

# Efficient Individualization Method of HRTFs Using Critical-band Based Spectral Cue Control

Yoomi Hur\*, Young-cheol Park\*\*, Seok-Pil Lee\*\*\*, and Dae Hee Youn\*

\*Dept. of Electrical and Electronic Engineering, Yonsei University, Seoul

\*\*Computer and Telecommunications Eng. Division, Yonsei University, Wonju

\*\*\*Korea Electronic Technology Institute (KETI), Seoul

(Received January 19, 2011; accepted April 11, 2011)

## Abstract

Recently, 3-D audio technologies are commonly implemented through headphones. A major problem of the headphone-based 3-D audio is in-the-head localization, which occurs due to the inaccurate Head-Related Transfer Function (HRTF). Since the individual measurements of HRTFs are impractical, there have been several researches for HRTF customization. In this paper, an efficient method of customizing HRTFs for the sound externalization is proposed. Firstly, it is determined which part will be customized in HRTF through psychoacoustical experiments. Then, the method controlling spectral notches and envelopes to provide individual localization cues are described. Since the proposed method is based on a critical-band rate, the structure is much simpler than that of previous studies, but still effective. The performance was evaluated through a series of subjective tests, and the results confirmed that the customized HRTF using proposed method could replace the measured individual HRTF successfully.

**Keywords:** 3-D audio, HRTF individualization, Sound externalization

**Subject classification:** Acoustic Signal Processing (1), Musical Acoustics and Psychoacoustics (8)

## 1. Introduction

Three-dimensional (3-D) audio technology enables sound images to be created anywhere around the listener. The core of this technology is digital filtering according to the Head-Related Transfer Function (HRTF). HRTFs represent an acoustic path from a sound source located in a free field to the listener's eardrums, and they have a unique frequency response for each direction and person-to-person variability.

In general, headphone or loudspeaker is used to reproduce the intended sound images, and

they have their own advantages and disadvantages [1]. The headphone listeners are not bothered by the situation of outside environment, and there is no sweet spot limitation. However, the reproduced sound images frequently remain inside the head, and there is a lack of richness in the quality of the sounds. On the other hand, loudspeakers can give a good sound externalization and a natural auditory environment including room responses. Instead, that system needs an additional crosstalk canceller to eliminate crosstalk between loudspeakers, and also has a critical problem of sweet spot.

Recently, as many portable devices offer various audio technologies, it is common that the users enjoy private listening by headphone or earphone at anytime and anywhere. In addition,

---

Corresponding author: Yoomi Hur (yumth@dsp.yonsei.ac.kr)  
DSP Lab., Yonsei University, 134 Shinchon-dong, Sudaemoon-gu,  
120-749, Seoul, Korea

the binaural system using headphone is considered optimal for reproducing 3-D sounds because of their relative immunity to the acoustical distortions [2]. Therefore, the problem of in-the-head localization is important, because the sound image cannot be located at a proper position if the image exists within the listener's head, which could degrade the effect of 3-D audio sounds.

The factors causing in-the-head localization have been reported in many articles [3-4], and the primary reason of in-the-head localization is known by that the spectral features of individual HRTFs are not well preserved [5]. To solve the problem, several researches have been suggested. Since the individual HRTF measurement is difficult, most approaches are based on additional information such as room reverberation, head movement and decorrelation technique. Although these methods can improve externalization perception, localization is blurred due to the inaccurate spectral cues. Thus, there have been other researches on individualizing general HRTFs [6-8]. However, those approaches are still problematic. Some of them need high computational cost, and some of them have the problem of requiring accurate process since the method is sensitive to relatively small changes.

In this paper, we propose an efficient method of customizing HRTFs, which achieves the customization by controlling spectral notches and envelope. Since the direct tuning of these spectral parameters at each frequency bin is time consuming, crucial parts of HRTF for sound externalization is investigated first. Then, those parameters are controlled based on a critical-band rate, which greatly simplifies the tuning process.

This paper is organized as follows. Section II reviews the problem of in-the-head localization, and summarizes various approaches to reduce the problem. A series of experiments to investigate the crucial parts of HRTF for sound externalization

are described in Section III, and the psychoacoustical evaluation for the effect of critical-band on spectral perception is presented in Section IV. In Section V, we propose an efficient method of customizing HRTFs. Finally, the conclusion is presented in Section VI.

## II. Background

### 2.1. In-the-head localization problem

One of the disadvantages on headphone listening is known as in-the-head localization, which means the sound images are localized within a listener's head. The factors that are supposed to cause in-the-head localization have been reported in many articles [9-13]. The major causes of the in-the-head localization are summarized below.

First, the existence of headphone, in itself, could induce the listeners to perceive sounds as originating inside the head [3]. However, several experiments made clear that it is not obvious parts of the problem. Secondly, it has been suggested that inaccurate acoustic cues could be a critical factor to the externalization perception. It was reported that the use of pinnae of others, not those of listener's own, will lead to inaccuracies and degraded externalization perceptions [10]. Thirdly, the lack of decorrelation between signals of both ears is known as an additional factor which appears to have an effect on perceived poor externalization [14]. Finally, the absence of reflections from walls and other obstacles within the recording and listening environments could also cause in-the-head localization [13].

### 2.2. Previous investigations

For many years, the quest of reproducing a natural external sound field through headphones has been studied. The principle of externalization is to reproduce the sound waveforms of the actual sound field at a listener's eardrums using headphones or earphones. However, this process

is difficult since the HRTFs must accord with the listener's HRTFs; therefore, the measurement of individual HRTFs is needed at every time the headphones are set on a listener's head.

