PAN ARCHIVES OF ACOUSTICS
Vol. **38**, No. 1, pp. 55–62 (2013)

DOI: 10.2478/aoa-2013-0007

HRTF Adjustments with Audio Quality Assessments

Shu-Nung YAO[(1)], Li Jen CHEN[(2)]

[(1)] *School of Electronic, Electrical and Computer Engineering, University of Birmingham*
Edgbaston, Birmingham, B15 2TT, UK; e-mail: SXY043@bham.ac.uk

[(2)] *Acoustic and Camera Technology Department, Wistron Corporation*
21F, 88, Sec.1, Hsin Tai Wu Rd., Hsichih
New Taipei City 22181, Taiwan, R.O.C.; e-mail: arlen_chen@wistron.com

(*received July 15, 2012; accepted January 4, 2013*)

There are an increasing number of binaural systems embedded with head-related transfer functions (HRTFs), so listeners can experience virtual environments via conventional stereo loudspeakers or headphones. As HRTFs vary from person to person, it is difficult to select appropriated HRTFs from already existing databases for users. Once the HRTFs in a binaural audio device hardly match the real ones of the users, poor localization happens especially on the cone of confusion. The most accurate way to obtain personalized HRTFs might be doing practical measurements. It is, however, expensive and time consuming. Modifying non-individualized HRTFs may be an effort-saving way, though the modifications are always accompanied by undesired audio distortion. This paper proposes a flexible HRTF adjustment system for users to define their own HRTFs. Also, the system can keep sounds from suffering intolerable distortion based on an objective measurement tool for evaluating the quality of processed audio.

Keywords: HRTF, PEAQ, cone of confusion, headphones, surround.

1. Introduction

The duplex theory (Rayleigh, 1907) provides a model for a listener to distinguish the location of a sound source by feeling interaural time differences (ITDs) and interaural level differences (ILDs). However, the model has a problem specifying a unique three-dimensional position, because the positions on the cone of confusion (Cheng, Wakefield, 2001) share the same ITD and ILD cues. Head-related transfer functions (HRTFs) subsume not only ITD and ILD information but also the spectral characteristics, frequency magnitude, and phase responses for perception of a source in a three-dimensional space. In Fig. 1, where $s(t)$ is the sound coming from the loudspeaker M, $e_l(t)$ and $e_r(t)$ are the sounds reaching the listener's left and right eardrums, respectively, which can be represented as

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

and

$$e_r(t) = s(t) * h_{Mr}(t), \tag{2}$$

where "$*$" symbolizes the convolution operator; $h_{Ml}(t)$ and $h_{Mr}(t)$ denote a pair of head-related impulse responses (HRIRs), the inverse Fourier Transforms of HRTFs $H_{Ml}(f)$ and $H_{Mr}(f)$, containing the characteristic of sound affected by the head, pinna, shoulder, and torso.

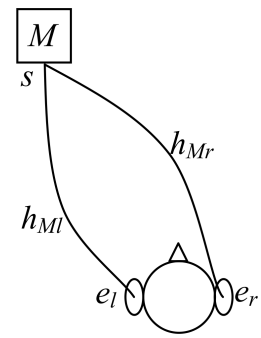

Fig. 1. A pair of HRTFs in the transaural listening: $s(t)$ – the original mono sound from the speaker M, $h(t)$ – HRIRs, $e(t)$ – signals at ears.

To date, some labs have measured several kinds of HRTFs by using artificial models (Gardner, Martin, 1994) or real human beings (Algazi *et al.*, 2001). However, when users use those databases, they might be unfamiliar with the HRTFs which do not actually belong to them. Then, two negative effects may hap-

pen, the so-called front-back confusions and up-down confusions. Figure 2 illustrates an example of front-back confusions. ITDs are important variables for human beings to locate the positions of sounds, but listeners perceive the same ITD if two sounds are virtually placed in the front hemisphere and symmetrically in the back hemisphere. In this case, listeners tend to confuse the locations of the front sound and the behind sound. On the other hand, Fig. 3 shows an illustration of up-down confusions. When human beings distinguish the elevations of sounds, the peaks and notches of HRTFs in frequency domain are quite important (Hebrank, Wright, 1974). If the spectrum characteristics of the non-individualized HRTFs stored in 3D audio systems rarely match those of the listener, the listener may have difficulties in locating the correct elevation of the virtual sound, thereby feeling it coming from a position slightly upper or lower than the actual one.

