Virtual Surround Sound on Headphone

DIWU (440524407)

Digital Audio Systems, DESC9115, Semester 1 2015 Graduate Program in Audio and Acoustics, Faculty of Architecture, Design and Planning, The University of Sydney

1. Abstract

The application of this proposal is to provide compositors or music arrangers with the opportunity to present their composition of different instruments synthesized in different positions especially for orchestra bands so that it could achieve a spatial virtual sound. The mainly purpose of this application is to provide an easy software for compositors to use headphone virtualize the different instruments.

In the reality, people sitting in the concert hall could hear various sounds from instruments in different positions such as fluent comes from left side and goes into left ear first and double bass placed on the right side and goes into right ear earlier. Thus, this sound application could implement that mixing multi-channel sound into just two audio channels that also allow compositors to distinguish the spatial direction of the sources when they are using headphone.

2. Introduction

Human beings extract a lot of information about their environment using their ears. During this process, sound signals are modified by the human peripheral hearing system dependent on spatial cues, processed by the brain to infer most likely position of the sound source. In order to build such virtual surround sound, we need to identify how the human spatial hearing operates.

Firstly, we must be able to understand the influencing factors of acoustic source localization. Auditory organ to determine the spatial location of the sound source is mainly dependent on the comparison of the binaural signal. Interaural Time Differences and Interaural Level Differences play the most important roles in determining the location of a sound source; and second, we have to consider the filtering effect of pinna-head-torso system and how they distinguish the differences of
most of cues from front and back as well as up and down. Beside these reasons, propagation effects and reflections also should be taken account as sound travels through an environment before a listener receives it. Head Related Transfer Functions can be used in recreate portions or all of the listening cues.

2.1 ITD and ILD

Our hearing system use two primary localization cues called interaural intensity difference and interaural time difference to estimate the apparent direction of a sound source.

- **Interaural Intensity Difference**
  
  IID refers to the fact that a sound is louder at one ear that is close to the localization cue because the intensity of sound at this ear will be higher than the other ear (see fig.1).

- **Interaural Time Difference**
  
  ITD refers to the fact that a sound will arrive earlier at one ear than the other due to the distance between sound source and two ears (see fig.2).
In fact, the role of ILD and ITD on the spatial sound source localization is limited because of the inaccurate of distance location and the phenomenon of cone of confusion (see fig.3).

2.2 Head – Related Transfer Functions

Pinna effects can be considered a function of acoustic space filter of our human auditory system, which could change the spectral characteristics of the spatial
direction of the sound. In the acoustic field, the frequency characteristics of human hearing is called Head-Related Transfer Function which contains all the listening cues that are applied to a sound as it travels from the sound’s position, through the environment, and arrives at the listener’s ear drum.

1. Implementation
Here two aspects need to be deal with is data reduction and spatial filter design in order to synthesis the model. Many audio engineerings has developed different approaches to perform this effect in this process. In this sound application development, I will review the method of Brown and Duda to model the structural properties of the system pinna-head-torso, which is mainly divided into three parts (Udo Zölzer, 2002):

- Head Shadow and ITD
- Shoulder Echo
- Pinna Reflections

First of all, considering the shadowing effect, we should assume that the head as a rigid sphere that reflects plane waves. The approximation of showing effect can be illustrated by a first order continuous time system with a pole and a zero in the Laplace complex plane:

\[ S_x = \frac{-2\omega_0}{\alpha(\theta)} \] [1]

\[ S_p = -2\omega_0[2] \]

where \( \omega_0 = c/a \) where \( c \) is the speed of sound in air, \( a \) is the radius of the head represented by a sphere. The position of zero is represented the position of sound in the Laplace plane where it could vary in different axis. The variable of position zero with the azimuth \( \theta \) according to the function:

\[ \alpha(\theta) = 1.05 + 0.95\cos\left(\frac{\theta}{150^\circ} \cdot 180^\circ\right) \] [3]

This can be translated into an IIR digital filter represented by the following function:

\[ H_{hs} = \frac{(\omega_0 + \alpha F_\theta) + (\omega_0 - \alpha F_\theta)z^{-1}}{(\omega_0 + F_\theta) + (\omega_0 - F_\theta)z^{-1}} \] [4]

In order to add the delay to a signal, zeros to the beginning of the signal will be added.
Because of the ITD, they can be obtained by a first order all pass filter. The angle where sound is positioned determines the group delay.

\[
\tau_h(\theta) = \begin{cases} 
-\frac{a}{c}\cos\theta & \text{if } 0 \leq |\theta| < \frac{\pi}{2} \\
\frac{c}{a} |\theta| - \frac{\pi}{2} & \text{if } \frac{\pi}{2} \leq |\theta| < \pi 
\end{cases} [5]
\]