Several studies have been described as possible means of which the in-the-head localization may be reduced without individual cues. Some of the researches have been conducted into externalization of sound images by observing the importance of the room reflected sound energies. In 1991, Rubak conducted the experiment for out-of-head localization using synthesized early reflections [9]. As an experimental result, Rubak suggested that the addition of only the 4 early reflections could improve the spatial impression. Begault also reported that the synthetic reverberation could give an effect on the degree of perceived externalization [15]. Although these approaches have advantages of ameliorating externalization perception significantly, adding reflected energy to recording decreases localization accuracy [16]. There have been other researches on the relation between the sense of externalization and head movement. As the experimental results of previous studies show, the introduction of dynamic HRTFs does contribute to reduce the front-back localization misjudgment and inside head localization. However, in the study of Begault, it was mentioned that the head-tracked system did not significantly increase externalization rates, nor did it yield more accurate judgments of azimuth angle [15]. Even if such a system can give good performances, the system size is too big and the computational complexity would be high. Decorrelation between the left and right channel signals is also known as a factor which appears to have an effect on external perception [14]. In 1995, Kendall created dynamic decorrelation which could produce a spatial effect similar to that of an environment. However, this method also has a problem of localization accuracy.

The possible methods as explained above

can give an improved externalization perception, especially in the case of using reverberation, but those methods still have a serious problem when the very frontal and back side positions are simulated. It is well known that this median-plane localization is difficult due to the lack of interaural differences, unless the individual spectral features are accurately reproduced [3].

HRTFs have acoustic cues that represent the filtering characteristics of the head and pinnae. Thus, the specific spectral characteristics are significantly different among the people, and they cannot have an accurate localization perception unless their own spectral cues are provided. It has been proved that the individual HRTFs can improve localization accuracy, increase externalization, and reduce reversal errors [10]. Since the process of measuring every listener's HRTFs is very impractical, it is necessary to find a method that can provide individual spectral cues without the measurements.

This leads to the need for individualizing HRTFs, and there were several trials to achieve the purpose [10,17]. These researches include database interpolation method which finds the best matching one in a given database of HRTFs [6]. Unfortunately, it is known that individual features of HRTF are sensitive to relatively small changes of the other ear's shape; therefore, finding the best match remains a research problem [17]. Other approaches are numerical methods, which compute HRTFs by solving wave equation with the individual body shape [7]. However, accurate surface meshes are needed for the solution and the computation cost is very high. There is also HRTF modeling method by approximating the head by a rigid sphere [8]. But this method cannot generate the feature of pinnae and result in impractically large number of terms needed for the shape of head, torso and pinnae [17].

In this paper, we propose an efficient method of customizing HRTFs based on the critical-

band rate. Our work was motivated by the fact that the critical-band rate is a frequency scale with which people perceive sounds. In order to set the frequency range for the customization, we first investigate which parts in HRTFs are crucial on the sound externalization.

### III. Crucial parts of HRTFs on the sound externalization

It is well known that the response of individual external ear is reflected mostly in high frequency region above 4 kHz, and the spectral notches of HRTF in that region stand for individual localization cues [10]. Thus, it is assumed that the individual spectral cues above 4 kHz are crucial on the externalization perception. A series of psycho-acoustical experiments were designed and performed to verify our hypothesis.

#### 3.1. Experimental environment

For the experiments, eight non-paid students from 20 to 30 years of age were participated. They all have normal hearing and some of them were participated in similar hearing or localization tests before, some of them were not. All the measurements and tests were performed in a small anechoic room to exclude any other factors that can affect the sense of externalization. The size of anechoic room was 2.0 m x 2.4 m x 2.5 m, and the listener sat in the center of the room. Single loudspeaker (BOSE 101 MUSIC MONITOR SPEAKER) was used placed at 0° azimuth and a distant of 1m away from the listener. The loudspeaker was uncovered thus visible to the subjects at all times.

Before starting the tests, each listener's individual HRTFs were measured. The measurements were made at the blocked ear canals with a small microphone (TURNER TP-845) which had a diameter of approximately 0.5 cm. Due to safety reasons, we did not try to place

microphone near the eardrum. The impulse responses were measured via the loudspeaker at 0° azimuth and a distance of 1 m, and a headphone (AKG K240 DF STUDIO MONITOR HEADPHONE). The B&K PULSE 3560 was used to the measurements, and the transfer functions were computed by the B&K FFT Analyzer 7770. The impulse response was obtained by averaging 20 times repeated measurements on each subject. Data was taken at 4096 points with quantization at 16 bits and a sampling frequency of 51200 Hz which is the nearest value bigger than 44100 Hz that can be supported by the equipment. Some example magnitude responses of measured HRTFs for three subjects are shown in Fig. 1. It can be seen that the spectral shapes are highly variant in the high frequency region relative to the low frequency region.

Then, the same subject wore the headphone, and the transfer function from the headphone to the in-ear microphone was also measured to compensate the distortion induced by the headphone worn by a listener.

In general, speech signals are commonly used in the localization test because of its familiarity. Speech signals were, however, incongruent with our experiments, since they hardly have energies in the frequency range above 4 kHz. Thus, the monophonic source signals with wide band spectra were carefully chosen. Our test materials consisted of five source signals: single instrument (SI), multiple instruments (MI), two complex music (CM, CM2), and white noise (WN) signal as summarized in Table 1. The stimuli used in the experiments were basically obtained by convolving these five source signals with the listener's HRIR.

For each experiment, listeners were heard two different stimuli that were binaurally synthesized, and asked to estimate the distant perception of reproduced sounds by headphone, especially, in a view of 'out-of-head' or 'in-the-head' localization. Before the experiments, no subjects had a

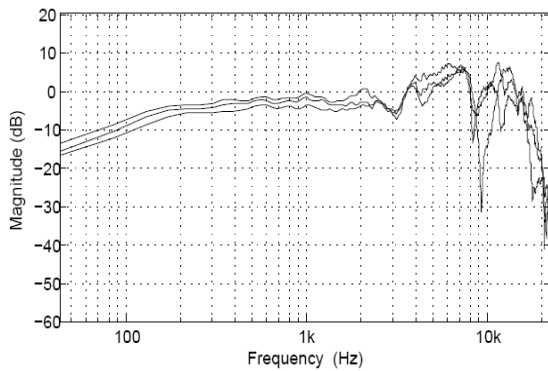


Fig. 1. Example magnitude responses of measured HRTFs from three subjects.

Table 1. Test source description.