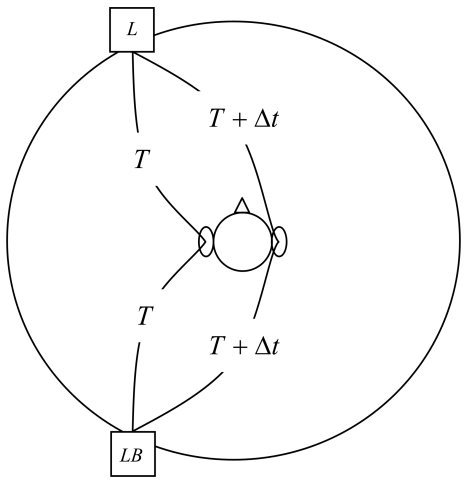

Fig. 2. Front-back confusions: T – the time delays from speakers to the left ear, $T+\Delta t$ – the time delays from speakers to the right ear.

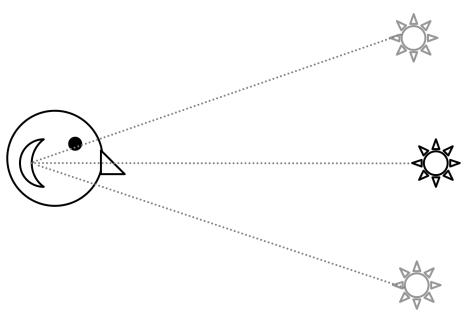

Fig. 3. Up-down confusions.

For medical electronics, Dobrucki *et al.* (2010) have introduced an efficient and reliable way to measure HRIRs in the visually impaired. Dobrucki and Plaskota (2007) have also indicated the measurement of head and ear geometry could provide sufficient information for modeling a numerical model. For consumer electronics, there have been several kinds of algorithms proposed to help listeners overcome the negative effects without any measurements. Tan and Gan (1998) have set up a 3D sound system by using a high-pass filter, several band-pass filters, and few variable gains, so users can boost or attenuate the frequency components in each filter band which is associated with front, back, up, and down perceptions. Gupta *et al.* (2002) obtained HRTF modification functions by using a solid sphere with cardboard flaps functioning as a human being's head with ears. Zhang *et al.* (1998) used weight functions to refine HRTFs, exaggerating the difference between front and back transfer functions. Park *et al.* (2005) noticed that the weight functions introduced by Zhang *et al.* (1998) overamplify spectral peaks and notches, and thus proposed more moderate functions. Even if those algorithms enhance the spatial effects, they may also cause distortions in terms of audio quality. Nevertheless, those studies rarely used any objective measurements to evaluate the distortions.

In this paper, we propose an HRTF adjustment system equipped with perceptual evaluation of audio quality (PEAQ), the International Telecommunications Union (ITU) standard for audio quality assessment (ITU-R BS.1387, 1994), so that objective measurements of perceived audio quality can be assessed. Moreover, the proposed system is composed of parametric filters which provide flexibility in the adjustment process.

2. Spatial effect enhancement

Based on several psychoacoustic studies (Hebrank, Wright, 1974; Iida *et al.*, 2007) and the frequency responses of HRTFs (Gardner, Martin, 1994; Algazi *et al.*, 2001), we have observed that the spectrum components in some frequency bands are closely associated with the subjective impression of direction.

Hebrank and Wright (1974) summarized the following results. Firstly, a sound passing by a 1-octave notch filter with the center frequency located at 7.5 kHz and a high pass filter with the cut-off frequency from 13 kHz to 14 kHz is perceived as a source located directly ahead. Second, a sound filtered by a 1/4-octave peak filter with an 8 kHz center frequency is perceived as a source located straight above. Finally, a peak filter at 11 kHz makes a sound perceived as a source located directly behind.

Iida *et al.* (2007) found that on the median plane, the peak of an HRTF always happens at about 4 kHz. Therefore, they suggested the peak is the reference information for human beings to analyze other peaks and notches. Moreover, they showed that two notch filters located at 9 and 16 kHz make a source appear as to be behind.

As a result, the adjustment filters for non-individualized HRTFs are designed according to the information in Table 1 which presents the summary of using the special frequency bands related to the subjective impression of direction. Figures 4a, b, and c indicate the magnitude frequency characteristics of the filter structures for making a sound coming from ahead, above, and behind, respectively.

