3D sound effect enhancement of front and back

Lu, Dongdong[1]
College of Marine Science and Technology, Northwestern Polytechnical University
No. 127 Youyi West Road, Beilin District, Xi'an, Shaanxi, China

Zeng, Xiangyang[2]
College of Marine Science and Technology, Northwestern Polytechnical University
No. 127 Youyi West Road, Beilin District, Xi'an, Shaanxi, China

ABSTRACT

In general, when the virtual 3D stereo synthesized with Head Related Transfer Function （HRTF) replayed by headphones, it is difficult for the listener to accurately distinguish the position in which the sound comes from the front (0º, 0º) or the back (0º, 180º). Even with the personalized HRTF data, front and back confusion still exists, which means the phenomena of front-back confusion are caused by the characteristics of the HRTF itself. Conventional methods are mainly achieved by changing the original spectrum.

In this paper, we put forward a method to reduce the confusion cone of front and back by amendment ITD of the HRTF. In this method, we change the time domain information of HRTF instead of changing the spectrum. Subjective hearing experiments show that the proposed method can improve the front-back localization performance.

Keywords: HRTF, enhancement, front-back
I-INCE Classification of Subject Number: 79
incluir el link http://i-ince.org/files/data/classification.pdf

1. INTRODUCTION

The head-related transfer function (HRTF), or the equivalent head-related impulse response (HRIR) in the time domain, describes how a sound is filtered by the head, torso, and pinnae of a listener as it propagates from the source to the listener's eardrum in free space. Traditional, HRTF is used to locate the sound source. Typically, ITD and inter-aural level differences (ILD) which can easily calculate from HRTF are applied to locate the horizontal plane, and the spectrum characteristics exist on the amplitude spectra of the HRTFs are applied to locate the median plane. It is known, HRTF mainly depends on the listener's anthropometric parameters such as head, torso, and pinnae, which reflect, diffract, disperse and shadow the source signal, like fingerprints. They are also personally unique to each listener. Therefore, in virtual reality, people often perform stereo playback through the headphones or loudspeakers.

[1] 1157835731@qq.com
[2] zenggxy@nwpu.edu.cn

In the mid-vertical plane, the time difference (ITD) and the interaural level difference (ILD) are the same. At this time, the method of applying the duplex theory for positioning is no longer feasible. It is generally accepted that interaural cues prescribe information on the horizontal plane while spectral cues help resolve the cone of confusion by giving front-back and elevation perception [1]. There are two critical aspects to HRTFs: they vary according to sound source position and are unique for each. When non-individualised HRTFs are used, the spectral cues may degrade or become unfamiliar to the listener. For non-individual HRTF, Chong-Jin Tan, Woon-Seng Gan[2] proposed a method which consists of fine-tuning the default choice by spectral manipulation and additional filtering of the HRTF. Similar in its fundamental principle to the spectral difference amplification method in [2] Man-ho Park, Song-in Choi [3] proposed the same idea of using weighting function to enhance perceptual differences for the sounds. Compared with [2], [3] analyze the pattern of the spectral difference of two confusion direction HRTFs to solve this problem.

All these methods use spectral difference method to achieve front-back hearing enhancement. However, these methods change the spectrum of the original HRTF signal. By analyzing the correlation between HRTF and physiological parameters, it is known that the physiological parameters of HRTF at (0^o, 0^o) are different from the physiological parameters of influence (180^o, 0^o). At the same time, the correlation between HRTF and physiological parameters before and after the application of spectral difference method was analyzed. The results showed that the application of spectral difference method would change the relevant physiological parameters, which is inconsistent with the actual situation. This method is feasible from the perspective of auditory perception, but in a physical sense, the method cannot be explained clearly.

This paper attempts to propose a new method to achieve front-back enhancement. The patent of the present invention is inspired by the slight rotation of the head which can help achieve azimuth enhancement at the front-back positions. As the minimum orientation perception of a person is 3.6^o in the horizontal orientation, it is difficult for the human ear to clearly distinguish (0^o, 0^o) and (0^o, 1^o). When the orientation is (0^o, 0^o), a human cannot percept the sound in front, instead of hearing at the center of the head. The invention patent knows the ITD of the front or rear azimuth and the nearest azimuth angle and uses the interpolation method to obtain the ITD when the head offset from the front or the back by 1^o, and the original HRTF is modified by the ITD to obtain a new one. HRTF. The HRTF synthesized by this method has no difference in the spectrum from the original HRTF, but the application of the HRTF can enhance the listener's azimuth positioning effect.