Abbreviation	Duration	Description
SI	2sec	Hihat cymbal (Drum Sound)
MI	6sec	“Black bird on the wire” (Instrumental Music)
CM	20sec	My Little Brid O.S.T (Korean Pop)
CM2	25sec	“Fade to Black” (Metal)
WN	4sec	White noise (Noise)

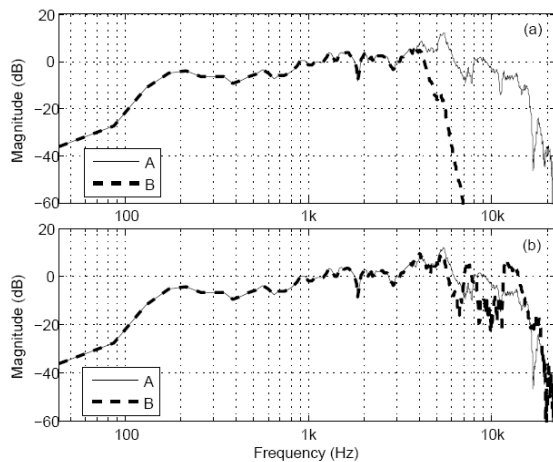


Fig. 2. Example magnitude responses of HRTFs for stimulus A (solid line) and stimulus B (dotted line): (a) Experiment I (b) Experiment II.

training session since the training could lead them to a biased judgment [18]. The overall test took about 30 minutes per a listener.

### 3.2. Psychoacoustical evaluation

A set of psychoacoustical tests were carried out. The purpose of Experiment I is to examine the effect of the existence of high frequency cues on the sense of externalization in the

median plane. For this test, two stimuli were synthesized using the same listener’s own HRTF but with different frequency range. In stimulus A, the frequency range of HRTF was maintained as the original spectrum of the measured transfer function. On the other hand, the frequency range of stimulus B was limited below 4 kHz by low-pass filtering. Fig. 2 (a) shows the two transfer functions which were applied to stimulus A (solid line) and B (dotted line), respectively. The spectral features of frequency range above 4 kHz were reflected to stimulus A, whereas they were not reflected in stimulus B.

Experiment II is designed to verify the hypothesis that the individual spectral features above 4 kHz are dominant to the sense of external perception. The experimental setup was same as Experiment I, except that the stimuli were changed. The frequency range of HRTFs was not limited for both of the stimuli in the test. Instead, stimulus A was generated using the listener’s own HRTF defined over full frequency range, and stimulus B was generated using a synthesized HRTF having the listener’s own HRTF below 4 kHz and randomly chosen HRTF above 4 kHz. Fig. 2 (b) shows the magnitude responses of HRTFs used for stimulus A (solid line), and that used for stimulus B (dotted line) for Experiment II. It can be seen that the spectral features of high frequency are different in both stimuli although the features below 4 kHz are same.

Two different test methodologies were used for psychoacoustical evaluation. First one is absolute distant test, and the second one is relative distant test.

#### 3.2.1. Subjective test for an absolute distant perception

The test of absolute distant perception was carried out to ascertain that, viewed in an absolute perception, how distant listeners sense a specific stimulus. As illustrated in Fig. 3, two reference

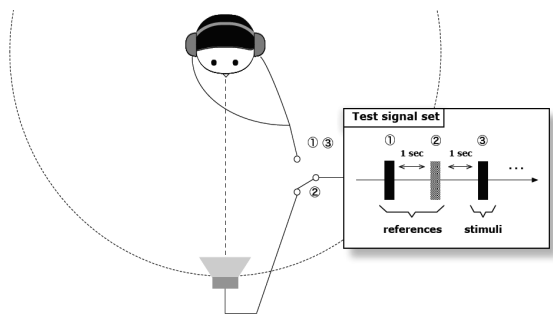


Fig. 3. Absolute distant test procedure.

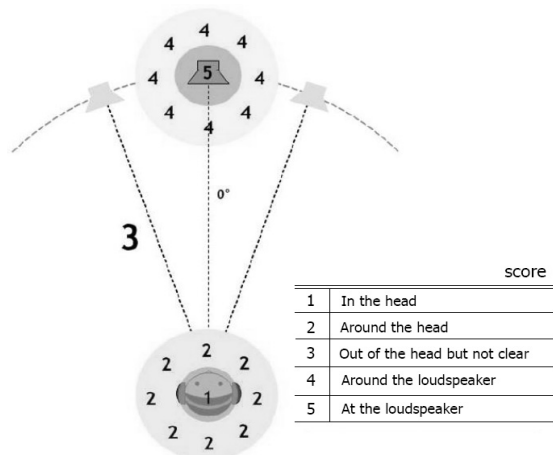


Fig. 4. Scores for absolute distant test.

signals were given with explanation before the stimulus to help an absolute distant judgment: an in-the-head localized sound through headphone and a real external sound coming from a loudspeaker at 1 m distance. Fig. 4 shows the absolute test scores on a scale of 1 to 5, which was given to the subjects during the listening test.

Test results for Experiment I and II are shown in Fig. 5 and Fig. 6, respectively. In each figure, the average score of all responses and 95 % confidence interval are marked for each test source signals. The square and circle represent the scores for stimulus A and B, respectively.

It can be seen from Fig. 5 that the average scores of stimulus A are always higher than stimulus B for all the source signals, which means that the absolute distant sensed from stimulus A are farther than that from B by the subjects. Overall, stimulus A has the mean score of 3.03 and stimulus B has that of 2.04. There is general understanding that existence of high frequency

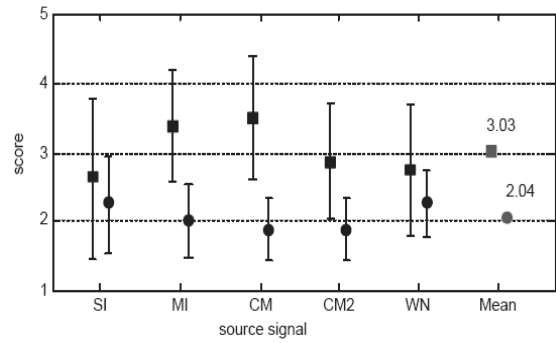


Fig. 5. Absolute test results of Experiment I.