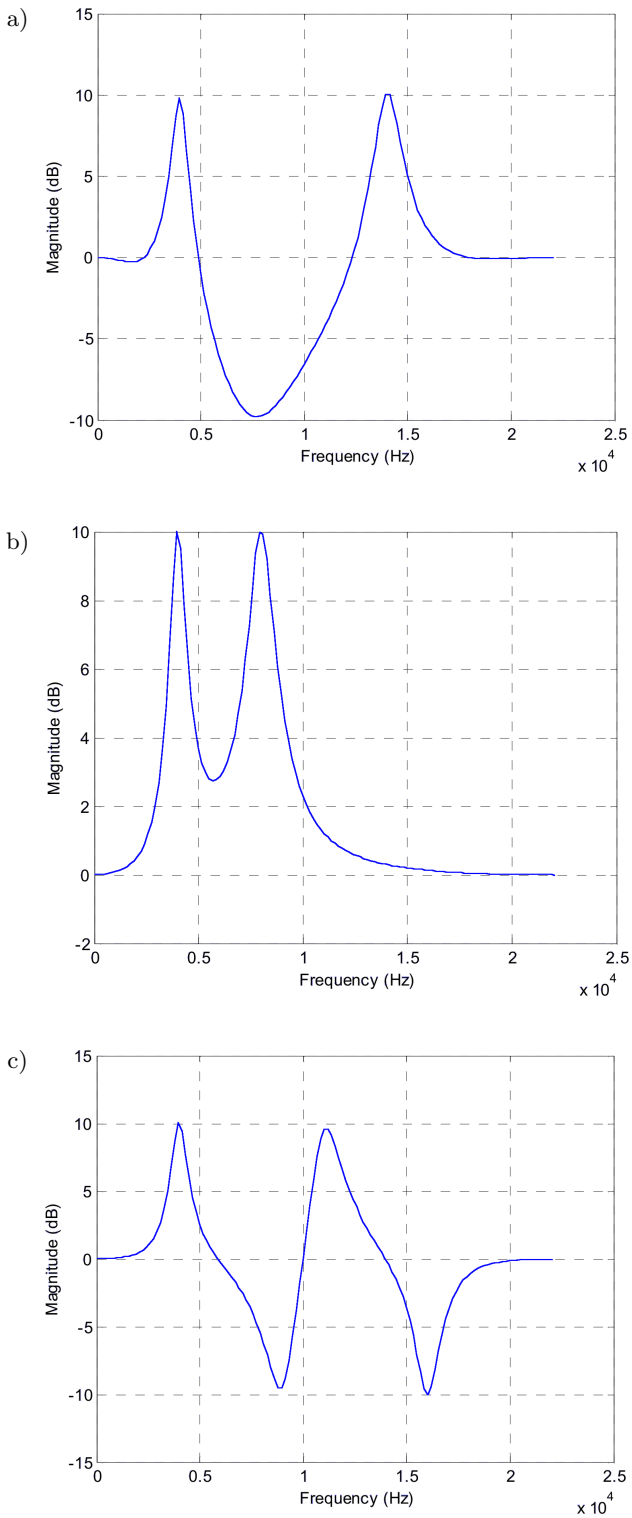

Fig. 4. The magnitude frequency characteristics of the filter structures: a) ahead, b) overhead, and c) behind.

Table 1. The characteristics of filters used in the proposed system.

"Frontness"		
Filter Type	Center Frequency	Band Width
Peak	4 kHz	1/4 octave
Notch	7.5 kHz	1 octave
Peak	14 kHz	1/4 octave
"Aboveness"		
Filter Type	Center Frequency	Band Width
Peak	4 kHz	1/4 octave
Peak	8 kHz	1/4 octave
"Behindness"		
Filter Type	Center Frequency	Band Width
Peak	4 kHz	1/4 octave
Notch	9 kHz	1/4 octave
Peak	11 kHz	1/4 octave
Notch	16 kHz	1/4 octave

The filter structures are composed of second-order parametric equalizers. The detailed design procedure is described by Zhang *et al.* (2010), and thus only the results are given in this paper. Zhang *et al.* (2010) designed three user-defined variables associated with digital audio equalization for octave bands. The first variable is used for determining the radian center frequency ω_c. The second variable is for the filter gain, G. The third variable Q is used to design the quality factor. Finally, the three parameters form the biquadratic filter transfer function shown as

$$H(z) = \Big[(1+M_1) + (M_2 - M_3 - 1)z^{-1} + (M_3 - M_1)z^{-2}\Big] \Big/ \Big[1 + (M_2 - M_3 - 1)z^{-1} + M_3 z^{-2}\Big]. \tag{3}$$

When we design a peak filter, let

$$M_1 = \frac{\dfrac{(G-1)k}{Q}}{1 + \dfrac{k}{Q} + k^2}, \quad M_2 = \frac{4k^2}{1 + \dfrac{k}{Q} + k^2}, \quad M_3 = \frac{1 - \dfrac{k}{Q} + k^2}{1 + \dfrac{k}{Q} + k^2}, \tag{4}$$

where $k = \tan(\omega_c/2)$.

When designing a notch filter, we choose

$$M_1 = \frac{-\dfrac{(G-1)k}{Q}}{1+\dfrac{kG}{Q}+k^2},$$
$$M_2 = \frac{4k^2}{1+\dfrac{kG}{Q}+k^2}, \qquad (5)$$
$$M_3 = \frac{1-\dfrac{kG}{Q}+k^2}{1+\dfrac{kG}{Q}+k^2},$$

where $k = \tan(\omega_c/2)$.

