Solution for Spatial Sound Realization in MIDI Specification

Sang-Jin Cho, Alexander Ovcharenko, Jin-wook Chae and Ui-pil Chong
School of Computer Engineering and Information Technology
Univ. of Ulsan., Ulsan

Abstract

Panning is the way in which to realize a spatial sound in MIDI by moving sound images by the loudness of each channel. However, there is a limitation for the natural spatial sound. The HRTF (Head Related Transfer Function) has been widely known as one of the ways to realize spatial sound using the two channels, but it needs much processing power. It is very hard to implement a real time processing structure. In this paper, we propose an improved 3D sound model for the spatial sound location by changing the acoustic parameters. We could get a good result from the experiment with MIDI Pan and our Model.

I. Introduction

The techniques of an efficient 3-D sound system based on the psycho-acoustics of spatial hearing with multimedia or virtual reality have been developed by many researchers [1-5].

HRTF (Head Related Transfer Function) has been widely known as one of the ways to realize spatial sound using just 2 channels. The HRTF filters make it possible to perceive a sound source in any direction. You will be easily able to experience an enhanced 3D sound using a headset and 2 speakers. Yet, the disadvantage of HRTF is that it needs so much processing power.

An alternative way of 3D sound implementation is through PANPOT (Panoramic Potential) in MIDI specification. The PANPOT in MIDI specification can control the loudness of each channel but the sound image is still inside the person's head [6]. For making spatial sound in MIDI environment, we should realize in real time process. So, we propose a new method for realization of spatial sound in MIDI specification in this paper.

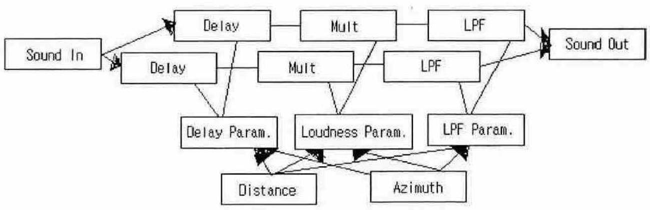

Figure 10 . New model for sound location

II. New Model

We consider the model based on musical acoustics with 3 basic parameters: pitch, loudness, and timber [7]. These parameters are changed as the sound source moves. The following model is suggested for the new model for spatial sound location.

We concern the distance and azimuth only for sound localization in this paper. That is why the elevation is very hard point to control sound.

This model is very simple compared to the HRTF method and is more exquisite compare to the PANPOT method. Now we will report each part of this model. We generate the 3 control parameters, which are delay control parameter, loudness control parameter, and control parameter of low pass filter's cutoff frequency. Each control parameter controls the delay and multiplier changing loudness and low pass filter.

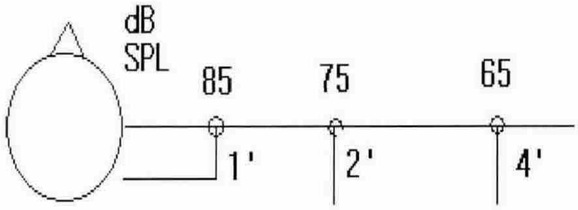

Figure 2. Decay of intensity using loudness scale

1. Delay Control

The average diameter of a human head is 20~30 cm, therefore we must create a time delay to account for each ear. We called it ITD (Inter-aural Time Difference). We can get a maximum time difference from following formula.

0.022 (m) / 340 (m/s) = 0.000065 sec = 65 μs.

When the sound source goes away from the ears, a time delay will be added at 0.065 ms.

Distances (m) / 340 (m/s) = additional time delay

2 Loudness Control

Loudness is similar to ITD for listening to the sound. We call it IID (Inter-aural Intensity Difference). According to widely known experiments, efficient IID is 10dB[8]. What loudness goes to half means distances goes to half by the research results of Stevens and Guiraro in 1962[8].

Using this factor, we made the loudness control block model.

3 Low Pass Filter

A human head acts like a low pass filter. When a sound source is located by the right side of ear, we can hear the direct sound of sound source in the right ear. In this case, the left ear hears a low pass filtered sound (cutoff frequency is 1.5KHz.) by shadow effect[8].

This model is similar to the two previously discussed models. So as the sound source moved, so did the cutoff frequency: from 1.5KHz to 22KHz. Thus, sound without LPF filtering was created.

III. Realization in Commercial Synthesizer K2661

The goal of this paper is implementation of spatial sound in MIDI [9]. So we programmed our model in a commercial synthesizer. (Kurzweil Music Systems' Synthesizer K2661). By using a commercial synthesizer, we can realize our model in real-time and get a stable result.

1. Design a Programming Model.

We realize our model in K2661 as shown on Fig.3.

MIDI control number 10 makes a control value of azimuth, and MIDI control number 28 makes a control value of distance. Then each control value controls a low pass filter's cutoff frequency and amplitude.

2. Realization of Delay Line

Almost every commercial synthesizer updates all sound parameters every 20ms [10], because humans cannot feel a difference in 20ms. (If humans can recognize the difference in 20ms, a musician would not use a digital synthesizer.) But as we show our delay model in 2.1, ITD is just 0.065 ms. We cannot make this delay value in a commercial synthesizer and this value has no meaning in the realization of spatial sound. So we ignore the delay parameter of ITD.

