MULTI-USER HAND GESTURE BASED MUSICAL ELEMENT MAPPING WITH TRADITIONAL MUSICAL ELEMENTS

To Yip Sang, Kam Wong
{yipsto, smkam}@cityu.edu.hk
School of Creative Media, City University of Hong Kong
Tat Chee Avenue
Hong Kong

ABSTRACT

This paper describes the possibility of creating a multi-user instrument by mapping traditional musical elements with hand gesture recognition. Our implementation will be using the “Multi-Touch Surface” as our hand gesture analysis device. Unlike common vision-based musical interface researches which mainly create sound in a generative and random ways, we provide performers with a means of obtaining precise, intuitive and learnable control on the musical elements. The only input is from users’ hands, no any external markers are needed. Complete musical pieces can be performed in a controllable way using our interface.

1. INTRODUCTION

There are thousands of researches in the field of music mapping based on gesture and vision tracking “Audiopad” from James Patten [1]. “Reactable” [2] from Martin Kaltenbrunner’s music research group [3] and “The Manual Input Sessions” Golan Levin [4] are important examples. They mainly focus on mapping specific pre-recorded sound sampling or generating sound in an automatic or rule based environment. Many of them depend on external “markers” which does not exactly provide an intuitive user experience. Each of them owns their art values. But we can observe from these cases that we are lacking a virtual music control interface that provides both precise control and generative methods for performers to play. Discounting for the time being the needs of “non-electronic music” performers such as classical musician, developing that control and generative system for music performance is the primary goal of our research. Our interface enables the operator to use their own hands as input to generate musical elements. Traditional musical pieces can be generated precisely by combining several hand gestures. Multiple input in real time from different performers can also be achieved by our multi-touch system.

2. UTILITIES

2.1 The tracking system

Our interface employs the Diffuse-illumination multi-touch tracking as our hand gesture recognition system. This is an excellent system which helps us screen out ambient lights and cut-out the exact hand-shapes for our analysis (Figure 1).

![Figure 1. Setup diagram of our multi-touch surface. Image from Multi-touch Dev [5].]

2.2 Sound generation system

We utilize foremost the sound library “Audicle” which is also called “Chuck” developed by the audicle team [6] to generate music notes and corresponding chords. In this research, we are generating several music elements and this library meets our needs.

3. IMPLEMENTATION

To combine complex musical elements with only our hands and avoid using overly complicated gesture recognition sets which may make the learning curve of user too steep, we borrow several concepts from drawing. Point-line-plane are the three basic elements from every drawing; they are simple in themselves but by combining them, complex shapes can be formed. Our instruments are divided into those three categories too; the Line In-
strument, the Plane Instrument and the Point Instrument. Each of them covers basic musical elements: Notes, Chords/Harmonics and Pitch.

3.1 Musical Notes: Fingertip gestures (The line instrument)

This instrument generates musical notes by analyzing the movement of fingertips. The tracking starts when users touch the multi-touch surface with their fingertip. There are in total 12 notes in one octave (7 whole notes, 5 semi-notes). We divide a 360 degrees angle into 12 portions as (Figure 2) and each portion represents specific notes in one octave.

![Figure 2](image1.png)

**Figure 2.** Map of movement of fingertips with the note generation table according to their moving paths.

We track each individual fingertip continuously and assign IDs to them as identifiers (Figure 3). There are three actions defined for the movement of those fingertips: Press, Drag, Turn, Release. The “Press” action starts when players begin touching the surface, it measures when will our notes start and provide the volume information for each note (Figure 4). “Drag” action is triggered when users start moving their fingertips (for example if the directional vector is 270, an A note will be generated, and we provide a range of “deviation”. The user does not need to have an exactly 270 degrees angle in order to trigger this note), the movement rate and distance from the starting point is the basis for calculating our music timbre. We then employed the basic theory of mouse gesture recognition to keep track of the position of each fingertips in step-frames, determining the precise moment that they are “Turning”. The “Turn” action is actually when the next notes should be start. Finally, the “Release” action defines when these sequences of notes come to an end. (Table 1. Summarizes all music elements mapping with different fingertip actions)

![Figure 3](image2.png)

**Figure 3.** Assign each fingertip with IDs and keep tracking its movement.

![Figure 4](image3.png)

**Figure 4.** Fingertips gesture analysis.

<table>
<thead>
<tr>
<th>Fingertips Movement</th>
<th>Sound Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size/Pressure</td>
<td>Volume</td>
</tr>
<tr>
<td>Direction Vector</td>
<td>Notes to be generated</td>
</tr>
<tr>
<td>Dragging Speed</td>
<td>Timbre</td>
</tr>
<tr>
<td>Timing between press/release</td>
<td>Note Duration</td>
</tr>
</tbody>
</table>

