Generate expressive music from picture with a handmade multi-touch music table

1 author:

Simon Lui
Singapore University of Technology and Design

41 PUBLICATIONS   **374** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   An audio content based semantic music search system View project

Project   Teaching and Learning English Pronunciation by Generating the Vocal Tract Shapes from the Frequency Domain Information View project

music scales with a hue color circle. He found the correspondence by comparing the wavelength and frequency of the visual and audio signal respectively. The frequency and wavelength data are as shown in Table 1. Notice that in the table, pitch (cycle per second) multiplied by wavelength (nm per cycle) equal to a constant of 200640 (nm per second). This single formula is not a complete proof of the association, but it shows a certain relationship between color and pitch.

Note	Pitch Frequency (Hz)	Color	Corresponding Wavelength (nm)
C	264	Red	760.0
D	297	Orange	675.6
E	317	Yellow	632.9
F	352	Green	570.0
G	395	Turquoise	507.9
A	440	Blue	456.0
B	495	Purple	405.3

Table 1. Pitch to color association by people with Synesthesia.

Tsang [6] performed an experiment to verify the pitch to color association. He invited participants to first rate the emotional content of several color stimuli, then associate the color with music, and finally associate the emotion with music. He found that the emotional response to color was closely related to the emotional response induced by music. Palmer [7] did a cross modal matches between color and music. He found that fast music in major key was usually related to light, yellow and saturated color, while music in minor key was usually related to de-saturated, darker and blue color. Poast [8] suggested that music emotion provokes color. For example, red refers to exciting music, and blue refers to clam music.

The pitch to color association is useful for many applications. Ciuha [9] applied the association for better visualization of harmony. McDonnell [10] explored the design of color organ with different color assigned to corresponding key. Colored object is a major component in picture. We will make use of this association to generate music from picture.

2.2 Generate music from picture

There are several previous works on generating music from picture. One of the famous examples was the *Augmented Groove* [11]. They used physical card to perform music. The physical motion of the card changed the musical features. The *Music in images* application [12] generated music from picture using a very simple and direct mapping algorithm. It is not interactive. It can only generate one single song from one single picture. Go [13] generated music from picture using a transform system. He first converted a picture into a 1-dimensional array. Then he converted the data into beats and assigned chord to each beat by some defined rules. The *Picture composer* [14] allowed some interaction between the performer and the picture. Performer used a mouse cursor to point and click on a certain pixel to produce sound. There is only one mouse cursor so it can only produce a monophonic melody.

In most previous works, people generated music from the whole picture. We generate music from any part of a picture in user's own

---

ABSTRACT
The multi-touch music table is a novel tabletop tangible interface for expressive musical performance. User touches the picture projected on the table glass surface to perform music. User can click, drag or use various multi-touch gestures with fingers to perform music expressively. The picture color, luminosity, size, finger gesture and pressure determine the music output. The table detects up to 10 finger touches with their touch pressure. We use a glass, a wood stand, a mini projector, a web camera and a computer to construct this music table. Hence this table is highly customizable. The table generates music via a re-interpretation of the artistic components of pictures. It is a cross modal inspiration of music from visual art on a tangible interface.

Author Keywords
Interaction design and software tools, Multimodal expressive interfaces, Musical mapping strategies, Novel controllers and interfaces for musical expression

ACM Classification
H.5.2 [Information Interfaces and Presentation] User Interfaces, H.5.5 [Information Interfaces and Presentation] Sound and Music Computing, I.7.5 [Document and Text Processing] Document Capture --- Graphics recognition and interpretation.

1. INTRODUCTION
This music table allows user to perform music with fingers using various multi-touch gestures. The table displays a picture of user's choice. User touches the picture with fingers. It generates music by associating with the color, the luminosity and the size of the touch object, as well as the position, the gesture and the pressure of the finger touches.