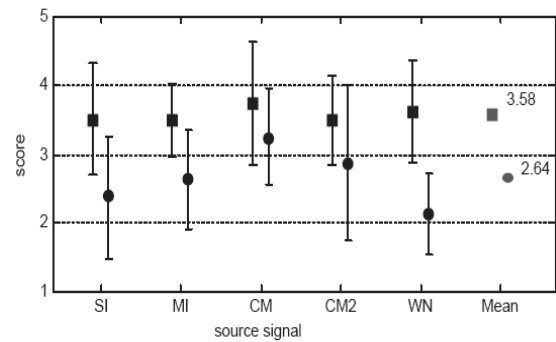


Fig. 6. Absolute test results of Experiment II.

tends to improve the externalization perception, and this has been supported by our experiment. Similar results can be found in the research of Weinrich [10]. They reported that the high frequency region from 4 kHz to 12 kHz affects to the external and front perception.

The test results of Experiment II can be found from Fig. 6 that stimulus A, which includes individual cues, has higher scores than stimulus B which includes non-individual cues above 4 kHz. The scores of stimulus B are far from satisfactory; although there exists higher frequency energies and the individual cues below 4 kHz were preserved. Therefore, it can be proved that the crucial factors of externalization perception are not only the existence of higher frequency energy but also the individual spectral features in it. In Fig. 6, differences of scores are relatively small for complex music signals. That is because, in complex music signal, most of the energies exist in the lower frequency mainly due to the vocal sound. Therefore, the spectral features above 4 kHz are sensed relatively weak. Note that the difference

is large especially in case of white noise signal having equally distributed energies along the frequency. The overall mean score is 3.58 for stimulus A, and 2.64 for stimulus B.

### 3.2.2. Subjective test for a relative distant perception

Next, paired comparison tests were carried out to explore the relationship between absolute and relative auditory perception. In the relative distant test, the reference signals were not given and only the two stimuli were alternately heard to the subject as illustrated in Fig. 7. They were asked to compare the stimuli A and B directly, and judge which one is more external or distant than the other.

Table 2 shows the test results for Experiment I and II. The percentage of perceptual relative distance of two stimuli out of the total number of responses is listed. In Experiment I, most subjects perceived stimulus A, which contains the higher frequency energy, more external than stimulus B. The percentage of listeners who answered stimulus B is more external than stimulus A is just 15 %. In addition, the 15 % comes from only the responses of three subjects, and especially one subject make up the half percentage of the 15 %. Thus, it can be confirmed that the existence of high frequency energy can improve the externalization perception for the most subjects. The results of Experiment II indicate that individual spectral cues above 4 kHz play a dominant role on the externalization perception. Listeners could distinguish their own spectral cues from the other's one, thus the percentage of response that stimulus B, which does not reflect individual features, are more external than stimulus A is only 12.5 %. Moreover, the percentage of 12.5 consists of only three subjects, and the 60 % of the 12.5 % is made by one person. However, more listeners answered that the two stimuli were similarly distant in a relative distant test compared to the absolute

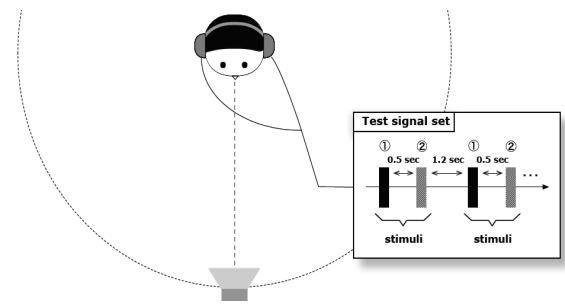


Fig. 7. Relative distant test procedure.

Table 2. The results of relative distant test.

Experiment I		Experiment II	
Relative distance	Percentage (%)	Relative distance	Percentage (%)
A>B	65	A>B	52.5
A=B	20	A=B	35
A<B	15	A<B	12.5

distance test. It was found that listeners had a difficulty to judge the distance without reference signals, and they were apt to perceive the paired sounds more similarly distant than listening them separately.

As a whole, the results from paired comparison test are coincided with the absolute distant test, confirming that the frequency range above 4 kHz is crucial for the sound externalization perception.

## IV. The effect of critical-band on spectral perception

Although the frequency range for the customization could be specified to the higher frequency through the previous experiments, it is still difficult to directly control the spectral cues at each frequency bin. Thus, we performed a psychoacoustic experiment to validate the critical-band effect on spectral perception especially for the notch frequency.

For the experiments, the spectrum of individual HRTF is replaced by that of digital notch filter for the frequency range from 4 kHz to 12 kHz, which is known as the range of pinna notch frequency. Since the purpose of experiment is

to examine the perceptual influence of notch frequency when the notch position is varied across the critical-band, the information of spectral envelope, which can influence to the result, was not provided in that frequency range. Fig. 8 shows an example of measured HRTF (dotted line), and synthesized HRTF with a digital notch filter for the experiment.

Two stimuli were generated with the synthesized HRTF in which a subject's own spectrum is replaced by digital notch filter for the frequency range from 4 kHz to 12 kHz. Stimulus A was generated by convolving source signal with the synthesized HRTF whose individual notch frequencies were preserved. On the other hand, stimulus B was generated by using the synthesized HRTF whose individual notch frequency was carefully moved to other frequency as in Fig. 9. Randomly selected one notch frequency of each subject was moved in the test, whereas the positions of all the other notch frequencies were fixed as they were. Then, subjects were asked to judge the perceptual similarity between two stimuli based on the score in Table 3. If the subjects perceive the sound of two stimuli very similar, it can be said that the two sounds are localized at the same location. This inference is based on the fact that people determine localization by the sound timbre [19].

Fig. 10 shows the experimental results for three subjects as an example. In each figure, the vertical dotted line shows the boundary of critical band, and the results show that subjects had a difficulty in distinguishing a difference if the notch frequency varied within a critical-band, whereas it was easily recognized if the notch frequency was moved outside the critical-band. For the experiment, we carefully moved the notch frequency to distinguish the fact that the cause of perceptual difference between stimuli is whether the frequency distance between the notches or the out-critical-band movement.