2. METHODOLOGY

Since the azimuth enhancement algorithm obtained by the spectral difference method can enhance the sense of orientation of the listener, the spectrum of the anteroposterior orientation obtained by the method has changed the physiological factors affecting the spectrum, which is already related to the HRTF obtained by the measurement - two completely different HRTFs. The invention patent proposes a new orientation enhancement algorithm based on the HRTF spectrum so that the listener can accurately distinguish the signal source.

In this paper, we proposed a method by changing the ITD of the HRTF at the front and back. The specific steps described below:

The rising edge method is used to obtain the ITD of the orientation that needs to be enhanced and the adjacent azimuth angle in the database with the same pitch angle. The ITD describes in equation 1.

$$\mathrm{ITD}_{\mathrm{lead}}(\theta,\phi)=t_{L,\eta}-t_{R,\eta} \tag{1}$$

Where, η indicates the percentage of the maximum value. In this paper, η equals to 10%. $t_{L,\eta}$ and $t_{R,\eta}$ indicates the start time when the HRIR of the left and right ear exceeds 10% of their maximum value at the first time.

θ_0 indicates the orientation that need to be enhanced, while θ_1 indicates the azimuth of the nearest neighbor. In this paper, we obtain the ITD of the orientation of $\theta_0+\Delta\theta$ with the method of linear interpolation. The interpolation equation describes in equation 2.

$$\mathrm{ITD}_{(\theta_0+\Delta\theta,\phi)}=ITD_{(\theta_0,\phi)}\frac{\Delta\theta}{\theta_1-\theta_0}+ITD_{(\theta_1,\phi)}\frac{\theta_1-\Delta\theta\text{-}\theta_0}{\theta_1-\theta_0} \tag{2}$$

Compensate the ITD at the orientation (θ_0,ϕ). After compensation, the ITD of (θ_0,ϕ) varies to $I\widetilde{T}D_{(\theta_0,\phi)}$.

$$I\widetilde{T}D_{(\theta_0,\phi)}=ITD_{(\theta_0,\phi)}+ITD_{(\theta_0+1,\phi)} \tag{3}$$

Transform the HRTF at the orientation of (θ_0,ϕ).

If the nearest neighbor（θ_1,ϕ) locates at the left side of（θ_0,ϕ), pad zero on the left side of the left channel of the HRIR at（θ_0,ϕ). At the same time, to ensure the length of the left channel and the right channel are consistent, the same length of zero is added to the right side of the right channel of the HRIR at（θ_0,ϕ). Detailed formula expression in equation 4.

$$\begin{cases} H\widetilde{R}IR_R(\theta_0,\phi)=[zeros(I\widetilde{T}D_{(\theta_0,\phi)},1);HRIR_{\mathrm{R}}(\theta_0,\phi)] \\ H\widetilde{R}IR_L(\theta_0,\phi)=[HRIR_{\mathrm{R}}(\theta_0,\phi);zeros(I\widetilde{T}D_{(\theta_0,\phi)},1)] \end{cases} \tag{4}$$

3. EXPERIMENTS

In this part, we have designed two sets of experiments, (1) minimum audible angle experiment; (2) front and back enhancement experiment.

Figure 1: Experimental site and equipment

A total of 20 subjects (11 males, and 9 females) aged between 21 and 27 years, were invited for this experiment. The 3D audio signals are generated by MATLAB, which imports into the audio playback software, then output to the headphone power amplifier PRO-XL HA4700 through the sound card, and finally played to the subject via the Sennheiser HD280 Pro high-fidelity headphones. Figure 1 shows the experimental site and equipment.

3.1 Minimum audible angle experiment

As the azimuth interval is not 1^{o} when the elevation is 0^{o} in the CIPIC database, in order to achieve the minimum auralization experiment, the interpolation method is needed to obtain the HRTF with the azimuth interval of 1^{o}.