2. DESIGN PRINCIPLE
2.1 The association between pitch and color
Synesthesia is a phenomenon in which one sensory such as music can lead to involuntary experience in a second sensory such as color [1]. Some people with Synesthesia can see a color when they listen to a music note [2]. Synesthesia can even assist some people to achieve the absolute pitch [3]. The pitch to color association of people with synesthesia is quite unique. There are only a few possible combinations. One of the most common associations is as shown in Table 1. Wells [4] proposed a correlation between music and visual color. He obtained a similar result as in Table 1.

Caivano [5] explained the association by comparing different

choice. Our system can generate music with more variety. Hence it can be used as a musical instrument instead of just a transforming machine. We add human sensation to the system and hence it can be used for expressive musical performance.

Picture is similar to music that they both have dynamic changes, melody contour and line contour, and there exists a color and pitch association as discussed in the previous section. However, a picture usually doesn't have a timeline but music is always sequential to time. So a direct mapping that converts everything from picture to music is not practical, since they are somehow different in nature. The main purpose of a picture to music generation system should be either as a novel musical instrument that inspire musical imagination from visual art, or as a tool to inspire idea for music composition. Since music and picture are created with different mindset. The artistic component from visual art may generate novel, stunning and surprising outcome in the musical format.

2.3 Music table

People have been using music tables to perform music. Berry [15] used an augmented reality approach to build a music table. User composed music with the table by placing cards on the table surface. The card changed the sound nature, and the position determined the pitch and timing.

The *Reactable* [16] used a similar concept as Berry's music table. It is now a very popular and successful commercial product. The *Reactable* allowed user to play music by placing objects on its surface. A visual light pulse was generated regularly and it collided with the physical object to produce sound. Stunning visual feedback and music were produced according to the position and orientation of the objects. The objects could be moved and rotated where the music would be changed in real time accordingly.

We would like to construct a music table in a different approach. The previous works used movable object to generate sound. We use finger instead of object since the touch experience is more direct and interactive. Fingers allow more gestures than objects and hence it can give more variety in the music performance.

3. IMPLEMENTATION

3.1 Table design

The multi-touch music table is handmade and hence highly customizable. To construct the table, we placed a glass of dimension 60cm x 45cm with an antiglare film on a wood stand. Then we installed a mini projector and a mirror at the bottom of the wood stand. The projector was connected to a computer. It projected the computer display on the mirror and then reflected on the glass surface. The resolution of the projected picture was limited to 640 x 480 pixel due to the specification of the projector. We installed a mini web camera at the bottom of the wood stand to capture the finger touch information. Figure 1 shows the hardware setup of the table.

We used an open-source visual tracking software *Community Core Vision* (CCV) [17] to capture and process the finger touch information from the glass. CCV can detect the finger touch pressure by measuring the image size of the fingertip on the glass. It required some calibration after the hardware setup. Then we used the *Tangible User Input / Output protocol (TUIO)* [18] to send the finger touch information to the main program via a communication port 3333. Finally the main program analyzed the finger touch and played the music.

There are two major reasons for us to make our own music table. Firstly it is highly customizable. We can create our own touch gestures and adjust the touch sensitivity. Secondly it greatly reduces the cost. We need a huge surface for the best user experience. However huge multi-touch devices are usually very expensive. Our table is not expensive at all. We only used a piece of glass with antiglare film, a wood stand, a mini projector, a web camera and a cheap computer to build it.
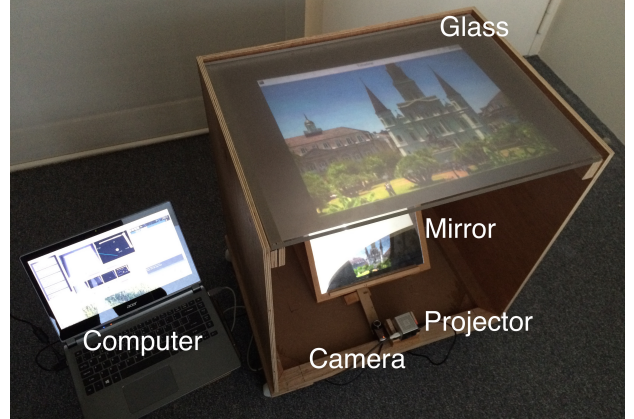

Figure 1. The Multi-touch music table.