**Table 1.** Audiovisual mapping for fingertips

3.2 Chords and Harmonics: Fingertip surface ellipse fitting (The plane instrument)

The detection of line instrument starts when the total contact surface area (referred to as blob size) of users’ fingertip reaches to a pre-defined size. We then employ an ellipse fitting procedure [8] to those fingertip surfaces to get the finger orientation vector (Figure 5). The music mapping procedure is very similar and related to the “point instrument”. We define “press”, “drag”, “turn” and “release” for each fingertip and keep tracking their moving paths. We treat it each as a point and a line at the same time. When players move their fingers, they are providing two types of information: the fingertip surface movement direction vector and the fingertip surface orientation vector that is gained by using ellipse fitting.

The fingertip direction vector defines which notes should be generated as usual. The fingertip surface orientation is newly defined and is used to map and decide which “chord”/“harmonics” should be generated. There are 5 most common “chords” defined here in ascending order from minor to Majors (Figure 6). For example, if the player is dragging in the direction to produce a C and its
fingertip surface orientation is pointing to “Major7”, it will generate a “C Major7”.

![Figure 5](image)

**Figure 5.** Getting finger orientation by using ellipse fitting, using the directional vector as chord mapping.

![Figure 6](image)

**Figure 6.** Relationship between chord mapping and fingertips surface orientation vector

<table>
<thead>
<tr>
<th>Fingertip surface</th>
<th>Sound Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size/Pressure</td>
<td>Volume</td>
</tr>
<tr>
<td>Surface Orientation</td>
<td>Chords to be generated</td>
</tr>
<tr>
<td>Direction Vector</td>
<td>Notes to be generated</td>
</tr>
<tr>
<td>Dragging Speed</td>
<td>Timbre</td>
</tr>
<tr>
<td>Timing between press/release</td>
<td>Note Duration</td>
</tr>
</tbody>
</table>

**Table 2.** Audiovisual mapping for fingertip surface

### 3.3 Pitch Controller and Wave Generator: Fingertips (The point instrument)

The function of the point instrument is totally different from the other two instruments. It acts as a controller for the sound, which is generated by the other instruments. At the same time it is also a sound generator. There will be three sets of combination. These are triggered when we detect 3, 4 or 5 fingertips together within a specific distance range. Three fingers represent the “Pitch Controller” for the current sound generated by user. Rotate those three fingers right effects an increase in pitch, while rotating left decrease the current pitch (Figure 7). The combination of 4 (Figure 8) and 5 (Figure 9) fingers represent the “Square Wave” and “Sine Wave” generators respectively. Rotating fingertips to the right adjusts to a higher frequency. Rotating fingertips to the left lowers the frequency. The volume of the sound generated is directly related to the average area width and height towards the fingertips configuration.

![Figure 7](image)

**Figure 7.** The Pitch Controller.

![Figure 8](image)

**Figure 8.** The Square Wave generator

![Figure 9](image)

**Figure 9.** The Sine Wave generator

<table>
<thead>
<tr>
<th>Fingertips number</th>
<th>Fingertips movement</th>
<th>Sound Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 (Pitch Controller)</td>
<td>Rotate to right</td>
<td>Pitch Up</td>
</tr>
<tr>
<td></td>
<td>Rotate to left</td>
<td>Pitch Down</td>
</tr>
<tr>
<td>4 (Square Wave)</td>
<td>Rotate to right</td>
<td>Frequency Up</td>
</tr>
<tr>
<td>Generator</td>
<td>Rotate to left</td>
<td>Frequency Down</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>5 (Sine Wave Generator)</td>
<td>Rotate to right</td>
<td>Frequency Up</td>
</tr>
</tbody>
</table>

Table 3. Audiovisual mapping for fingertip groups

4. CONCLUSION

We presented several methodologies and instrument methods that meet the needs of traditional performers as well as normal users. Providing an intuitive, convenient and easy to learn system for virtual music generation. In this research, we have just covered basic musical elements like: notes, chords/harmonics, pitch and simple wave sound generation. We plan to conduct deeper implementation between musicians and normal users to find out the most efficient user-control interface. In future research progress, we will continue to investigate the possibility of mapping more musical elements to our recognition system and enhance the value of this performance mechanism, such as adding more controls to musical expression. Our aims is for musicians to be able to change the emotion and mood of their musical pieces dynamically. We will also consider enhancing our visual interface so that it can become a more complete audio-visual performance tool.

5. REFERENCES


http://iad.projects.zhdk.ch/multitouch/


Acknowledgments. The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. CityU 121205]