In this paper, we choose two classical interpolation methods: PLS (Spartial Least Squares, PLS) regression and BPNN (Back Propagation Neural Network, BPNN) regression combined with principal component analysis, to interpolate HRTFs near (0^{o}, 0^{o}) and (0^{o},180^{o}).

The specific experimental process is as follows:

(1) Two rounds of judgment are made for each area with the center point as a reference point. The first round of judgment is performed clockwise from the reference point as the starting point, and the second round of judgment is performed counterclockwise with the reference point as the starting point. Each round of judgment first performs a positive judgment and then performs a reverse judgment.

(2) Each round judges the audio signal of the center point as the reference audio signal, and the audio signal with the angular separation of the center point as the contrast audio signal, and the subject judges whether the reference audio signal and the contrast audio signal are the same respectively, when judging the orientation of the two audio signals The same is considered that the subject of the azimuth interval cannot be discerned. When it is determined that the orientations of the two segments of the audio signal are different, the subject of the azimuth interval is considered to be discernible.

(3) In the forward determination, the reference audio signal and the contrast audio signal spaced from the reference audio signal are the first set of audio signals, and the subject determines that the orientations are the same or different. When the subject is unable to discern the angular interval of the two sets of audio signals, the step size is increased, and the angular interval of the reference audio signal and the contrast audio signal is increased clockwise, so that the subject continues to judge until the subject can distinguish the reference audio signal. And contrast audio signals.

(4) When the subject judges that the orientations are the same, the interval between the two groups of sounds is reduced by the interval, and the subject is again judged to have the same or different sounds. The minimum audible angle of the positive judgment of the listener is finally recorded.

(5) In the reverse judgment, the sound in which the direction of the minimum audible angle is spaced from the center point is determined as the reference audio signal, and the interval is the first group, the step size is increased, and the angular interval is repeated counterclockwise. After the reverse judgment is completed, it is considered that one round of judgment is completed.

(6) The second round of judgment is started from the center point.

(7) Take two rounds of judgment, the mean of a total of 4 judgments as the minimum audible angle of the subject in this area.

Table 1 shows that the fitting effect of BP neural network method combined with the principal component analysis is better than that of partial least squares regression method. The minimum audible angle is 5.2o in front by the method of BPNN regression, and the angle is 9.2o in back with the method of BPNN regression.

Table 1: *The experiment result of minimum audible angle*

Interpolation method	front	back
PLS regression	8.9°	16.8°
BPNN regression	5.2°	9.2°

3.2 Front and back enhancement experiment

The subjects scored the orientation perception of the synthesis signal with original HRTF, enhanced HRTF with the method of [3] and the enhanced HRTF with the method that we proposed. To synthesize the measurement signal, we select white noise, speech and music as the dry signal, especially.

The specific experimental process is as follows:

During the test, the subject is first informed about the actual orientation of the audio to be played, and the subject takes a level scale [5] to evaluate the azimuth of the azimuth audio signal. To characterize the perceived effect of the subject on the orientation, the grade was divided into 5 levels. When the subject's listening orientation is closer to the actual sensed orientation, the azimuth perception is more obvious, which is 5 points. When the subject's listening position is less close to the actual sensed sense of orientation, the azimuth perception is less obvious, which is 1 point. Table 2 is a table of orientation perception levels. Subjects scored the azimuth perception of each method and used the scoring results as a criterion for judging the performance of the method.

For testing, each test signal is 0.5 s in length and is repeated twice each time.

Table 2: *Orientation perception level*

Orientation perception level	Not obvious	A bit obvious	Generally obvious	Very obvious	Extremely obvious
Level	1	2	3	4	5