Especially in Fig. 10 (b), the subject perceived

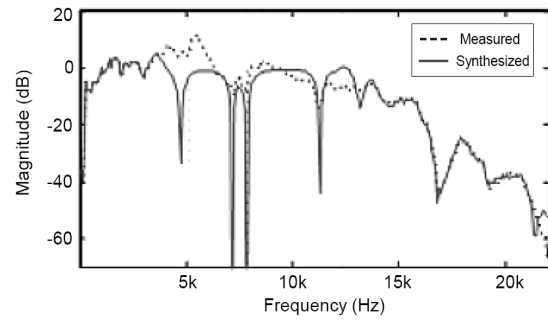


Fig. 8. Example of measured HRTF (dotted line) and synthesized HRTF with notch filter (solid line).

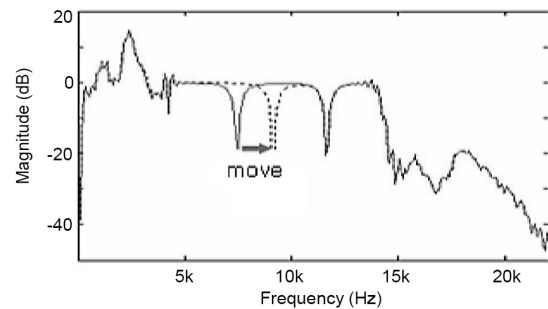


Fig. 9. Example of moving notch frequency.

Table 3. Perceptual similarity score for the test.

Score	Description
0	There is no difference
1	There is little difference
2	There is a little difference
3	There is difference

the sound similarly when the notch frequency was moved from 7.72 kHz to 8.5 kHz, while they perceived the sound differently when the notch frequency was moved to 7 kHz although moved distances were almost same in both of the cases. It is noticeable that the score was even higher when the notch frequency was moved to 7.3 kHz than when it was moved to 8.5 kHz. That is, the subject perceived the sound more differently when the notch frequency was moved to other critical-band although the frequency distance was shorter. Therefore, based on these results, it can be proved that the exact position of individual notch is not important, as long as the notch is retained within the same critical band. Since the frequency range from 4 kHz to 12 kHz can be represented by five critical bands, we need to



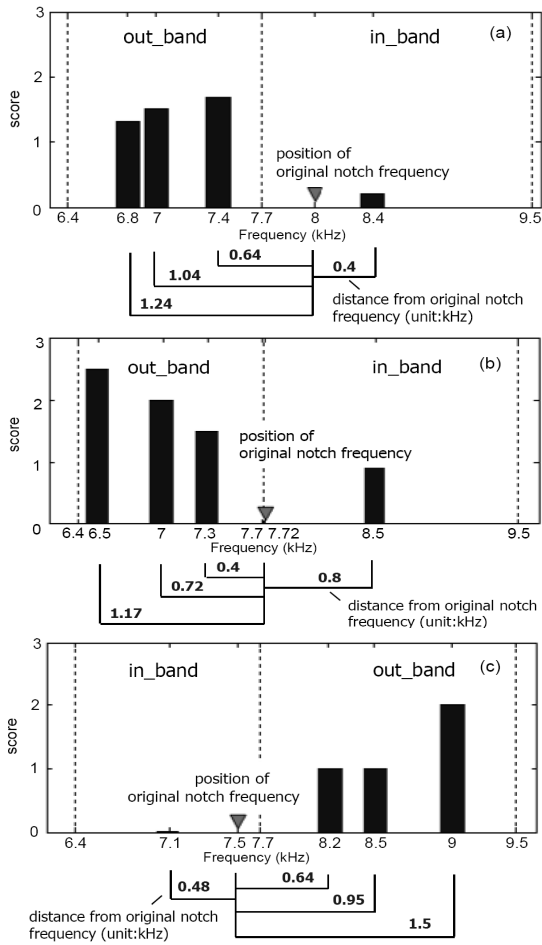


Fig. 10. Experimental results of moving notch frequency: (a) subject 2 (b) subject 6 (c) subject 8.

control at most five notches in those five critical bands, which makes our proposed method much simpler.

## V. Proposed method of customizing HRTFs

In this section, we describe our proposed HRTF customizing method controlling individual spectral cues based on critical-band rate.

### 5.1. Spectral notches

The individual notches are very different from person to person, such that, in which frequency the notches exist or how many notches the subject has. Fig. 11 shows example spectral notches in measured HRTFs of two subjects. Note that all

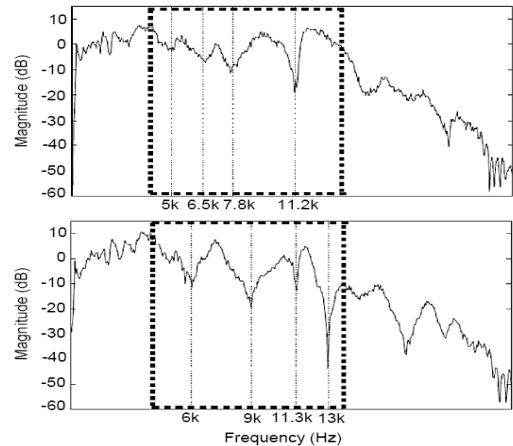


Fig. 11. Spectral notches in measured HRTFs.

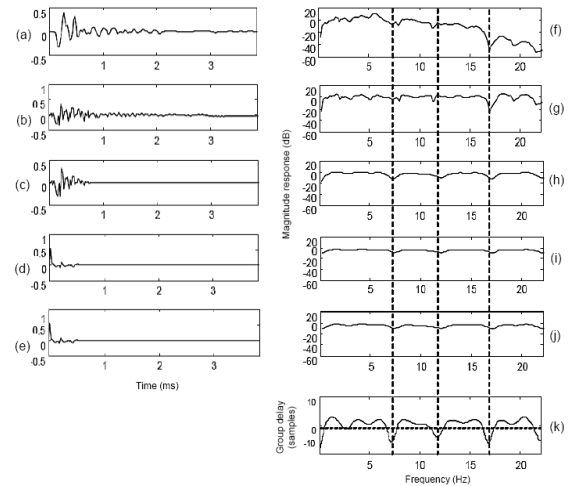


Fig. 12. The steps for extracting pinna notch frequencies: (a) Original HRIR (b) 12th order LP residual (c) windowed LP residual (d) autocorrelation function of windowed LP residual (e) windowed autocorrelation function. The plots (f), (g), (h), (i) and (j) show the magnitude spectrum corresponding to signals in (a), (b), (c), (d) and (e), respectively. (k) shows the group-delay function of the windowed autocorrelation function. Local minima with threshold  $-1$  are shown.

of the spectral notches are not caused by pinna. Since the reflections from shoulder or torso can also be responsible for some of them, the algorithm reported in 2004 [20] was applied to the measured HRTFs in order to extract pinna notch frequencies.