3.2 Finger touch design

The visual and musical components are associated according to their corresponding dimension: a 2-D sub-picture is mapped with a chord, a 1-D path is mapped with a melody, and a sequence of 0-D taps is mapped with a rhythmic loop pattern.

3.2.1 The 2nd dimension

We associate a 2-D sub-picture with a music chord. User can select any custom area on the table surface as a sub-picture by using two fingers. This is done by a two-fingers touch and release as shown in Figure 2. A chord is played when both fingers are released.

Figure 2. Play a chord by a two-fingers gesture. When both fingers are released, the red area will be used to generate a chord.

The size of the sub-picture determines the duration of the whole chord. The mapping of size to duration is customizable. Our current mapping is non-linear: a 10cm^2 area refers to 1 second, a 100cm^2 area refers to 2 seconds, and a 1000cm^2 area refers to 3 seconds. From our trial experience, we found that the non-linear approach was more natural, since it did not produce very extreme chord duration like 0.1 second or 10 seconds.

The color of the objects in the sub-picture determines the pitch. We follow the color to pitch mapping in Table 1 by Caivano [5]. Currently we use a chromatic scale with 12 notes in an octave. For example, a C# note refers to the red-orange color since C is red and D is orange. The scale is customizable.

We divide the whole picture (640 x 480 pixel) into 1024 segments of dimension 20x15 pixel. Each segment has a pitch and octave level. The values are pre-calculated during setup. When a sub-picture is created, a histogram of pitch value is plotted and only the top *n* pitch values are used as the chord components. In our current setting, *n* is a random variable of 4 to 8.

The luminosity of segments determines the octave. We allow six levels of octaves. The table plays music notes from C2 (65.41 Hz) to B7 (3951 Hz). A lighter color refers to a higher octave that follows the mapping rules of Palmer [7]. In each sub-picture histogram, an average octave level is calculated for each pitch.

People listen to music sequentially, but they can look at different parts of a picture in any order. A picture usually doesn't contain any sequential data, and hence the music generated from picture should be lack of sequential time information. Here we have the freedom to decide which note should be played first. In our current design, we select the note with the highest ranking in the histogram as the first note. Since this is the base note, we limit its octave to 2 and 3. We set it to 3 if it is 4 or above. Then the next three highest-ranking notes in the histogram follow the first note tightly in time. Finally the other notes are located randomly in the chord. Figure 3 shows an example of how a sub-picture can be converted to a music chord.

|	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
Average octave	3	4	3	4	4	2	5	4	3	3	3	4
Count	12	8	9	10	4	18	1	3	2	6	3	4
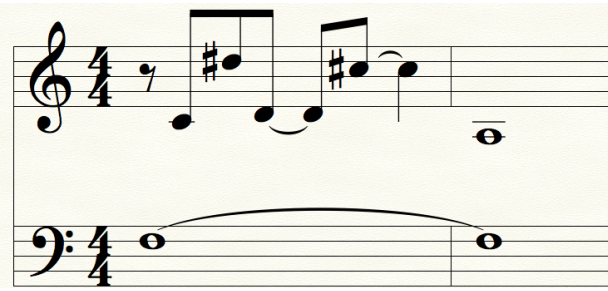
Figure 3. Top: the histogram data of a sub-picture. n=6 (arbitrary), hence F2 is chosen as the first note. C3, D#4 and D3 follow the first note. C#4 and A3 are randomly placed in the chord. Bottom: the resulting chord of duration 2.43s (the sub-picture size is $270cm^2$).

