3D AUDIO WITH HEADPHONES PRODUCED BY DIGITAL SIGNAL PROCESS

Zhibo Song (Boey) 311114059(SID)

Digital Audio Systems, DESC9115, Semester 1 2012
Graduate Program in Audio and Acoustics
Faculty of Architecture, Design and Planning, The University of Sydney

ABSTRACT

This Report showcases a review on using digital signal process to produce 3D audio in headphones. An introduction on the development of 3D audio with headphones is initially outlined. The rationale then focuses on how the hearing system works and how to simulate it using digital ways. By the end of this of this report there is a discussion about the application of this techniques as well as a analysis of its problems.

1. INTRODUCTION

3D techniques on images has become very popular recently, accordingly, the demand of 3D audio is also increasing now. Before, most audio products are designed for the loudspeaker playback system. One ironical factor is that conventional loudspeaker stereo is acceptable on headphones to most people. However, we have great opportunity to change this negative situation now. Within digital signal processors (DSP), it is now possible to create 3D directional sound cues for headphones within a reasonable cost. Besides, thousands of computer games and virtual reality applications have open the commercial application market for 3D audio techniques with headphones.

There are several methods to achieve 3D sound in the headphone listening environment. Two relatively direct ways are using real heads and Dummy heads for recording. However, for situations such like a computer game, which including a huge amount of virtual sound sources, engineers want simply create the localization information by digital process. Through making appropriate time delays and spectral characteristics, there are different methods to simulate natural spatial cues now. Below we will discuss the foundations of hearing system and how to stimulate HRTFs based on these studies following by an analysis of its limitations and conclusions.

2. RATIONALE

2.1. Acoustic and Perceptual foundations

2.1.1. IID and ITD

Researches have shown that two most important cues for localization a sound source are the interaural time differences (ITD) and interaural intensity differences (IID).

1500Hz plays an important role in frequency domain here. Above it, sine waves become smaller than the size of the head. For frequencies greater than 1500Hz, our head will acts as an obstacle, reducing the intensity of waves come to the farther ear. So IID is effective mainly above 1500Hz. Below this frequency, the phase differences produced by ITD will play the main role in determine the position of a sound source.

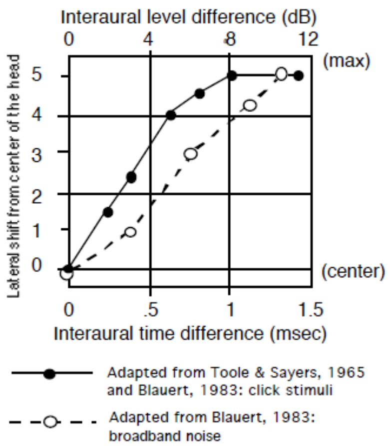

Figure 1 perception of interaural differences within a lateralization paradigm (Begault, 2003)

When two identical (monaural, also termed diotic) sounds been perceived by human ears, the spatial image will located in the center of the head. As IID and ITD increased, the position of the virtual sound will move to the ear. (Figure 1)

2.1.2. Head Movement and Source Movement Cues

In the real life, our heads are never static. Therefore, the head motions can obtain great localization cues. On the other hand, a moving sound can cause dynamic changes for a fixed head as well. One of these changes is Doppler shift, which shift the pitch of a moving sound.

2.1.3. Spectral Cues Provided by the Pinnae

In the below situation shown in figure 2, if we assume that our head is spherical, then the sound source from point A and B will theoretically produce same ITD and IID information and similarly for point x and y. The cone shown in the figure formed by the circle with the center of the head is called the cone of confusion. Although our head cannot be actually symmetry, the ITD and IID cues will be weaken inside this cone area. Instead, the pinna-head-torso system can provide more useful cues here.

Before a sound source reaches the ear drum, its spectrum will be characterised by the shape of outer ear. Thus, a function called head-related transfer function (HRTF) has been introduced to describe this effect. HRTF can be thought as a linear filter for a plane wave coming from a given direction. The magnitude and phase differences caused by the filter will help the listener to point out the sound source even within a similar ITD and IID information.