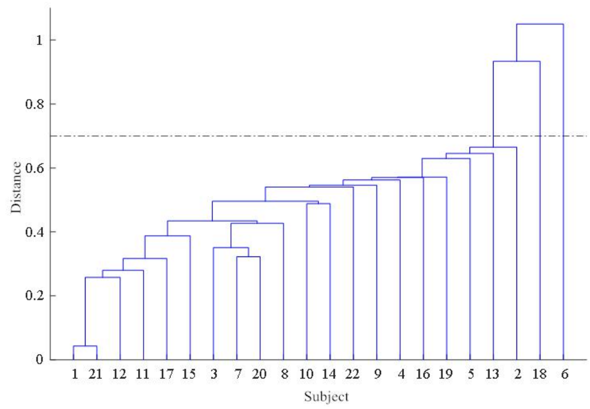

Figure 2: *Subject hierarchical clustering analysis*

In the experiment, 22 evaluation data were obtained, and the experimental data were clustered by hierarchical clustering method to judge the validity of the experimental results. Figure 2 shows the clustering results of all the data of 22 subjects. The ordinate is the distance of each subject's evaluation results. When the distance is smaller, the similarity between the two points is higher, and the distance is higher. Large, indicating that the sample data is not similar to the other samples. Observing the results of the cluster analysis, the evaluation data of subject 6 and subject 18 were far away from the evaluation data of other people, so the score data of the two subjects were excluded to obtain more reliable results.

The evaluation data of the remaining 20 subjects were analyzed. The results of the front-back enhancement experience showed in Figure 3. The enhancement method 1 corresponds to the method in [3], while the enhancement method 2 correspond to the method that we proposed.

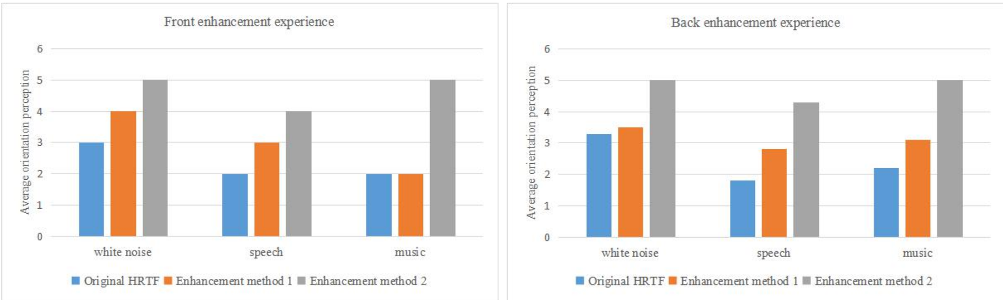

Figure 3: Front and back enhancement experience

Figure 3 shows that the the enhancement methods are better than the original HRTF in average orientation perception, and the enhancement method 2 is the best enhancement method. Since the speech signal only contains low frequency signals, the auditory perception of the speech signal is the worst; and since the white noise is a full-band signal, the auditory perception of white noise is the strongest.

4. CONCLUSIONS

In this paper, we proposed a novel algorithm to improve sound localization performance of front-back. First, we calculate the ITD of the HRTF of the orientation that needs to enhancement and the nearest orientation in the CIPIC database. Then an interpolation method is used to obtain the ITD that has an interval of 1^{o} with the orientation that needs to enhancement. The obtained ITD is used to change the original HRTF.

In subjective experience, we designed two sets of experiments, (1) minimum audible angle experiment; (2) front and back enhancement experiment. The minimum audible angle experience showed that in front the minimum audible angle is 5.2^{o}, and in back the minimum audible angle is 9.2^{o}. The result of the minimum audible angle experience provides theoretical support for the method proposed in this paper, while the front and back enhancement experience verify the accuracy of the method.

So we can conclude if we adopt this algorithm to virtual surround sound system using headphone then we could minimize front-back confusion of existing algorithm and could achieve a high-quality 3D sound.

5. REFERENCES

1. Chong-Jing Tan and Wwn-Seng Can. *"User-defined spectral manipulation of HRTF for improved localization in 3D sound systems"*. ELECTRONICS LETTERS, 34 (25), pp. 2387-2389, 1998.

2. Navarum Gupta, Armando Barreto. *"SpecIra1 Modification of Hear-Related Transfer Functions for Improved Virmal Sound Spatialization"*. ICASSPZOOZ, Vol. 2, pp. 1953-1956, 2002.

3. Ming Zhang, KahChye Tan and M.H.Er. *"A Refined Algorithm of 3D Sound Synthesis"*, ICSP98, pp. 1408-1411, 1998.
4. Man-ho Park, Song-in Choi. *"Improvement of front-back sound localization characteristics in headphone-based 3D sound generation"*. ICACT 2005, pp. 273-276, 2005.
5. Kean Chen. *"Auditory perception and automatic recognition of environmental sound"*. Science Press, 2014.