3. Algorithm

For programming in K2661, we use two layers where

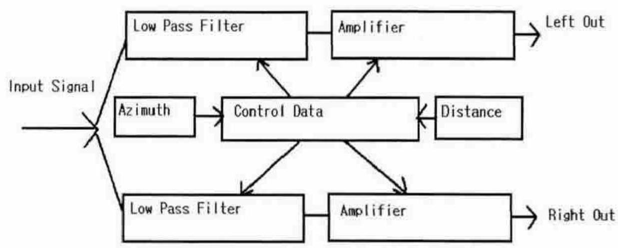

Figure 3. Block Diagram for Model Programming

```
EditProg:F2 FRQ(LOPASS)           <>Layer:1/2
Coarse:G10 25088Hz        Src1 :MIDI 10
Fine   : 0 ct             Depth:-4800ct
                          Src2 : Off
KeyTrk: 0ct/key           DptCtl :Off
VelTrk: 0ct               MinDpt: 0ct
Pad    : 0 dB             MaxDpt: 0ct
<more  F1 OFF  F2 FRQ  F3 OFF  F4 AMP more>
```

```
EditProg:F2 FRQ(LOPASS)           <>Layer:2/2
Coarse:G6 1568Hz          Src1 :MIDI 10
Fine   : 0 ct             Depth:4800ct
                          Src2 : Off
KeyTrk: 0ct/key           DptCtl :Off
VelTrk: 0ct               MinDpt: 0ct
Pad    : 0 dB             MaxDpt: 0ct
<more  F1 OFF  F2 FRQ  F3 OFF  F4 AMP more>
```

Figure 4. Programming display of LPF in K2661

one layer is for left audio out and the other one is for right audio out. Azimuth is assigned by MIDI control number 10. (It's defined by MIDI Implementation). Distance is assigned by MIDI Control Number 28. (It's Undefined yet.)

These two control numbers control low pass filter's cutoff frequency of each layer and the amplitude of each layer.

4. Controlling Low Pass Filter's Cutoff Frequency

We used just one control source. As sound moves left to right or right to left, the low pass filter's cutoff frequency is changed.

We programmed K2661 for realization of 2.3 as shown on Fig.4.

K2661 use a special unit, cent for Filter setting.

Cent is relative unit. 1200 Cent is 1 Octave.

G10 25088Hz means by-passed sound in the condition of 44100Hz Sampling rate. and Depth 4800 cent means 4 octaves .

By programming like this, we can change the cutoff frequency of low pass filter 1568Hz to 25088Hz.

This programming factor simulates the LPF from a human head.

```
EditProg:F4 AMP(FINAL AMP)        <>Layer:1/2
Adjust:6dB                Src1 :MIDI 28
                          Depth:-96dB
                          Src2 :MIDI 10
KeyTrk: 0.00dB/key        DptCtl :ON
VelTrk: 35dB              MinDpt: 5dB
Pad    : 0 dB             MaxDpt: -5dB
<more  F1 OFF  F2 FRQ  F3 OFF  F4 AMP more>
```

```
EditProg:F4 AMP(FINAL AMP)        <>Layer:2/2
Adjust:6dB                Src1 :MIDI 28
                          Depth:-96dB
                          Src2 :MIDI 10
KeyTrk: 0.00dB/key        DptCtl :ON
VelTrk: 35dB              MinDpt: -5dB
Pad    : 0 dB             MaxDpt: 5dB
<more  F1 OFF  F2 FRQ  F3 OFF  F4 AMP more>
Figure 5. Programming Display of amplitude in K2661
```

Figure 5. Programming display of amplitude in K2661

5. Controlling Amplitude

We used 2 control sources. One is MIDI control number 10, the other one is MIDI control number 28. MIDI control number 28 means distance. So when MIDI 28's depth set to -96dB, loudness changes from zero (K2661's -96dB is near to silence) to maximum amplitude. The other control number 10 is assigned to azimuth. When we programmed this parameter, we just set minimum depth and maximum depth based on IID. Scaling is followed by MIDI specification [6].

IV. Conclusions

1. Experimental Results of Our Model.

In our system, using a MIDI control message, we can naturally move a sound image as compare to MIDI Pan (MIDI Pan is realized by just changing the sound's loudness.) in real-time.

We did a comparison experiment with MIDI Pan and our model for 10 people. 8 people felt that the movement of our model's sound image is better than MIDI Pan.

2. Remained Research

As aforementioned, we should do more audience tests for reliability. We will realize externality sound using a hardware MIDI controller in next research. We will put more article for camera ready.

Thanks to KURZWEIL Music Systems for supporting this research.

This work was supported by grant No. R01-2005-000-10671-0 from Korea Science and Engineering Foundation in Ministry of Science & Technology.

References

[1] S. Carlile, "The Physical and Psychophysical Basis of Sound Localization", in Virtual Auditory Space: Generation and Applications, S. Carlile, Ed. Austin, TX: R. G. Landes, 1996

[2] HRTF Measurements of a KEMAR Dummy-Head Microphone MIT Media Lab Perceptual Computing - Technical Report #280.

[3] J. Blauert, "Spatial Hearing. The psychophysics of human sound localization', revised edition, MIT press, Cambridge MA, 1997

[4] D. Begault, 3-D Sound for Virtual Reality and Multimedia, Academic Press, Boston, MA, 1994.

[5] Willian G. Gardner, Transaural 3-D audio, M.I.T Media Lab. Technical Report No.342. 1995.

[6] MMA/AMEI, Default Pan Curve (RP-036)

[7] Hall, Donald E., Musical Acoustics, 2nd Ed, Brooks/Cole Publishing, 1991.

[8] S. H. Kang, Spatial Audio, Keejun Press, 1997.

[9] http://www.midi.org

[10] Hal. Chamberlin , The Musical Applications of Microprocessors, Hayden Book, 1985