Real Time Control of 3D Sound Space by Gesture

Tsutomu Harada, Akio Sato, Shuji Hashimoto, Sadao Ohteru
Department of Applied Physics
Naseda University
3-4-1 Okubo, Shinjuku-ku
Tokyo, 168 JAPAN

ABSTRACT

In most interactive performance systems, the music output's tone and tempo are controlled in real time to allow for artistic improvisation. However, the supposed spatial layout of instruments is fixed during the performance. In the proposed system, we tried to control the position and movement of the sound sources in 3D space, interactively, by the performer's natural gesture. The system consists of Data Glove, the gesture recognition system, the MIDI controller and three speaker pairs placed oppositely on X, Y, and Z axes.

A performer can use the proposed system to control the 3D sound interactively and thus integrate the improvised sound motion into his musical performance.

1. Introduction

Localizations and movements of sounds are often experienced in our daily life. If we can control the sound movements in a music performance, the limits of musical expression seem to be extended enormously[1]. However, at present only a few audio professionals can realize this by their highly advanced skills of musical sound mixing in electric amplified music performance.

For ten years, our laboratory has been engaged in the development of man-machine interfaces for computer music[2][3]. Recently we have developed a 3-D sound system which can control the musical sound space by gestulation with DATA GLOVE in real-time. The system makes it possible to enjoy musical sound localization and its movement without the complex techniques of the sound mixing engineer. In this paper, we introduce our new system with experimental results.

2. System Implementation

The system consists of a MIDI Instrument (E-mu Proteus 200, MIDI sequencer大陸), DataGlove system (VPL), personal computer (NEC

![System Implementation Diagram]

fig. 1 System Implementation
PC-9801), and 6 speakers with amplifiers (fig. 1). The MIDI instrument has 3 pairs of stereo outputs. The DataGlove system measures the bending of a performer's fingers and the position of his hand wearing the glove and sends these data to the computer in real-time. This personal computer recognizes the performer's gestures, calculates balances and amplitude for 3 pairs of speakers, and then produces a sound image at the expected position. These processes are all managed in real-time. For this system we use the algorithm of the gesturalization system which was developed for the articles "A Computer Music System that Follows a Human Conductor[2]" and "Virtual Musical Instrument[3]." This system also recognizes gestures signifying commands to adjust the parameters for system control and to select the instrumental sounds.

3. Sound Localization and Movement

3.1 Sound Localization

First we attempted to control sound movement along a line between a pair of speakers. The imaginary position of sound localized by the speaker pair is affected by a variety of factors, such as the characteristics of the speakers and MIDI instruments, the sound condition of the room and the human auditory system. For example, the MIDI velocity parameter is not linear to the power of speaker output, and the human auditory sense is not linear to the power of sound. To realize localization and movement of a sound image in high fidelity, our system compensates for these effects by making a compensating function (eq. 1).

\[ V = T(P) \quad \text{(eq. 1)} \]

\[ V : \text{MIDI parameter VELOCITY} \]

\[ P : \text{MIDI parameter PAN} \]

We can easily determine this function experimentally using the glove (fig. 2). The performer makes the function by simply drawing a curve with his finger while wearing the DataGlove, while he is listening to the sound and looking at his drawn curve in the display. When he ends up making the function, he has only to signal the 'end of edit' sign to the system (fig. 3).

To realize sound image movement in 3-dimensional space, the system uses the compensating function and an algorithm as

fig. 3 Examples of Control Gestures

fig. 2 Making of the Compensating Function
follows. In this algorithm the direction of sound image, which is pointed out by performer's first finger, is taken to represent a center of gravity \( O \) of 6 point masses.

\[
O = \sum a_i \cdot r_i / \sum a_i
\]

\( r_i \): position vector of \( a_i \)

\( a_i \): \( V_i \)

(eq. 2)

The weight of each point mass \( a_i \) is taken to be the power \( V_i \) of each speaker's output (Fig. 4).

The distance between the performer and the sound image, the position of which is indicated by the performer's finger, is also considered in order to realize 3-dimensional sound localization. The total power of the six speakers' output, represented by the total weight of the point masses, is calculated and compensated for adjusting the distance using the function \( T \) (eq. 3).

\[
\Sigma a_i + \Sigma V_i \cdot T(0) \quad (eq. 3)
\]

The system inversely calculates the amplitude of each speaker from the position of the performer's finger in real-time to localize and move the musical sound image. In our system, spatial resolution of finger positioning in performance space is \( 128 \times 128 \times 128 \), while output resolution of sound localization is \( 32 \times 32 \times 32 \).

