Enhancement of a 3D Sound Using Psychoacoustics

Kyosik Koo, and Hyungtai Cha

Abstract—Generally, in order to create 3D sound using binaural systems, we use head related transfer functions (HRTF) including the information of sounds which is arrived to our ears. But it can decline some three-dimensional effects in the area of a cone of confusion between front and back directions, because of the characteristics of HRTF.

In this paper, we propose a new method to use psychoacoustics theory that reduces the confusion of sound image localization. In the method, HRTF spectrum characteristic is enhanced by using the energy ratio of the bark band. Informal listening tests show that the proposed method improves the front-back sound localization characteristics much better than the conventional methods.

Keywords—HRTF, 3D sound, Psychoacoustics, Localization

I. INTRODUCTION

THREE-dimensional (3D) sound is becoming increasingly important for scientific, commercial, and entertainment systems [7]. It can greatly enhance auditory interfaces to computers, improve the sense of presence for virtual reality simulations, and add excitement to computer games.

3D sound places the virtual sound source with a mono sound in a given 3D space by adding pitch, tone and sound color, sense of direction and sense of distance.

It is well known that audio systems using 5.1 channels are DVD standards and the most common implementation of surround systems. But for a system with 5.1 channel speakers, we need actually six speakers which take lots of space and money.[3] Hence, binaural systems that use 2 channels come into the spotlight. The most recent systems have used HRTF to compose virtual 5.1 channels.

HRTF is a series of algorithms utilized to synthesize simulated binaural signals from a monaural source. It includes the cues that arise from the scattering process of sound from the user’s body, head, and ears [4]. However, because of the listener’s physical characteristics and other individualistic qualities, the use of non-individual HRTF can create confusion in up/down and front/back directional perception. The listener especially cannot separate the difference of each direction in the cone-of-confusion with just a difference of time or difference of level. To solve this problem we need a specific HRTF for each individual which cannot be accomplished in the real world. So we propose an algorithm to improve the confusion of front/back sound localization to make realistic 3D sound.

The paper is organized as follows. The basic theory to create 3D sound is described in Section II. In Section III, the proposed algorithm is presented. The simulation and experimental results are given in Section IV. Conclusions are drawn in Section V.

II. BASIC THEORIES

A. Head Related Transfer Function

When sound is generated from a source in 3D space, human can perceive the direction of sound source. Many researches have been proposed how human perceives sounds around him. As a result, HRTF is known as the most useful tool.

Generally, HRTF is the transfer function that expresses sound which arrives to a person’s external auditory meatus through a fixed incidence angle in a free field. It is used in binaural signal composition in virtual reality application system or auralization system. In the 3D space, sound direction is defined using azimuth and elevation [1]-[8].

In the horizontal plane, the most important cues that help a person localize a sound source are the interaural time difference (ITD) and interaural intensity difference (IID). Fig. 2 shows the typical HRTF measured using KEMAR dummy head at MIT Media Lab.

Using HRTF, 3D sound is created through technologies such as sound localization technology, sound field reproduction technology, etc. Sound localization technology is a technique...
that freely locates free sound source to three-dimensional space.

This can be the biggest characteristic and advantage of an only 3D sound technology. HRTF convoluted with mono sound which has no direction creates 3D sound \((\mathbf{Z}_L, \mathbf{Z}_R)\) of specific direction \((\theta, \Phi)\) where \(\theta\) is azimuth and \(\Phi\) is elevation in a spherical coordinate system. And \(S\) is mono sound.

\[
X_r(z, i) = a_r(z)X_r(z, i) = a_r(z)\sum_{\omega} X(\omega, i)
\]

where, \(0 \leq z \leq Z - 1\)

\(\omega_{zh}\) and \(\omega_{zd}\) mean low and high frequency boundaries about any critical band, \(z\) from total critical bands, \(Z\), respectively. \(a_r(z)\) is frequency dependence attenuation caused by elements of various transmission factors from external to middle ear.

\[
X_r(z, i) \text{ convolutes with perceptible energy spread function. Perceptible energy spreading function is the displayed diffusion phenomenon of signal energy in the basilar membrane. Using this, we calculate hearing stimulation energy, excitation}_\text{energy}(z, i) \text{ for each critical band}\).
\]

\[
SF(v, z) = 15.81 + 7.5(\Delta z + 0.474) - 17.5\sqrt{1 + (\Delta z + 0.474)^2}
\]

\(\Delta z\) is \(z - v\), a difference value between critical bands, that is expressed by bark index.

**III. THE PROPOSED ALGORITHM**

There are some methods to solve the confusion of front-back direction caused by using individualized HRTF. The most typical method is a spectral band emphasis/deemphasis method or filterbank method on psychoacoustics. In these methods, specific spectral bands play a key role in direction determination of sound. Filterbank method depicted in Fig. 3 divides input spectrum into several bands by bandpass filters and emphasize or deemphasize each band by configuration to have maximum localization ability [1].
The second method uses the distinctive difference of the ear’s projection degree which is a key to front/back direction perception in a human’s physical characteristics [5]. The bigger the projected degree is the better to sense front/back direction and frequency characteristics of the bigger angle is highlighted.

These methods control the fixed frequency bands without considering characteristics of HRTF between each direction. If the energy of mono sound is excessive at some bands, loss of direction or bad sound quality are sure to result.