3.2.2 The 1st dimension

We associate a 1-D finger path with a melody. User can use a finger to drag a path to play a single line of melody. The color of each segment along the path determines the pitch of each note in the melody. The luminosity of each segment determines the octave of each note. The finger pressure determines the volume of each note. An example is as shown in Figure 4 and Figure 5. The multi-touch table supports unlimited number of concurrent melody lines. However in order not to mix up with other multi-touch gestures, we limit the number of concurrent melody lines up to 2, but it is customizable.
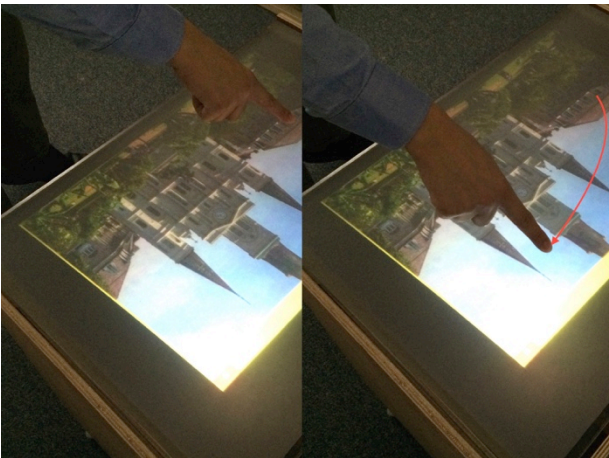
Figure 4. Generate a melody from a path. The segments along the path will generate a melody.
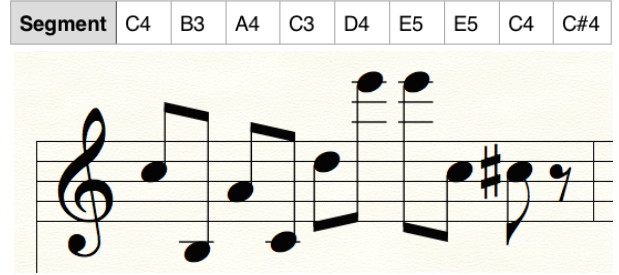
Figure 5. Top: the pitches of segments along the path (extract). Bottom: the melody produced by the path (extract).

3.2.3 The 0th dimension

We associate a sequence of 0-D single taps with a rhythmic pattern. User can tap on any part of the picture in his/her own rhythm. Each single tap must only consist of a quick tap and release without drag or sustain. After a sequence of at least 4 taps, user can press the picture with his/her palm to indicate the end of the pattern. The recorded tap pattern will be played in loop. The user can stop the loop by another palm press. An example is shown in Figure 6.

Figure 6. Generate a loop rhythm from a sequence of taps. When the user taps position 1 to 8 respectively, followed by a palm press, the eight notes will be played in loop.

3.3 Customization

According to Caivano [5], different color maps can be used to represent different music scales. So we customized our own color maps for several particular purposes. For example, we used the Japanese Akebono Scale to perform Japanese style music on a Traditional Japanese painting as shown in Figure 7 and Figure 8. We also used the Major Pentatonic Scale to perform Traditional Chinese style music on a grey scale Chinese brush painting as shown in Figure 9 and Figure 10.

4. CONCLUSION AND FUTURE WORK

The multi-touch music table can inspire novel and unexpected musical knowledge from visual art in a cross modal manner. It is an artistic re-interpretation of visual art. It is not just a direct mapping from visual to audio. It requires expressive input of the performer. The cost to construct a multi-touch music table is very low, which reduces the cost barrier for composer and performer to build their own table.

In this work we assumed that there is no sequential time information in a picture. However, people are usually first attracted by the main object in a picture. Hence there should have some object priority there. According to the principle of compressed sensing [19], information are usually sparse. Perhaps only 10% information in a picture is essential in constructing our visual perception in the first sight. In our next design, we expect to implement a fast and efficient object
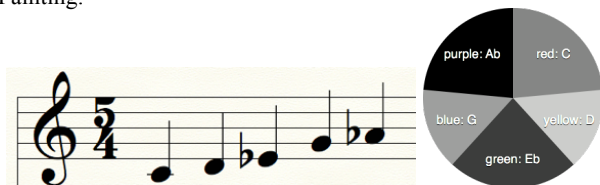
identification algorithm that can run in real time. Then we will ignore the less important component and only focus on the main object when it generates music.