The filter structures are realized by cascading such peak or notch filters. Because of using parametric filters, the magnitude and bandwidth of each peak or notch can be adjusted flexibly. In Fig. 4, we arbitrary tune the magnitude to 10 dB for peaks and −10 dB for notches and the bandwidth, if it cannot be clearly found in references, to 1/4 octave.

An example of enhancing non-individualized HRTFs from three directions on the median plane is shown in Fig. 5. In this example, we assume that the listener's head is perfectly symmetrical, so the left and right ear impulse responses are identical (Gardner, Martin, 1994). The non-individualized HRTFs corresponding to a point ahead, overhead, and behind are respectively filtered by the three filter structures whose frequency responses are shown in Figs. 4a, b, and c. Figure 5 indicates that the effects of HRTFs are exaggerated. That is, by using the proposed adjustment system, the peak values are modified to be higher, while the notch values are lower.

3. Audio quality assessment

It is generally accepted that either overamplifying peaks or attenuating notches causes sound distortions (Zhang *et al.*, 1998; Park *et al.*, 2005). Therefore, a computer-based objective algorithm, PEAQ, is introduced to analyze the relationship between sound distortions and HRTF enhancements. PEAQ was developed for objective measurement of the perceived audio quality, grading the quality of the audio signals. Unlike traditional objective measurement methods, such as signal-to-noise-ratio (SNR) or total-harmonic-

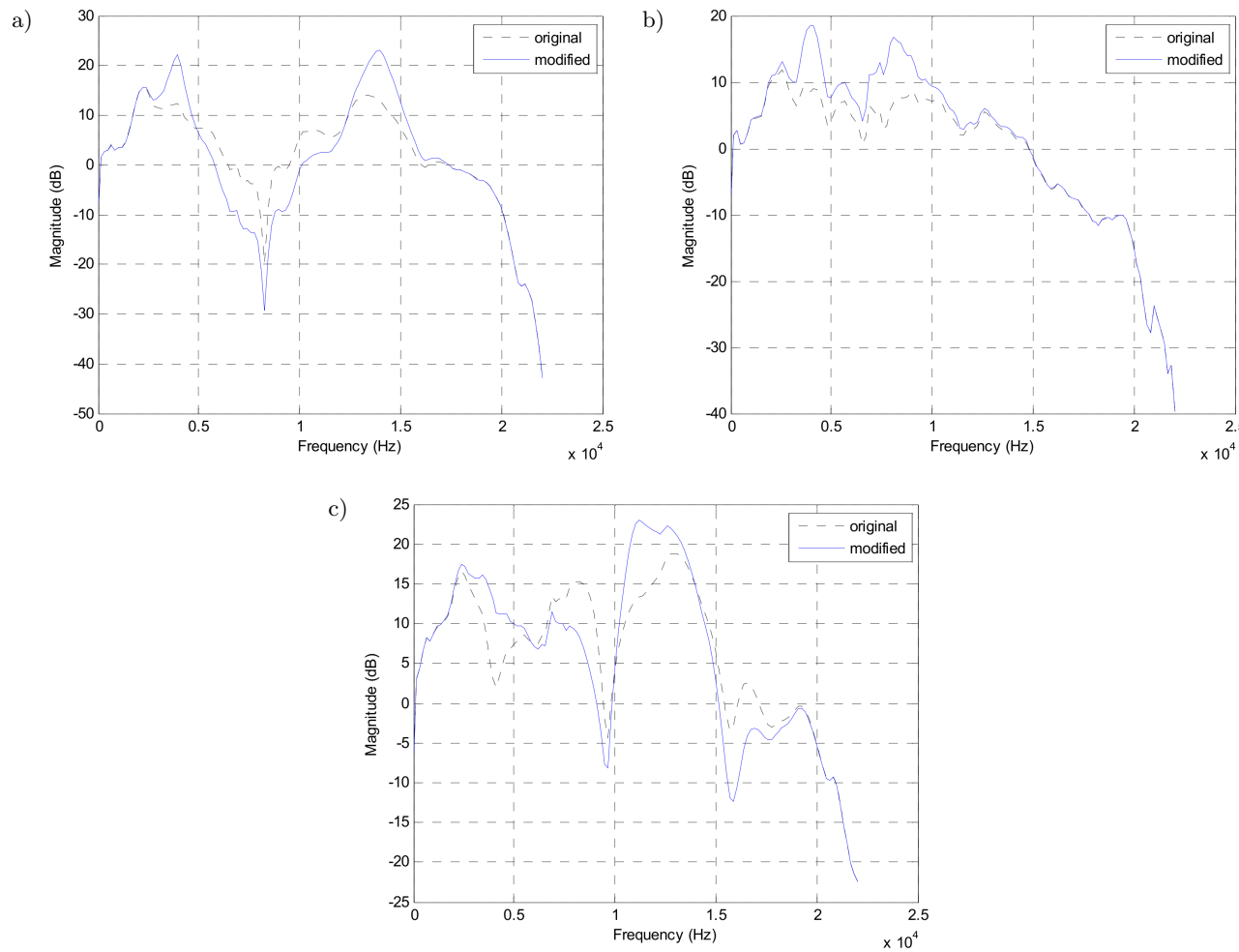

Fig. 5. An example of HRTF adjustment: a) ahead, b) overhead, and c) behind.