The extracting algorithm is based on several signal processing techniques including LP analysis, windowing, autocorrelation and group delay function as in Fig. 12 (a) to (k). The results of each procedure are shown with the time-domain

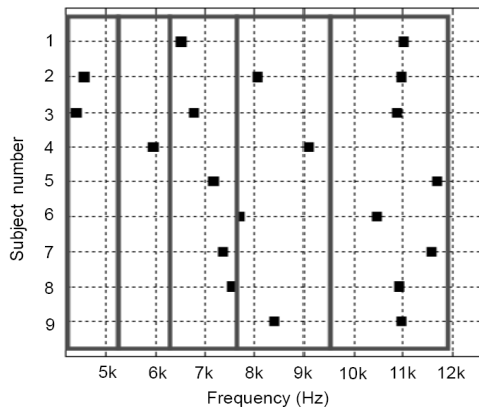


Fig. 13. Distribution of pinna notch frequencies.

representations in left column and the frequency-domain representations in right column. From the measured HRIR signal as in Fig. 12 (a), the group-delay of windowed autocorrelation function is plotted in Fig. 12 (k), and the local minima with the threshold of  $-1$  are shown indicating the pinna notch frequencies.

As the previous experiments showed that the notch should be retained within the same critical band, it is now important how the subjects can control their notch frequencies. Fig. 13 shows the distribution of pinna notch frequencies extracted from the measured individual HRTFs for nine subjects. The x-axis symbolizes frequency, y-axis symbolizes subject number, and each red box represents one critical band. According to the observation of Fig. 13, the individual notches can be expressed with the number and their positions. The number of notches is varied from two to three among the subjects, and the case of having two different notches was most common in our measurement: seven out of total nine subjects have two spectral pinna notches. Also the notches seem to be distributed in certain ranges although the positions are not exactly the same. For example, most subjects have one notch at the region from 6 kHz to 8.5 kHz and the other subjects have one notch around 11 kHz. This is reasonable in a sense since human ears have common shapes and similar size.

Since we proved that subjects perceive the

sound similarly if the notch frequency remains within the same critical-band, the notch frequencies of control system can be set to the center of each corresponding band. Then, the only thing the subjects have to do is to select several bands, which might include their notch frequencies, among five bands. A series of experiments (Experiment III to V) were performed to examine that the subjects could distinguish the difference, if they selected other critical band that did not include their notch frequency. Six subjects participated and white noise was used as a source signal.

In Experiment III, two stimuli were generated by convolving source signal. Stimulus A was generated using HRTFs in which individual notch frequencies were preserved, and stimulus B was generated using HRTFs in which two of the notch frequencies were randomly changed to the center frequencies of other critical bands. The two stimuli were heard to the subjects and judged the similarity of the sounds according to the score in Table 3. Fig. 14 shows the experimental result and it can be seen that almost all of the scores are over 2 with the average score 2.27. That is, most subjects perceived the sound of stimulus B to be very different from that of stimulus A.

Next, Experiment IV was carried out to see how subjects perceive sound if one of the notch frequencies was moved to the center frequency of other critical band. The subjects were also asked to score based on how they perceive the two stimuli. In Fig. 15, the scores are still over 1.5, which means that the subjects can distinguish the difference even only one of the notch frequencies is changed.

In the last Experiment V, the frequencies of pinna notches were slightly changed within their own critical-band and the stimuli were scored. The result is shown in Fig. 16. As expected from the results of the validation in Section IV, the scores are below 1 and the average score

is 0.6. Subjects perceived the modified sound very similarly to the original sound even though the notch frequency was varied.

Taking these results into consideration, it can be proved that people perceive the sounds similarly if we choose the correct critical-band where their individual notches are located. In other words, subjects can distinguish the critical bands that include their own spectral notches from the bands that do not include.

In order to simplify the tuning procedure, the five frequencies corresponding to the center frequencies of five critical bands, can be given to the subjects for possible choices. The only task is to choose several bands so that the sound can be heard as a reference externalized sound. Furthermore, most subjects, at least those participated in our experiments, have two pinna notches, and almost every subject has their

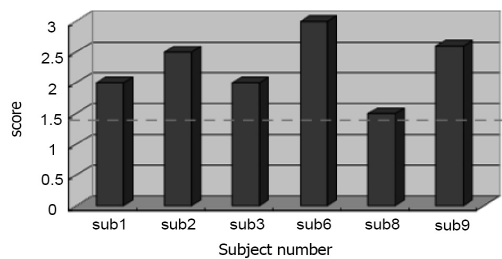


Fig. 14. The result of Experiment III.

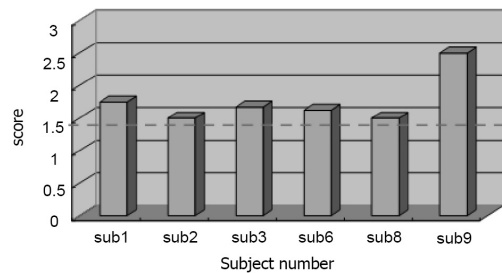


Fig. 15. The result of Experiment IV.