5. ACKNOWLEDGMENT

Figure 7. Perform Japanese music with a Traditional Japanese Painting.

Figure 8. Left: The Japanese Akebono Scale. Right: The hue circle of the Japanese Akebono Scale.

Figure 9. Perform Chinese music with a Traditional Chinese Brush Painting.

Figure 10. Left: the Major Pentatonic Scale. Right: The grey scale level circle of the Major Pentatonic Scale.

6. REFERENCE

[1] Cytowic, R. E. (2002). Synesthesia: A union of the senses. MIT press.

[2] Zamm, A., Schlaug, G., Eagleman, D. M., & Loui, P. (2013). Pathways to seeing music: Enhanced structural connectivity in colored-music synesthesia. Neuroimage, 74, 359-366.

[3] Loui, P., Zamm, A., & Schlaug, G. (2012). Absolute Pitch and Synesthesia: Two Sides of the Same Coin? Shared and Distinct Neural Substrates of Music Listening. In ICMPC: Proceedings/edited by Catherine Stevens...[et al.]. International Conference on Music Perception and Cognition (p. 618). NIH Public Access.

[4] Wells, A. (1980). Music and visual color: A proposed correlation. Leonardo, 101-107.

[5] Caivano, J. L. (1994). Color and Sound: Physical and Psychophysical Relations*. Color Research & Application, 19(2), 126-133.

[6] Tsang, T., & Schloss, K. B. (2010). Associations between Color and Music are Mediated by Emotion and Influenced by Tempo. The Yale Review of Undergraduate Research in Psychology, 82.

[7] Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. Proceedings of the National Academy of Sciences, 110(22), 8836-8841.

[8] Poast, M. (2000). Color music: Visual color notation for musical expression. Leonardo, 33(3), 215-221.

[9] Ciuha, P., Klemenc, B., & Solina, F. (2010, October). Visualization of concurrent tones in music with colours. In Proceedings of the international conference on Multimedia (pp. 1677-1680). ACM.

[10] McDonnell, M. (2007). Visual music. Visual Music Marathon, Boston Cyberarts Festival Programme.

[11] Poupyrev, I., Berry, R., Kurumisawa, J., Nakao, K., Billinghurst, M., Airola, C., ... & Baldwin, L. (2000). Augmented groove: Collaborative jamming in augmented reality. In ACM SIGGRAPH 2000 Conference Abstracts and Applications (p. 77)

[12] http://www.musicinimages.com

[13] Go, G., Song, T., Jung, S., Jeong, H., Kim, M., Kim, M., ... & Jeon, J. (2010, October). Generating music using a picture and a transform system. In Control Automation and Systems (ICCAS), 2010 International Conference on (pp. 2285-2289). IEEE.

[14] http://www.tophersaunders.com/wp/picture-composer-generate-music-with-images/

[15] Berry, R., Makino, M., Hikawa, N., & Suzuki, M. (2003, October). The augmented composer project: The music table. In Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality (p. 338). IEEE Computer Society.

[16] Jordà, S., Geiger, G., Alonso, M., & Kaltenbrunner, M. (2007, February). The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In Proceedings of the 1st international conference on Tangible and embedded interaction (pp. 139-146). ACM.

[17] http://ccv.nuigroup.com/

[18] Kaltenbrunner, M., Bovermann, T., Bencina, R., & Costanza, E. (2005, May). TUIO: A protocol for table-top tangible user interfaces. In Proc. of the The 6th Int'l Workshop on Gesture in Human-Computer Interaction and Simulation.

[19] Donoho, D. L. (2006). Compressed sensing. Information Theory, IEEE Transactions on, 52(4), 1289-13