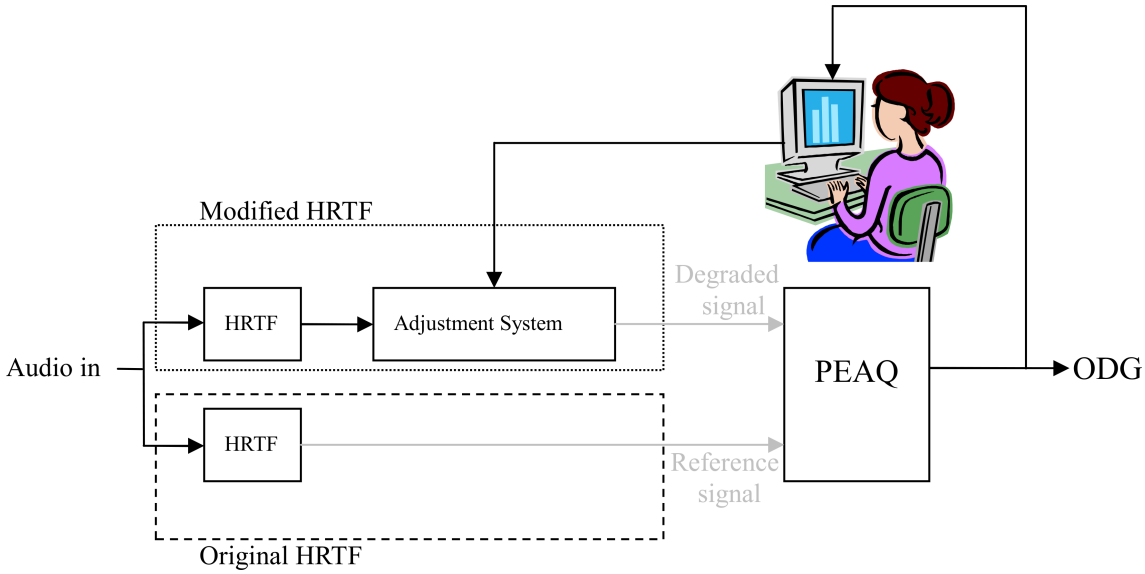

Fig. 6. The measurement scheme by PEAQ.

distortion (THD), PEAQ is designed based on the training of a neural network.

The measurement scheme is shown in Fig. 6. Users can determine the level of enhancement by adjusting the parameters of parametric filters after selecting the non-individualized HRTFs. Then the audio filtered by the modified HRTF will be compared with that filtered by the original one. Finally, overall difference grade (ODG) is generated. Normally, ODG is a value in the range of −4 to 0. As shown in Fig. 7, the more negative the score, the more perceptible the audio distortion. The ODG values provide the information about the tolerance of audio distortion, so HRTFs can be reasonably adjusted by users.

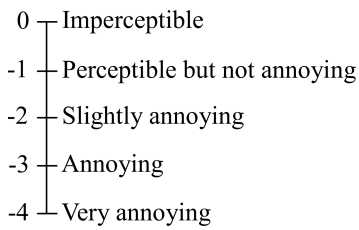

Fig. 7. Five-grade impairment scale.

4. Applications

Downmixing is an audio technique used to convert multi-channel surround to stereo format. The performance of downmix equations given by ITU recommendation (1994) had been well discussed (Kin, Plaskota, 2011). In addition to using downmix parameters, Hen *et al.* (2008) have exploited HRTFs to reproduce 5-channel audio over headphones. Recently, the conversion from 7-channel to 2-channel format has been proposed (Yao *et al.*, 2011). The main idea is to extend the situation from Fig. 1 to Fig. 8, so $e_l(t)$ and $e_r(t)$ will become (6) and (7), where $h_{Xy}(t)$ is the HRIR between the loudspeaker X (C, L, R, LS, RS, LB, or RB) and the listener's ear y (l or r), and $s_X(t)$ denotes the original audio signal produced by the loudspeaker X:

$$\begin{aligned} e_l(t) = {} & s_C(t)*h_{Cl}(t) + s_L(t)*h_{Ll}(t) + s_R(t)*h_{Rl}(t) \\ & + s_{LS}(t)*h_{LSl}(t) + s_{RS}(t)*h_{RSl}(t) \\ & + s_{LB}(t)*h_{LBl}(t) + s_{RB}(t)*h_{RBl}(t); \end{aligned} \tag{6}$$

$$\begin{aligned} e_r(t) = {} & s_C(t)*h_{Cr}(t) + s_L(t)*h_{Lr}(t) + s_R(t)*h_{Rr}(t) \\ & + s_{LS}(t)*h_{LSr}(t) + s_{RS}(t)*h_{RSr}(t) \\ & + s_{LB}(t)*h_{LBr}(t) + s_{RB}(t)*h_{RBr}(t). \end{aligned} \tag{7}$$