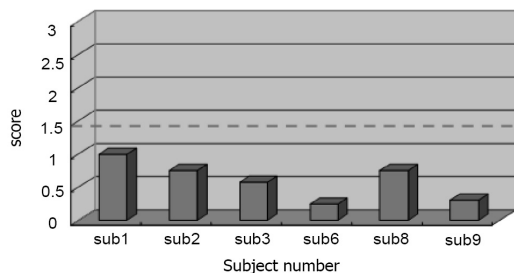


Fig. 16. The result of Experiment V.

pinna notch in the fifth critical band as shown in Fig. 13. Thus, the tuning procedure can be further simplified.

### 5.2. Spectral envelope

Although it is well known that the individual spectral notches are important to the sound localization, human auditory system uses not only the spectral notches but also the overall spectral envelope [21]. Thus, additional experiments were performed to find out how the information of spectral envelope can be provided to the listeners using critical-band rate. The basic idea is to adjust the spectral envelope by defining several reference points, which are the boundaries of each critical-band. In Fig. 17, dotted lines represent the boundaries of each critical band and the six black dots represent the reference points for controlling spectral envelope.

By adjusting the magnitude of reference points, the spectral envelope can be shaped as in Fig. 18.

The comparison between measured HRTF (thin line) and synthesized HRTF (thick line) without spectral shaping and with spectral shaping

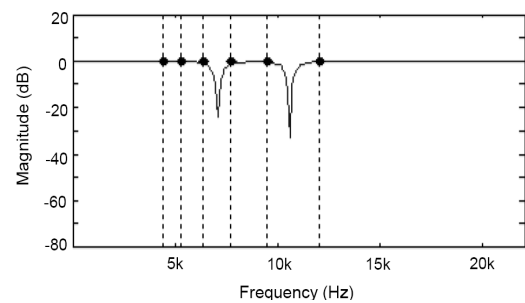


Fig. 17. Six reference points at the boundaries of each critical-band for spectral envelope adjustment.

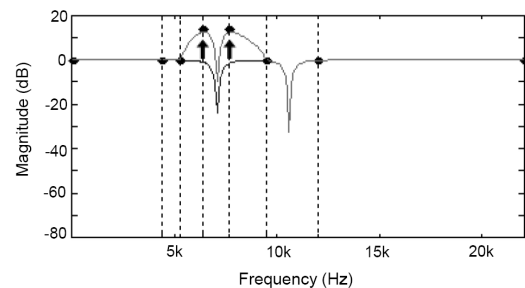


Fig. 18. Example of spectral envelope shaping with the reference points.

are shown in Fig. 19 (a) and (b), respectively. If the synthesized HRTF without spectral shaping is applied to generate sound, listeners may have a difficulty in localization due to the severe timbre change. On the other hand, subjects can localize sound correctly with the spectral shaped synthesized HRTF since the overall spectral shape is well captured by the reference points.

To evaluate the effect of spectral envelope shaping based on the reference points, another psychoacoustical experiment was performed. In this experiment, several stimuli were generated by randomly adjusting spectral envelopes with the six reference points. These stimuli were heard to the subjects, and then they were asked to find out the most effective one. Fig. 20 shows an example of adjusted four HRTFs for one subject. The thin solid line in all of the figures represents the subject's own HRTF, whereas the thick solid line in each figure represents the randomly shaped synthesized HRTF based on the reference points.

Six subjects participated for the experiment, and all of them could easily find out the stimulus which was generated from the most similarly shaped HRTF. Since the spectral envelope is directly connected to the sound timbre, it was not a difficult work compared to that of selecting critical bands for notch frequency to the listeners.

Through a series of psychoacoustical experiments, we showed that the individual localization cues such as spectral notches and envelope can be controlled by applying critical-band rate. Subjects may easily tune the parameters to find their individual cues by selecting one or two critical band among five bands for spectral notches and by adjusting magnitude of six reference points for spectral envelope. Since our proposed method has a relatively small number of parameters, and they are controlled on a basis of critical-band rate, the individualization process becomes simpler and user-friendly than other customization methods.

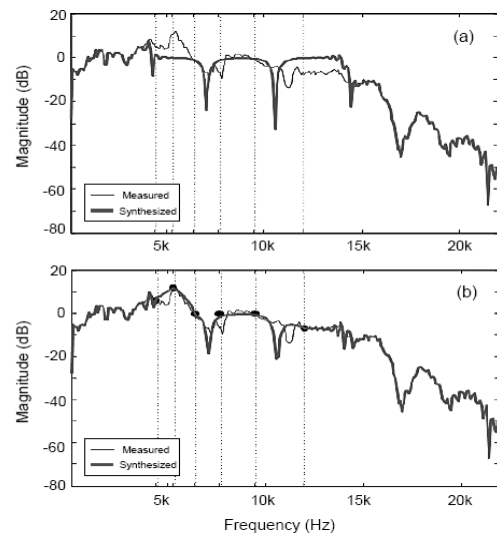


Fig. 19. Measured HRTF (thin line) and Synthesized HRTF (thick line): (a) without spectral shaping (b) with spectral shaping.

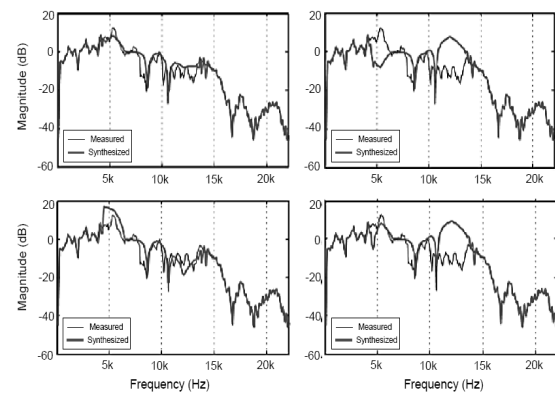


Fig. 20. Example of randomly shaped HRTFs for a subject.

## VI. Conclusions

An efficient method of customizing HRTFs by controlling spectral notches and envelope was proposed in the paper. First, we performed psychoacoustical experiments to validate which parts in HRTF are crucial on the externalization perception. Then, the effect of critical-band based spectral parameter control was investigated, and it was proved that the exact position of individual notch was not important, as long as the notch was retained within the same critical band. Based on these results, the proposed system identified a single notch per critical-band, and its frequency was set to the center of the corresponding band.

