.

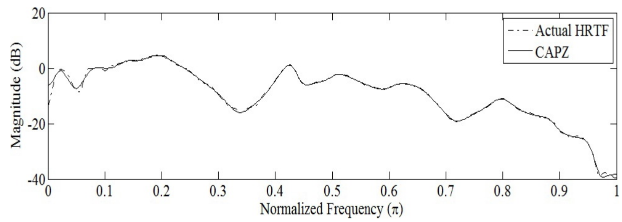

Fig. 11 Magnitude response of left ear at 30°.

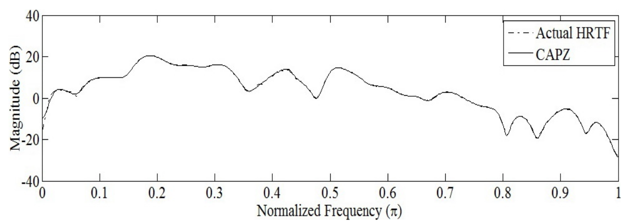

Fig. 12 Magnitude response of right ear at 30°.

Table 1 Implementation cost.

	HRTF	CAPZ
Number of multipliers	200	120
Number of adders	199	119
Number of states	199	79

Table 2 Requirement of coefficient quantity.

Model	Directions	Poles	Zeros	Total	Saving
HRTF	4	0	200	800	0%
CAPZ	4	40	80	360	55%

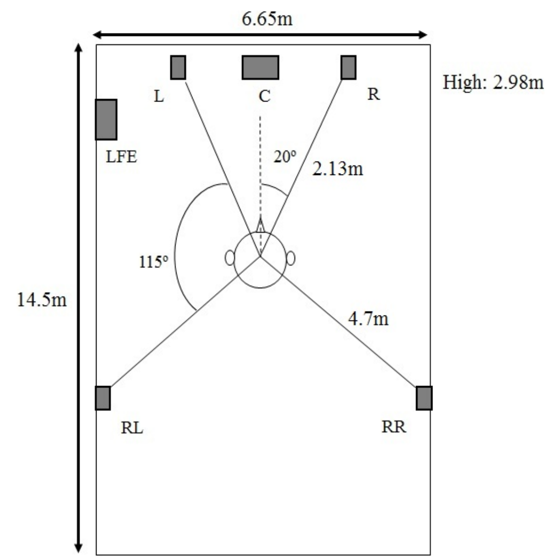

Fig. 13 A conference room with a multichannel system.

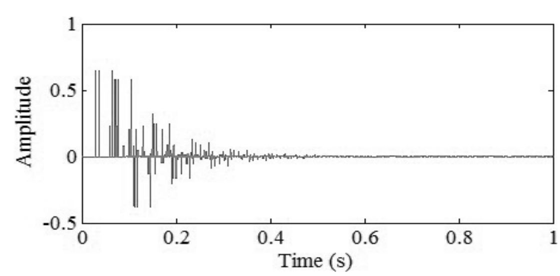

Fig. 14 The impulse response of simulation space.

spectively. The impulse responses of CAPZ for the left and right ears are very close to the traditional HRTF. Figure 11 and Fig. 12 show the magnitude responses of actual HRTF and CAPZ to the ears at 30 degrees. It can be observed that CAPZ shows excellent approximation with lower orders. From a power consumption viewpoint, CAPZ reduces the computation units of the multiplier, adder, and memory by 40%, 40%, and 60%, respectively. Table 1 lists the implementation cost with the direct-form II filter structure [17].

In the case of a four-channel system, it needs four different positions to represent the virtual loudspeakers. Table 2 shows that the proposed method decreases the CAPZ coefficient requirement by approximately 55% compared to the traditional HRTF.

3.2 Simulated Room Impulse Response by Moorer's Reverberator

We used a typical conference room measuring 14.5 m× 6.65 m × 2.98 m as the experimental space as shown in Fig. 13. The average absorption coefficient γ of the wall was set as 0.3. The reverberation time T is defined as the time required for the sound energy density to decay 60 db as show in (12) [18].

$$T = \frac{60V}{1.085c\hat{a}} \tag{12}$$

where V is the volume of room in cubic meters, c is the speed of sound, and $\hat{a}$ represents the metric absorption units which is defined as

$$\hat{a} = S[-2.3 * \log_{10}(1 - \gamma)] \tag{13}$$

where S is summation of the areas of absorbing surfaces. The reverberation time of the simulation space was approximately 0.41 s, and the early reflection delays from right speaker to the six walls and then to the listener are approximately 21.1 ms, 28.1 ms, 28.8 ms, 29.8 ms, 36.6 ms, and 62.8 ms, respectively. The impulse response of the simulation space can easily be calculated, as Fig. 14 shows.

3.3 Subjective Hearing Test

Table 3 shows the score rating ranges between −3 and 3 used to compare the two systems: our proposed system A and compared system B [19]. Listeners are given the instructions to rate the performance in the range from −3 to 3, where "0" means no difference, "1" means just noticeably better, "2" is obviously better, and "3" means absolutely better.