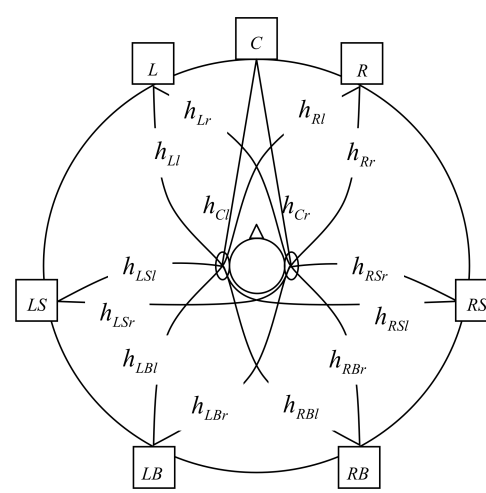

Fig. 8. A 7-channel surround system, h – HRIRs.

Since the conversion algorithm is based on HRIRs, front-back confusions are expected to happen. To be more specific, the loudspeaker L and loudspeaker LB

are virtually placed in the front hemisphere and symmetrically in the back hemisphere, respectively. The similar situation can be found by looking into the positions of the loudspeaker R and loudspeaker RB. As a result, the proposed HRTF adjustment technique is applied to improve the perceived localization of these four loudspeakers. The experimental results are described in the following section.

5. Experiments and results

During the experiments, subjective listening tests are carried out to evaluate spatial effects, while PEAQ is used to assess audio quality. The system is currently being run on a PC with the operating system Windows 7. The functions of HRTF adjustments and audio quality assessments were technically implemented in Matlab programming. The sounds are played via headphones. We compare the experimental results by tuning the magnitude of each peak and notch to three different settings, ±5 dB, ±10 dB, and then ±15 dB.

For subjective listening tests, 15 untrained subjects, 8 males and 7 females, are involved. They are asked to locate white noise filtered through the original HRTFs and the modified ones. Each piece of white noise comes from any of the four positions corresponding to the loudspeaker R, the loudspeaker RB, the loudspeaker L, and the loudspeaker LB in Fig. 8. The purpose of the subjective tests is to determine the best sound localization of the four settings presented as follows:

Test A – Original HRTFs measured by Gardner and Martin (1994);

Test B – Modified HRTFs by adjusting the peak filter values to 5 dB and the notch filter values to −5 dB;

Test C – Modified HRTFs by adjusting the peak filter values to 10 dB and the notch filter values to −10 dB;

Test D – Modified HRTFs by adjusting the peak filter values to 15 dB and the notch filter values to −15 dB.

There are 24 stimuli in each test. A stimulus is composed of a 1-second pause followed by a 2-second burst of white noise coming from different positions in random order. The same directional white noise was presented 6 times per test. The total duration of each test is 72 seconds. It takes about 5 minutes for each subject to complete the four tests. When a subject correctly recognizes the direction of the sound source, a point is accumulated in this current test. The average scores scaled from 0% to 100% are shown in Fig. 9. The means and standard deviations may appear in connection with the hypothesis that the more exaggeratedly we reshape HRTFs, the more easily listeners can distinguish the locations.

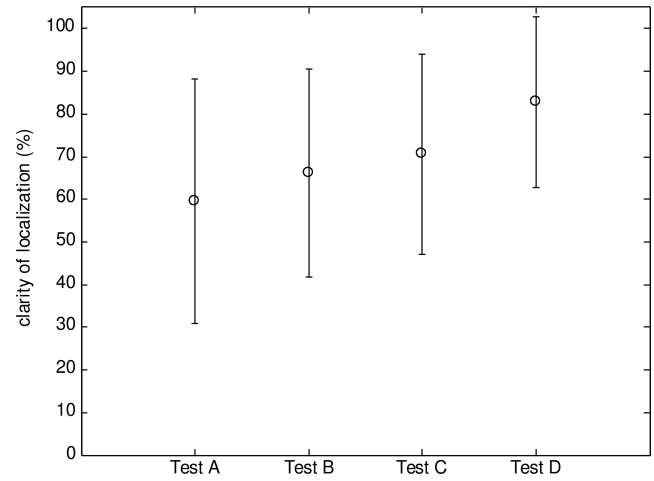

Fig. 9. Perceived clarity of sound locations. The circles are the mean values and the vertical lines symbolize the standard deviation values.