Matlab code (Udo Zölzer, 2002):

%%% For left ear
theta_L = theta_L + 90; % angle adjustment so that range is: -90<theta<90
theta0 = 150;
alfa_min = 0.5;
c = 340; % Speed of sound in air
a = 0.08; % the radius of a human head
w0 = c/a;
% The variation of the zero to express the change of angular position of the
% sound source is given by alfa
alfa = 1 + alfa_min/2 + (1-alfa_min/2)*cos(theta/theta0*pi);
% B gives the numerator of the transfer function
B = [(alfa+w0/fs)/(1+w0/fs),(-alfa+w0/fs)/(1+w0/fs)];
% A gives the denominator of the transfer function
A = [1,-(1-w0/fs)/(1+w0/fs)];
% Group delay is calculated through the function below for the ITD
if (abs(theta_L)<90)
gdelay = -fs/w0*(cos(theta*pi/180)-1);
else
gdelay = fs/w0*((abs(theta)-90)*pi/180+1);
end;
a = (1-gdelay)/(1+gdelay); % the coefficient of allpass filter for ITD
out_magn = filter(B,A,data); % filter to simulate head shadowing effects
output_L = filter([a,1],[1,a],out_magn); % filter for ITD

%%% For Right ear
\[ \theta_R = \theta_R + 90; \quad \% \text{angle adjustment so that range is: } -90 < \theta < 90 \]
\[ \theta_0 = 150; \]
\[ \text{alfa}_\text{min} = 0.5; \]
\[ c = 340; \quad \% \text{Speed of sound in air} \]
\[ a = 0.08; \quad \% \text{the radius of a human head} \]
\[ w_0 = \frac{c}{a}; \]
\[ \% \text{The variation of the zero to express the change of angular position of the sound source is given by alfa} \]
\[ \text{alfa} = 1 + \frac{\text{alfa}_\text{min}}{2} + \left(1 - \frac{\text{alfa}_\text{min}}{2}\right) \cos\left(\frac{\theta}{\theta_0} \pi\right); \]
\[ \% \text{B gives the numerator of the transfer function} \]
\[ B = \left[\left(\text{alfa} + w_0/\text{fs}\right)/(1+w_0/\text{fs}),\left(-\text{alfa} + w_0/\text{fs}\right)/(1+w_0/\text{fs})\right]\];
\[ \% \text{A gives the denominator of the transfer function} \]
\[ A = \left[1,\left(-1-w_0/\text{fs}\right)/(1+w_0/\text{fs})\right]\];
\[ \% \text{Group delay is calculated through the function below for the ITD} \]
\[ \text{if } (\text{abs(}\theta_R)<90) \]
\[ \text{gdelay} = -\text{fs}/w_0*(\cos(\theta*\pi/180)-1); \]
\[ \text{else} \]
\[ \text{gdelay} = \text{fs}/w_0*((\text{abs(}\theta)-90)*\pi/180+1); \]
\[ \text{end}; \]
\[ a = (1-\text{gdelay})/(1+\text{gdelay}); \quad \% \text{the coefficient of allpass filter for ITD} \]
\[ \text{out_magn} = \text{filter}(B,A,\text{data}); \quad \% \text{filter to simulate head shadowing effects} \]
\[ \text{output}_R = \text{filter}([a,1],[1,a],\text{out_magn}); \quad \% \text{filter for ITD} \]

Shoulder echoes:

Second, we can get the shoulder echoes by the time of delay of sound.

\[ \tau_h = 1.2 \frac{180^\circ - \theta}{180^\circ} (1 - 0.00004((\phi - 80^\circ) \frac{180^\circ}{180^\circ + \theta})^2)[6] \]

Where \( \phi \) is the elevation of sound source.
Using the characteristic of FIR filter can implement shoulder echo and pinna reflections.

Pinna reflections:

Finally, multiple reflections provided by the pinna, which can be obtained by a series of delays.

\[ \tau_{pn} = A_n \cos(\theta / 2) \sin(D_n(90^\circ - \phi)) + B_n[7] \]

Where the parameter \( D_n \) is related to the individual characteristics of the pinna.

- Here I measured each instrument in these three steps and finally add them up to achieve a final spatial audio.

2. Evaluation

There are several elements need to be considered in the process.

First, when we measure a virtual 3D sound, the virtual sound sources move as your head turns, so the sound maintains a constant direction relative to your head, rather than remaining at stable positions within the listening environment.

The delay time of the echo generated by the shoulders is assumed in the matlab. Thus, the accurate data need to be more precise simulation of hearing system and may need more digital processes.

3. References


Appendix

Structural model of this pinna-head-torso system for every instrument.