Twenty listeners were invited to participate in these subjective listening tests. Five 5.1-channel audio sequences are fragments of the DVD, each sampled at 44.1 kHz and

Table 3 Score rating.

Description	Score
A is much better than B	3
A is better than B	2
A is slightly better than B	1
A and B are both the same	0
A is slightly worse than B	−1
A is worse than B	−2
A is much worse than B	−3

Table 4 Description of audio sequences.

Name	Format	Audio sequences
Audio-1	5.1	Male speech with contemporary music
Audio-2	5.1	Female speech with contemporary music
Audio-3	5.1	Male speech with classical music
Audio-4	5.1	Male and Female speech with classical music
Audio-5	5.1	Male speech with computer music

with six seconds long, were tested. The characteristics of the test sequences are presented in Table 4. The first test was the comparison between a four-channel headset versus a stereo headphones, i.e., to test whether a four-channel headset is better than stereo headphones in various positions of sound source. The 5.1-channel audio sequence is first down-mixed to be either 4-channel or stereo signals. Then all channels are processed by HRTF with different sound directions at front (0°), right (90°), rear (180°), and left (270°). The test was blind in the sense that the same four-channel headset was used for both cases. Only the signals sent to the headset are different. In the stereo case, only one earphone element at one side receives signal. The five-to-two down-mixing structure of the standard ITU-R BS.775-1 was used to produce signals for stereo headset [20]. The second comparison was to determine whether a four-channel headset can rival the multichannel system in space hearing perception, i.e., the capability to locate the sound source. The same scoring system as in Table 3 is used for rating the performance. A conference room was used to furnish a 5.1-channel system that was described in Sect. 3.2. The speakers of Right, Rear right, Rear left and Left are placed at 20°, 135°, 225°, and 340°, as shown in Fig. 13. Next, our proposed system reconstructed the experiment environment in a four-channel headset to evaluate the sound effect. The gains of output signal are set to be α=0.7 and β=0.5. The results for comparisons with stereo headphones and 5.1-channel systems for five audio sequences are shown in Fig. 15 and Fig. 16, respectively.

In Fig. 15, the distributions of grades are majorly at the positive side. It shows that a four-channel headset clearly performed better sound localization than the stereo headset. Our enhancement system can clearly give the sound locations, especially the sounds coming from the front part versus the rear parts. However, the result of the comparison with a 5.1-channel system is mixed. Figure 16 shows that the average grades are on both sides around zero. In a well-controlled environment, such as a quiet room, most listeners prefer the loudspeaker system. In addition, the proposed system is able to reconstruct arbitrary space image, from

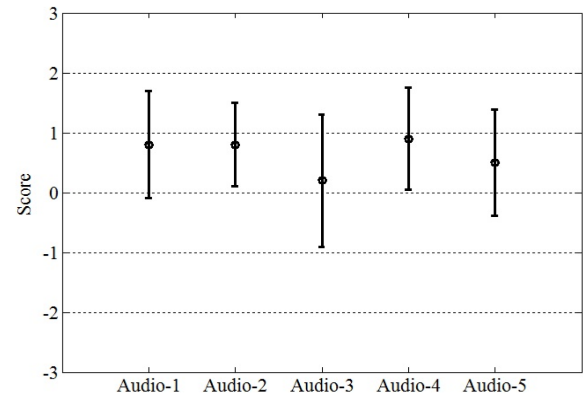

Fig. 15 Comparison with stereo headphones on the position difference.

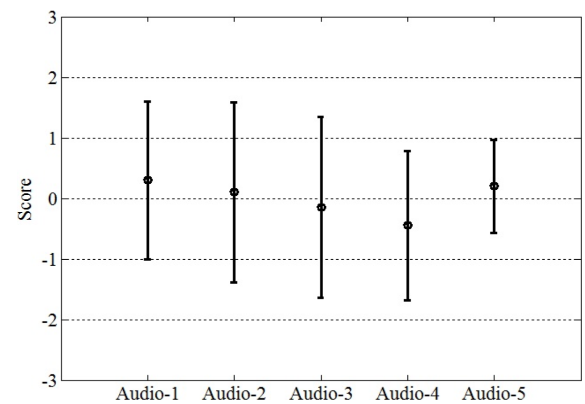

Fig. 16 Comparison with 5.1-channel on spatial sound field.

a tiny room to a concert hall, on the headset for personal preferences. From subjective listening tests, the proposed four-channel system provides superior sound localization to stereo headphones and comparable spatial sound field to a 5.1-channel.

4. Conclusion

This study proposes a four-channel headset capable of simulating multichannel surround audio while maintaining privacy. The proposed system combines low-complexity HRTF and artificial reverberation. The four-channel system uses the CAPZ model to reduce computation complexity and adopts an efficient artificial reverberator to simulate the reverberation of a real environment. Subjective listening tests show that the proposed four-channel headset performs much better than stereo headphones. Conversely, the four-channel headset preserves privacy at a low cost. Hence, our method that provides a multi-channel effect embedded in headset with lower power consumption and possesses great commercial potential.