For objective measurement of perceived audio quality, the input signals are several types of audio pieces including string, wind, percussion, and piano music. As shown in Fig. 6, input audio filtered by an original HRTF functions as a reference referring to undistorted signal, while input audio filtered by a modified HRTF is the distorted signal under evaluation. After processing by PEAQ, we summarize the resulting ODGs in Table 2. Through looking into the values of ODGs one can see, as we expect, that the larger the filter gain, the worse the grade.

Table 2. The resulting ODGs by adjusting the magnitude of parametric filters.

Filter setting	Cello	Flute	Cymbals	Piano
±5 dB	−0.522	−0.357	−0.909	−0.288
±10 dB	−1.502	−1.181	−1.142	−1.163
±15 dB	−2.151	−1.983	−2.152	−2.511

6. Discussion

The proposed filter structures are developed according to spectral cues on the median plane (Hebrank, Wright, 1974; Iida *et al.*, 2007), but we find the proposed system still work pretty well when it is applied to the virtual 7-channel surround. This is because positions of loudspeakers R, RB, L, and LB are not far from the median plane, a reason which can be verified by comparing the HRTF characteristics at the positions R, RB, L, and LB and those on the median plane. Figure 10 illustrates an instance of the comparison. Through looking into the magnitude of $H_{Rl}(f)$, $H_{Rr}(f)$, $H_{Cl}(f)$, and $H_{Cr}(f)$ one can see that there is not much difference among the HRTFs.

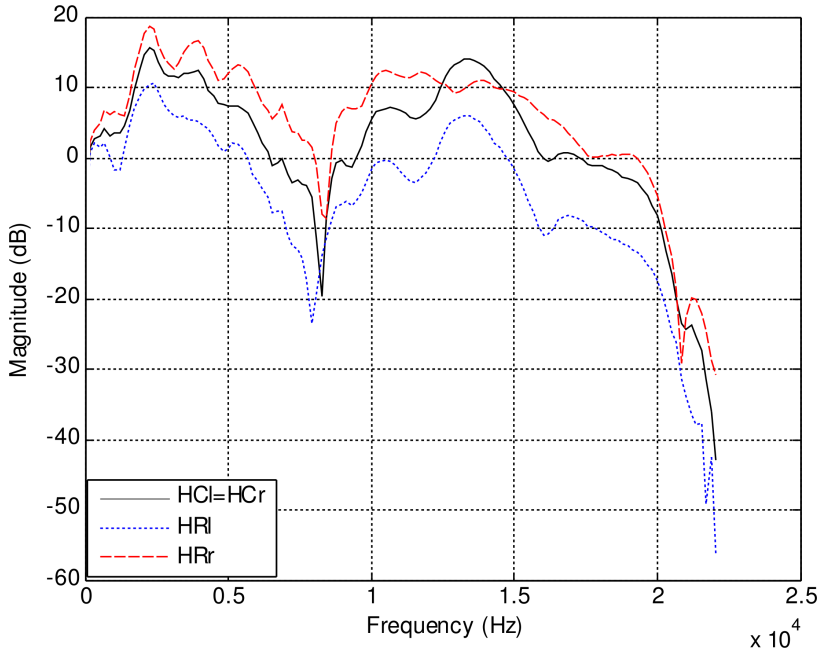

Fig. 10. The HRTFs with elevation 0° and azimuth 30° (the position of the loudspeaker R) and the HRTF with elevation 0° and azimuth 0° (the position of the loudspeaker C). Detailed description of symbols in text.

Broadly speaking, the experimental results indicate a negative correlation coefficient between the spectral difference in HRTFs and the grade of audio quality, while the correlation coefficient between the spectral difference in HRTFs and the spatial effect of sound is positive. When reinforcing the localization performance, one should be careful about audio distortion. Through looking into the resulting ODGs in Table 2 together with the physical meanings of ODG scale in Fig. 7 it can be seen that the treated audio becomes slightly annoying when HRTF magnitude variations reach ±15 dB. On the other hand, the ODGs are in ±5 dB setting range between 0 and −1 which means the treatment is perceptible but not annoying.

Although there is a negative relationship between audio quality and localization performance, some of the subjects got a higher score in Test B than in Test D. The possible explanation is that the filter setting used in Test B provides reasonable spectral cues which coincidentally match their real HRTFs, so they are similar to the modified ones. That the unchanged HRTFs match the listener's HRTFs could also happen, but rarely. This can be verified by the high standard deviation in Test A in Fig. 9.