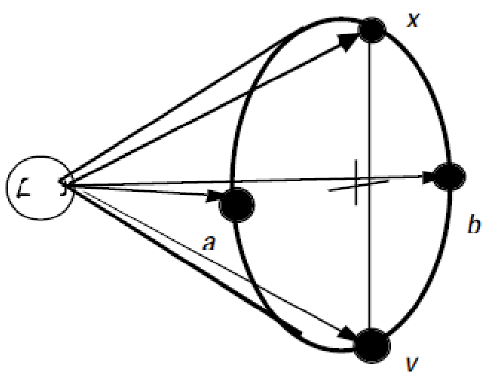

Figure 2 the cone of confusion (Begault, 2003)

2.2. Simulate the spatial cues

2.2.1. Interaural differences

A simple way to manipulating IID is adjusting interaural level difference (ILD) in dB. Feed the two channels with different gain using the data from the previous figure 1 can provide the IID cues, as shown in figure 3 (a). On the other hand, a two channel delay system can create ITD cues as well. Based on extensive research data, both of this two cues can be made into frequency dependent curves and thus we can use formulas to obtained the values we need.

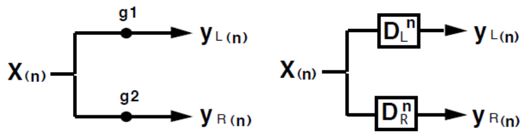

Figure 3. (a) *Figure 3 (b)*

Figure 3 (a) two channel gain scheme (b) two channel delay scheme.

2.2.2. decorrlation

Another method helping move the sound source out of the centre of the head is decorrlation. In natural situations, the signals reaching the ears are decorrelated especially because of room reverberation. To avoid two coherent sound, we can built decorrelator.

There are several ways to achieve the decorrlation. Gerzon proposes linear phase FIR filter with irregular magnitude. Kendall recommends using flat-magnitude decorrelation filters. However, both of the design have limitations in practical use. A third way is using a feed back delay network. The two relatively uncorrelated output can be taken by using two distinct sets of output coefficients and this is also effective within the loudspeaker system in some cases.

2.2.3. Head-related transfer functions

Head-related transfer functions have been measured by several different researchers, a popular one is a group of data measured from KEMAR dummy head. Measurements usually use signals which can easily be deconvolved from the measured response. The result then becomes a set of Head-related Impulse Response(HRIR). Mathematically, X(z) · H(z) in the frequency domain is equivalent to x(n) * h(n) in the time domain (* symbolizes convolution). This is exactly what the HRTF does in the natural hearing. Thus, the HRIRs can be directly used as coefficients of a pair of digital filter.

However, different people have different shape of ears, therefore, HRIRs vary widely. Some reported that using HRIR of someone else can resulting front-back reversals. Besides, the raw catalog of HRIRs are not efficient for implementation. Therefore, a model of the external hearing system is more desirable since it also has the potential for people to understanding the hearing theory.

In order to synthesis the model, two things need to be deal with is data reduction and filter design. Many researches show different methods to deal with the problems happens in this process. Here I will review Brown and Duda's approach of modeling the structural properties of the system pinna-head-torso. (Udo Zölzer, 2002) The model can be divided intro three parts:

- Head Shadow and ITD
- Shoulder Echo
- Pinna Reflections

First, they use a first-order continuous-time system, which is a pole-zero couple in the Laplace complex plane as an approximation of the head shadow effect.

$$s_z = \frac{\omega_0}{\alpha(\theta)} \quad (1)$$

$$s_p = -2\omega_0 \quad (2)$$

Where ω_0 is related to the effective radius a of the head and the speed of sound c by

$$\omega_0 = \frac{c}{a} \quad (3)$$

The position of the zero varies with the azimuth angle θ according to the function

$$\alpha(\theta) = 1.05 + 0.95cos\left(\frac{\theta}{150°}180°\right) \quad (4)$$

The pole-zero couple can be directly translated into a stable IIR digital filter by bilinear transformation, and the resulting filter is

$$H_{hs} = \frac{(\omega_0+\alpha F_s)+(\omega_0+\alpha F_s)z^{-1}}{(\omega_0+F_s)+(\omega_0+F_s)z^{-1}} \quad (5)$$

The ITD can be obtained by means of a first-order allpass filter whose group delay in seconds in the following function of the azimuth angle θ

$$\tau_h(\theta)\begin{cases} -\frac{a}{c}cos\theta \quad if\ 0 \le |\theta| \le \frac{\pi}{2} \\ \frac{a}{c}\left(|\theta| - \frac{\pi}{2}\right) \quad if\ \frac{\pi}{2} \le |\theta| \le \pi \end{cases} \quad (6)$$

Actually, the group delay provided by the allpass filter varies with frequency, but for these purposes such variability can be neglected.