Since the crucial parts for spectral cues of HRTFs, which include the pinna notches, can be represented by five critical bands, we need to control five notches in those five critical bands at most. After determining notch frequencies, the spectral envelope is tuned based on the same critical-band rate. It was shown that, by controlling the energy of each critical band, the sound timbre of individual HRTF could be maintained.

Since the proposed method controls both spectral notches and overall spectral envelope based on a critical-band rate, the method is much simpler than the previous customization methods, but still effective. However, the performance of overall process has not been carried out yet, although the performance of each tuning process for spectral notches and envelope was evaluated separately in this paper. The additional research and experiments for the overall tuning process should be done for future works.

---

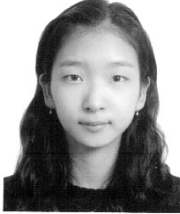
## References

---

1. C.-J. Tan, W.-s. Gan, "Direct concha excitation for the introduction of individualized hearing cues," *J. Audio Eng. Soc.*, vol. 48, no. 7/8, pp. 642–653, July, 2000.
2. D. R. Begault, "Auditory and non-auditory factors that potentially influence virtual acoustic imagery," *AES 16th International Conference*, paper no. 16-002, Mar, 1999.
3. J. Blauert, *Spatial Hearing*, MIT Press, Cambridge, 1983.
4. D. M. Green, "Psychoacoustics," *CHABA Symposium on Sound Localization by Humans, National Academy of Sciences (unpublished)*, 1988.
5. W. M. Hartmann, A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3678–3688, June, 1996.
6. S. Shimada, N. Hayashi, and S. Hayashi, "A clustering method for sound localization transfer function," *J. Audio Eng. Soc.*, vol. 42, no. 7/8, pp. 577–584, July, 1994.
7. Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, "Numerical modeling of the transfer functions of a dummy-head and of the external ear," *AES 16th International Conference*, paper no. 16-029, Mar, 1999.
8. M. J. Evans, J. A. S. Angus, and A. I. Tew, "Spherical harmonic spectra of head-related transfer functions," *AES 103rd Convention*, paper no. 4571, Sep, 1997.
9. P. Rubak, "Headphone signal processing system for out-of-head localization," *AES 90th Convention*, paper no. 3063, Feb, 1991.
10. Weinrich, S. Gert, "Improved externalization and frontal perception of headphone signals," *AES 92nd Convention*, paper no. 3291, Mar, 1992.
11. F. E. Toole, "In-head localization of acoustic images," *J. Acoust. Soc. Am.*, vol. 48, no. 4B, pp. 943–949, 1970.
12. L. Wightman, D. J. Kistler, "Headphone simulation of free-field listening I : Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858–867, Feb, 1989.
13. N. Sakamoto, T. Gotoh, Y. Kimura, "On out-of-head localization in headphone listening," *J. Audio Eng. Soc.*, vol. 24, no. 9, pp. 710–716, Nov, 1976.
14. G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, 1995.
15. D. R. Begault, E. M. Wenzel, A. S. Lee, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized Head-Related Transfer Functions on the spatial perception of a virtual speech source," *AES 108th Convention*, paper no. 5134, Feb, 2000.
16. D. R. Begault, "Perceptual effects of synthetic reverberation on 3-dimensional audio systems," *J. Audio Eng. Soc.*, vol. 40, no. 11, pp. 895–904, Nov, 1992.
17. P. Satarzadeh, V. R. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," *AES 122nd Convention*, paper no. 7098, May, 2007.
18. D. Griesinger, "Binaural techniques for music reproduction," *AES 8th International Conference*, paper no. 8-026, May, 1990.
19. A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, no. 4, pp. 1152–1165, Apr, 1978.
20. V. C. Raykar, R. Duraiswami, B. Yenarayanan, "Extracting frequencies of the pinna spectral notches in measured head related impulse responses," *Perceptual Interfaces and Reality Lab., University of Maryland, Technical report CS-TR-4609*, 2004.
21. B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 820–836, 1989.

## [Profile]

### •Yoomi Hur



Yoomi Hur received B.S and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2004 and 2006, respectively. She is pursuing her Ph.D. degree at Yonsei University in the area of audio signal processing. Her research interests include 3-D audio signal processing, adaptive filter, array signal processing, and multimedia applications such as mobile phone, home theater, etc. At present, she researches about various array signal processing techniques for 3-D audio at the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA as a visiting researcher since 2008.

### •Young-cheol Park



Young-cheol Park was born in Korea in 1964. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1986, 1988, and 1993, respectively. From September 1993 to October 1995, he was a Postdoctoral Research Scholar at The Pennsylvania State University, University Park, where he worked on active noise control. From March 1996 to August 1998, he was with the Samsung Electronics Company, Ltd., Kiheung, Korea, and from April 1999 to February 2002, he was with InTime Corporation, Seoul, where he worked on digital hearing aids, digital audio processors, and VoIP systems. In 2002, he joined the Computer and Telecommunications Engineering Division, Yonsei University, where he is currently a Professor. His research interests include three-dimensional audio signal processing, high-quality audio coding, adaptive filtering, digital hearing aids, and their implementations.

### •Seok-Pil Lee



Seok-Pil Lee received B.S., M.S., and Ph.D. degrees all in electrical engineering from Yonsei University, Seoul, Korea in 1990, 1992, and 1997, respectively. He worked for Daewoo Electronics Corp., Korea as a senior research engineer from 1997 to 2002. From 2002, He works as a director at Digital Media Research Center, Korea Electronics Technology Institute (KETI), Korea. His main research area is digital TV, Personalized Service, and IPTV.

### •Dae Hee Youn



Dae Hee Youn received the B.S. degree in electronic engineering from Yonsei University, Seoul, Korea, in 1977 and the M.S. and Ph.D. degrees in electrical engineering from Kansas State University, Manhattan, in 1979 and 1982, respectively. From 1982 to 1985, he was an Assistant Professor at the University of Iowa, Iowa City. Since 1985, he has been with the Department of Electrical and Electronic Engineering, Yonsei University, where he is currently a Professor. His research interests include adaptive filters, speech/audio coding, and real-time implementation of DSP algorithm.