Many researches show that sound direction is dependant on specific frequency band called the direction decision band. That is, irrespective of the sound direction, sound image is localized at the direction of frequency band emphasized in the spectrum of the signal reaching the eardrum.

Therefore, we proposed the algorithm that revises spectrum characteristics using psychoacoustics.

First we calculated the critical band energy of HRTF, that α is direction localizing sound image using frequency spectrum. Next, we calculated the critical band energy of HRTF that direction is β, HRTFβ. It is the symmetrical direction of α that causes confusion in direction perception.

Next, cband_energy(α,z,i) convolutes with perceptible energy spread function. Perceptible energy spreading function is the displayed diffusion phenomenon of signal energy in the basilar membrane. Using this, we calculate hearing stimulation energy, excitation_energy(α,z,i) for each critical band. Similarly, calculate hearing stimulation energy, excitation_energy(β,z,i) for direction β.

Direction decision band of HRTF exists for each direction as previously explained. As a result, if HRTF reaches in eardrum, critical band is changed according to superior direction decision.

\[
rate_α(z,i) = \frac{excit_α(z,i)}{excit_α(z,i)}
\]  

(5)

The resulting calculated weight is applied to the original HRTFα frequency spectrum to amplify the superior band for each direction. Then, ingredient such as spectrum’s peak or notch is emphasized. As a result, excitation energies of two direction’s HRTF are embossed.

\[
HRTF_β(f,z,i) = HRTF_α(f,z,i) \times rate_α(z,i)
\]

(6)

Here, \(f_α\) is the frequency band equivalent to each critical band.

In this paper, we suggest an algorithm to improve the confusion arising from front/back direction characteristics that happens in the process of HRTF.

Sound signals used in simulation are 44.1kHz, 16bits acoustic signals that are abstracted in audio CD and recorded using sound program. And we applied HRTF DB measured in MIT Media Lab.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>SOUN D SOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Sources</td>
</tr>
<tr>
<td>Source 1</td>
<td>Recorded voice</td>
</tr>
<tr>
<td>Source 2</td>
<td>Bomb sound</td>
</tr>
</tbody>
</table>

We carry out experiments to demonstrate the potential of the proposed method. The performance is determined via subjective evaluation and the assessment criterion is the ability to localize sound source. A small group of people whose major isn’t speech and sound processing is involved in the
experiments and evaluation. Four types of mono sound sources are employed for the experiments like Table I.

The objective of the set of experiments is to determine which of the three localization algorithm, namely the simple HRTF filtering, the filterbank method which shows the best performance among the conventional algorithm, and the proposed method, enables observers to localize sound sources more accurately. We graded the algorithms for feeling of localization in some directions.

<table>
<thead>
<tr>
<th>Table II</th>
<th>Simulation Test Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>Comment</td>
</tr>
<tr>
<td>1</td>
<td>Worst</td>
</tr>
<tr>
<td>2</td>
<td>Worse</td>
</tr>
<tr>
<td>3</td>
<td>Similar</td>
</tr>
<tr>
<td>4</td>
<td>Better</td>
</tr>
<tr>
<td>5</td>
<td>Best</td>
</tr>
</tbody>
</table>

As confirmed from Figs. 7 and 8, the proposed algorithm can revise the HRTF very effectively by calculating and applying the excitation energy in the direction decision band.

Using the proposed algorithm, we can revise front/back characteristics of HRTF in upper or lower plate as well as horizontal plane. Also, it is applied to revise lower/upper characteristics of HRTF easily.

V. CONCLUSION

In this paper, we propose an algorithm to improve the confusion arising from front/back direction to make real stereophonic sound. We use information such as level or time difference of sound that reaches the human ears in the case of creating 3D sound using HRTF in binaural system of 2 channels. But we cannot distinguish the differences of each direction in cone of confusion when the distance between source and listener’s two ear becomes the same.

Therefore, we propose an algorithm correcting the band which greatly influences the human ear effecting each direction using frequency characteristics of psychoacoustics and HRTF. It controls sound influence to the human hearing directly and emphasizes direction of HRTF. So we could create an accurate sense of direction which gives a true, real sound than the existing method that uses frequency spectrum.

Also, we conducted listening tests with audio non-professionals to confirm the improved results and checked proposed algorithm’s improvement results.

This algorithm add real sense of feeling sound technology modeled from sound sources to human eardrums considering that humans perceive all sound with only 2 ears. However, this hearing model has limits since it affects everyone equally because of its subjective characteristics. Therefore more research is needed to reduce errors and accomplish a more perfect hearing model.

REFERENCES


**Kyosik Koo** was born in Paju, Korea in 1979. He received the B.S. and M.S. degree from the School of Electronics Engineering, Soongsil University, Korea in 2005 and 2007, respectively. He is currently pursuing the Ph. D. degree at Soongsil University. His current research interests are audio signal processing, spatial audio and psychoacoustics.

**Hyungtai Cha** received the M.S. and Ph.D. degree in dept. of Electrical Engineering from the University of Pittsburgh in 1988 and 1993 respectively. He is currently an Associate Professor in the School of Electronic Engineering, Soongsil University. His recent research interests include Multimedia Systems and Applications, Audio and Video Signal Processing, ASIC and DSP Implementation of Digital System, and Communication System.