3.2 Selecting Instruments through Gestures

With this system instruments can also be selected through an appropriate gesture. After learning the gestures for playing certain instruments through the glove, the system recognizes the player's gestures and selects each instrument, such as a piano, trumpet, saxophone, and so on, all of which are interchangeable in real time. The performance can be enjoyed according to one's musical preferences and can also be "experienced" through the feeling of playing these instruments.

3.3 Gesture Comprehension

At first, the user sets the limits of the left-hand position (up, down, right, left, front and back), velocity (vertical or horizontal), and finger flexing (grasping or stretching), so that the system suits the user's physique. Next it quantifies the movement data from the DataGlove relative to these limits to produce a movement-pattern vector \( S \). This vector consists of the 14 components shown in Table 1. The discriminating function \( KT_i \), expresses the relationship between the movement-pattern vector \( S \) and the meaning vector \( C \), which consists of the 9 components shown in Table 2. \( KT_i \) transforms the more meaningful part of \( S \) into certain positive values, the less meaningful part of \( S \) into certain negative values, and the other part into 0 (fig. 5). This \( KT_i \), for each \( S_i \), which determines the meaning \( C \), can be made up automatically from several samples of the same gesture. The meanings can

![fig.4 The Method using 'Center of Gravity']
be determined with the help of the
discrimination function $R_T$, by the following
formula.

$$ C_j = E R_T (S_j) \quad (eq. 4) $$

If the value of the meaning $C_j$ is maximum and
over a predetermined threshold value, the
computer understands that the user is
indicating the meaning $C_j$. If the meaning $C_j$
expresses the value plane, for example, the
system sends the message to generate a piano
tone as the MIDI signal.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{0j}$</td>
<td>$j$ position</td>
</tr>
<tr>
<td>$S_{1j}$</td>
<td>$1j$</td>
</tr>
<tr>
<td>$S_{2j}$</td>
<td>$V_j$ velocity</td>
</tr>
<tr>
<td>$S_{3j}$</td>
<td>$V_j$</td>
</tr>
<tr>
<td>$S_{4j}$</td>
<td>$V_j$</td>
</tr>
<tr>
<td>$S_{5j}$</td>
<td>$V_j$</td>
</tr>
<tr>
<td>$S_{6}$</td>
<td>$y$</td>
</tr>
<tr>
<td>$S_{7}$</td>
<td>$pitch$</td>
</tr>
<tr>
<td>$S_{8}$</td>
<td>$roll$</td>
</tr>
<tr>
<td>$S_{9}$</td>
<td>$thumb flex$</td>
</tr>
<tr>
<td>$S_{10}$</td>
<td>$index flex$</td>
</tr>
<tr>
<td>$S_{11}$</td>
<td>$middle flex$</td>
</tr>
<tr>
<td>$S_{12}$</td>
<td>$pink flex$</td>
</tr>
<tr>
<td>$S_{13}$</td>
<td>$winkle flex$</td>
</tr>
</tbody>
</table>

Table 1. Movement-pattern vector $S = (S_0, S_1, \ldots, S_n)$

$R_{10}$($S_0$) $R_{11}$($S_1$) $S_0$ $S_1$

fig.3 Examples of Discriminating Function

4. Performance

The performer can control the sound
movement very easily by pointing the direction
in which he wants to locate the sound wave. At
any time in the performance he can select
instruments by means of a playing gesture. The
system also displays a simple picture of the
selected instruments moving in the room as an
animation. All these processes are managed in
1980s. We enjoyed Bach's invention so i with
this system, changing instruments and moving
their positions improvisationally.

If the performer wants to modify the
proposed compensating function, he has only to
indicate so by gesture, and the system enters
the editing mode to modify the table. The
readiness of controlling system can make him
perform freely and creatively.

5. Conclusion

With this system we can control a
3-dimensional virtual musical sound space by
natural gesture - the localization and movement
of the sound image and the selection of
instruments in real time. Furthermore, the
system can control video disc recorder to
display the performance scene of each
instrument to improve the virtual reality.
Musical performers and composers have longed
for: controlling their performance space, and
now they can create music pieces which also
consider movements of sound along with the
progress of Music.

This system can also be applied to
performing plays with high sound presence. For
example, if performers wear certain position
detectors, then the system produces movements
of sound as they move. Real time sound space
control will play a more important part in
musical performance in near future to overcome
the limitations of instrument positioning and
the corresponding sound movement.

6. References

[1] Bost, M. "An Interactive Real-Time System for
the Control of Sound Localization," ICNC
Glasgow 1986 Proceedings, pp.112-114
Computer Music System that Follows a Human
"Jinging & Playing in Musical Virtual Space,"