In order to generate the precise spectral cues, the filter setting for the loudspeaker L and LB may be different from that for the loudspeaker R and RB. This can be illustrated by the following example. Subject No. 7 has no difficulty in identifying the source positions of her left hand side during Test B, as well as the positions of her right hand during Test C, while the sounds from the right hand side in Test B and the left hand side in Test C lead to poor localization. In this case, we use ±5 dB setting to refine the HRTFs, $H_{Ll}(f)$, $H_{Lr}(f)$, $H_{LBl}(f)$, and $H_{LBr}(f)$, and use ±10 dB setting to refine $H_{Rl}(f)$, $H_{Rr}(f)$, $H_{RBl}(f)$, and $H_{RBr}(f)$, a new asymmetric setting which is used for further examination. Upon the extra informal listening tests, the results are quite encouraging, showing better localization performance. Therefore, we hypothesize that asymmetric adjustments sometimes are needed because of the distinct physical characteristics from an individual's left ear to right ear.

7. Conclusion

In this paper, we design a system for HRTF customization with an audio quality assessment technique. The HRTF adjustment system is implemented by few parametric filters, so users can flexibly adjust the bandwidth and the magnitude of each filter. Moreover, the computer-based objective algorithm, PEAQ, provides the results of objective measurement for users to evaluate audio quality. As a result, users can clearly assess the trade-off between sound distortion and localization performance. Through subjective localization listening experiments and objective audio quality measurements, the proposed system can improve the sound spatialization in virtual environments without suffering too much annoying sound distortion.

There are some areas of future work concerning the externalization of sounds. The room model chosen for the system and the effects caused by the reflections

and reverberation on sound distortion and localization performance will be investigated. For headphone-based spatial sound, the sense making the sound outside of the head will be of great interest.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. The authors would also like to thank Tim Collins and Peter Jančovič for their guidance and helpful advice in this study.

References

1. Algazi V. R., Duda R. O., Thompson D. M., Avendano C. (2001), *The CIPIC HRTF database*, Proceedings of IEEE workshop Applicat. Signal Process. Audio Acoust., 99–102.
2. Cheng C. I., Wakefield G. H. (2001), *Introduction to head-related transfer functions (HRTFs): representations of HRTFs in time, frequency, and space*, J. Audio Eng. Soc., **49**, 4, 231–249.
3. Dobrucki A. B., Plaskota P. (2007), *Computational modelling of head-related transfer function*, Archives of Acoustics, **32**, 3, 659–682.
4. Dobrucki A. B., Plaskota P., Pruchnicki P., Pec M., Bujacz M., Strumillo P. (2010), *Measurement system for personalized head-related transfer functions and its verification by virtual source localization trials with visually impaired and sighted individuals*, J. Audio Eng. Soc., **58**, 9, 724–738.
5. Gardner B., Martin K. (1994), *HRTF Measurements of a KEMAR Dummy-Head Microphone*, MIT Media Lab.
6. Gupta N., Barreto A., Ordonez C. (2002), *Spectral modification of head-related transfer functions for improved virtual sound spatialization*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1953–1956.
7. Hebrank J., Wright D. (1974), *Spectral cues used in the localization of sound sources on the median plane*, Journal of the Acoustical Society of America, **56**, 1829–1834.
8. Hen P., Kin M. J., Plaskota P. (2008), *Conversion of stereo recording to 5.1 format using head-related transfer functions*, Archives of Acoustics, **33**, 1, 7–10.
9. Iida K., Itoh M., Itagaki A., Morimoto M. (2007), *Median plane localization using a parametric model of the head-related transfer function based on spectral cues*, Applied Acoustics, **68**, 8, 835–850.
10. ITU-R BS.775-1 (1994), *Multichannel stereophonic sound system with and without accompanying picture*, ITU, Geneva.
11. ITU-R Rec. BS.1387 (1998), *Method for objective measurement of perceived audio quality*, ITU, Geneva.
12. Kin M. J., Plaskota P. (2011), *Comparison of sound attributes of multichannel and mixed-down stereo recordings*, Archives of Acoustics, **36**, 2, 333–345.
13. Park M.-H., Choi S.-I., Kim S.-H., Bae K.-S. (2005), *Improvement of front-back sound localization characteristics in headphone-based 3D sound generation*, Proceedings of IEEE International Conference on Advanced Communication Technology, 273–276.
14. Rayleigh L. (1907), *On our perception of sound direction*, Philosoph. Mag., **13**.
15. Tan C.-J., Gan W.-S. (1998), *User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems*, Electronics Letters, **34**, 25, 2387–2389.
16. Yao S. N., Collins T., Jancovic P. (2011), *A dual-mode architecture for headphones delivering surround sound: low-order IIR filter models approach*, Proceedings of IEEE International Symposium on Consumer Electronics, 62–66.
17. Zhang M., Tan K.-C., Er M. H. (1998), *A refined algorithm of 3-D sound synthesis*, Proceedings of IEEE International Conference on Signal Processing Proceedings, 1408–1411.
18. Zhang X., Zhang R., Chen W. (2010), *Design of digital parametric equalizer based on second-order function*, Proceedings of IEEE International Conference on Image Analysis and Signal Processing, 182–185.