References

[1] P. Rubak, “Headphone signal processing system for out-of-head localization,” The 90th Convention Audio Engineering Society, preprint 3063, Feb. 1991.

[2] Y. Fujinami, "Improvement of sound image localization for headphone listening by wavelet analysis," IEEE Trans. Consum. Electron., vol.44, no.3, pp.1183–1188, Aug. 1998.
[3] O. Kirkeby, P.A. Nelson, and H. Hamada, "The stereo dipole: A virtual source imaging system using two closely spaced loudspeakers," J. Audio Engineering Society, vol.46, no.5, pp.387–395, May 1998.
[4] C.I. Cheng and G.H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency and space," J. Audio Engineering Society, vol.49, no.4, pp.231–249, April 2001.
[5] S. Yano, H. Hokari, and S. Shimada, "A study on personal difference in the transfer functions of sound localization using stereo earphones," IEICE Trans. Fundamentals, vol.E83-A, no.5, pp.877–887, May 2000.
[6] M. Okamoto, I. Kinoshita, S. Aoki, and H. Matsui, "Sound image rendering system for headphones," IEEE Trans. Consum. Electron., vol.43, no.3, pp.689–693, 1997.
[7] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," IEEE Signal Process. Lett., vol.4, no.2, pp.39–41, Feb. 1997.
[8] M.A. Blommer and G.H. Wakefield, "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," IEEE Trans. Speech Audio Process., vol.5, no.3, pp.278–287, May 1997.
[9] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," IEEE Trans. Speech Audio Process., vol.7, no.2, pp.188–196, 1999.
[10] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," IEEE Trans. Speech Audio Process., vol.2, no.2, pp.320–328, April 1994.
[11] M.R. Schroeder, "Digital simulation of sound transmission in reverberant spaces," J. Acoust. Soc. Am., vol.47, no.2, pp.424–431, 1970.
[12] L. Wang, F. Yin, and Z. Chen, "An out of head sound field enhancement system for headphone," IEEE International Conference on Neural Networks and Signal Processing, Zhenjiang, China, June 2008.
[13] R. Irwan and R.M. Aarts, "Two-to-five channel sound processing," J. Audio Engineering Society, vol.50, no.11, pp.914–926, 2002.
[14] L. Ljung and T. Soderstrom, Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1983.
[15] U. Zolzer, Digital Audio Signal Processing, John Wiley & Sons, 1997.
[16] J.A. Moorer, "About this reverberation business," Computer Music Journal, vol.3, no.2, pp.13–28, 1979.
[17] A.V. Oppengeim, R.W. Schafer, and J.R. Buck, Discrete-Time Signal Processing, 3rd ed., Prentice Hall, 2010.
[18] L.L. Beranek, Acoustics, American Insttitute of Physics, 1986.
[19] International Telecommunications Union Radiocommunication Sector, "Method for the subjective assessment of intermediate sound quality," ITU-R BS.1534-1, Geneva, Switzerland, 2001.
[20] International Telecommunications Union Radiocommunication Sector, "Multi-channel stereophonic sound system with and without accompanying picture," ITU-R BS.775-1, Geneva, Switzerland, 1994.

Tai-Ming Chang received the B.S. degree in electrical engineering and the M.S. degree in computer and communication from the Southern Taiwan University and Shu Te University, Taiwan, in 2005 and 2007, respectively. He is currently pursuing the Ph.D. degree at the Video-Audio Processing Laboratory in communication engineering, National Central University, Taiwan. His research interests include speech/audio processing, audio compression, multi-channel audio signal processing.

Yi-Ming Shiu received the B.S. degree in communication engineering and the M.S. degree in communication engineering from the Yuan Ze University and National Central University, Taiwan, in 2009 and 2011, respectively. His research interests include audio coding, multimedia audio system and multi-channel audio headset, and so on.

Pao-Chi Chang received the B.S. and M.S. degrees from National Chiao Tung University, Taiwan, in 1977 and 1979, respectively, and the Ph.D. degree from Stanford University, California, 1986, all in electrical engineering. From 1986 to 1993, he was a research staff member of the department of communications at IBM T.J. Watson Research Center, Hawthorne, New York. At Watson, his work centered on high speed switching systems, efficient network design algorithms, and multimedia conferencing. In 1993, he joined the faculty of National Central University (NCU), Taiwan, where he is presently a Professor in the Department of Communication Engineering and the Department of Electrical Engineering, as well as the Advisor of the Video-Audio Processing Laboratory (VAPLab). He was a visiting professor at Stanford University in 2000 and 2004, respectively. His main research interests include speech/audio coding, video/image compression, scalable coding, error resilient coding, digital watermarking and data hiding, and multimedia delivery over packet and wireless networks.