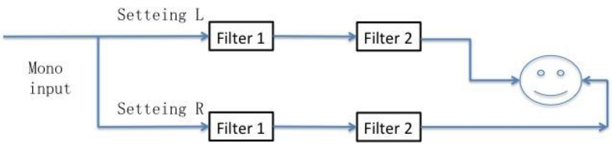

Figure 4 signal process figure of stimulate head shadow and ITD

Second, the approximation of shoulder and torso effects is synthesized in a single echo. An possible expression of the time delay can be

$$\tau_{sh} = 1.2\frac{180°-\theta}{180°}\left\{1 - 0.0004\left[(\emptyset - 80°)\frac{180°}{180°+\theta}\right]^2\right\} \quad (7)$$

Finally, the pinna provides multiple reflections that can be obtained by means of a tapped delay line. A formula for this time delay can be

$$\tau_{pn} = A_n cos(\theta/2)sin[D_n(90° - \emptyset)] + B_n \quad (8)$$

Figure 4 shows the structural model of this pinna-head-torso system. All the 3 blocks of functions will repeat twice for the two ears.

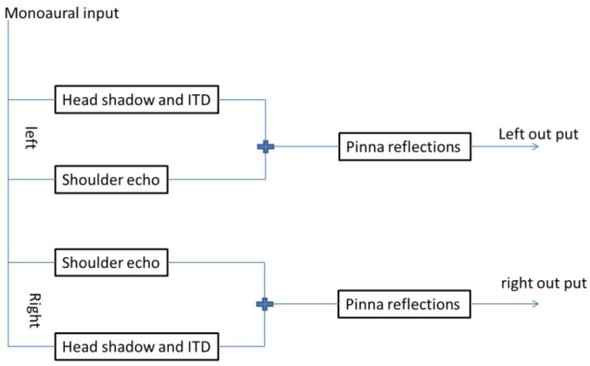

Figure 5 structural model of pinna-head-torso system

3. DISCUSSION

3.1. Problems

Just as the complex structures of our ears, the digital signal process to achieve 3D audio is also complex. One some element can affect the final result. There are some potential problems reported by some researches.

First, as mentioned before, our head never static in the real life, thus the ITD, IID and HRTFs cues are less effective when they are not related to movements of the listener's head. However, the head tracking control is very expensive and cumbersome.

Adding filters to audio signals cannot absolutely avoid coloration to the original sound source. Thus, the more precise simulation of hearing system may need more digital processes, therefore increase the potential of error as well as excessive coloration.

A recent experiment mentioned that our brains are quite insensitive to the HRTF magnitude spectrum, so the externalization mentioned before may actually cause incorrect coupling between the headphones and the ear canal. This hypothesis also leads to the next question: can the 3D DSP work most types of headphones, since the characters vary widely between each other.

3.2. Applications

Fortunately, as the technology developed rapidly, there are more and more applications using 3D headphone sound, therefore accelerate the development in this area. Flight simulators, computer games, virtual reality applications and architectural auralisation are all areas that are benefiting from these developments. Further studies may also benefit the loudspeaker system as we can create virtual loudspeakers using the binaural cues.

4. CONCLUSION

3D audio with headphone is achievable now since the wide using of digital signal processors. However, this field still has potential space for further development. A well understanding about the principles of human hearing system is essential to design the necessary processors for 3D audio, while an efficient simplify of the procedure is also required.

5. REFERENCES

[1] Francis Rumsey (2001) *Spatial Audio*, chapter 3

[2] Udo Zölzer, John Wiley & Sons, (2002), *Digital Audio Effects*, chapter 6

[3] Durand R. Begault (2000) *3-d Sound for Virtual Reality and Multimedia*, chapter 2&4

[4] Wiley (2007) *spatial sudio processing*, chapter 7