

# Computer Assisted Melo-rhythmic Generation of Traditional Chinese Music from Ink Brush Calligraphy

Will W. W. Tang, Stephen Chan, Grace Ngai and Hong-va Leong  
Department of Computing, The Hong Kong Polytechnic University  
Kowloon, Hong Kong  
{cswwtang, csschan, csgngai, cshleong}@comp.polyu.edu.hk

## ABSTRACT

CalliMusic, is a system developed for users to generate traditional Chinese music by writing Chinese ink brush calligraphy, turning the long-believed strong linkage between the two art forms with rich histories into reality. In addition to traditional calligraphy writing instruments (brush, ink and paper), a camera is the only addition needed to convert the motion of the ink brush into musical notes through a variety of mappings such as human-inspired, statistical and a hybrid. The design of the system, including details of each mapping and research issues encountered are discussed. A user study of system performance suggests that the result is quite encouraging. The technique is, obviously, applicable to other related art forms with a wide range of applications.

## Keywords

Chinese Calligraphy, Chinese Music, Assisted Music Generation

## 1. INTRODUCTION

Art can be considered a language of communication, while different art forms can be considered different modes, or languages, of communication. Human languages can normally be translated into one another, even though there are often words and concepts that are difficult to translate. By analogy, it is believed that it is also possible to translate, or at least transform, one art form into another. The difference between art forms is obviously far greater than the difference between human languages. Nevertheless, it is believed that certain elements of art forms can be transformed.

The rapid growth of the Internet in China has spurred on research interest in the computerization of traditional Chinese art forms. Previous work [3],[7] have demonstrated the possibility of transforming Chinese ink brush calligraphy into music, with varying degrees of complexity. This paper reports on a two-part transformation which advances the technology to the next level of complexity. The melodic part transforms the ink brush strokes into notes of different pitches, making use of the type of the stroke as well as a statistical model of note sequences of a popular style of Chinese melodies. The rhythmic part, on the other hand, transforms the timing information in the stroke making (speed, duration, gaps, etc.) into the rhythm of the music thus generated. The aim of this (“melo-rhythmic”)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*NIME'13*, May 27-30, 2013, KAIST, Daejeon, Korea.

Copyright remains with the author(s).

transformation is to transform Chinese ink brush calligraphy into music, while preserving as much as possible the physical and artistic properties of the original art form. The transformation is ultimately based on the concept of metaphoric transformation - the matching of simple metaphors such as up, down, fast and slow between ink brush calligraphy and music. The end result is that a user can generate traditional Chinese music by writing ink brush calligraphy in the traditional way. The system is usable for someone who knows nothing about computers, such as some of the elderly Chinese artists, and children.

We designed CalliMusic with the following projections. (1) The same person who writes the same sentence in the same style will generate essentially the same piece of music. (2) The same person who writes the same sentence in different styles will generate the same set of notes in different rhythms since the timing of the stroke-making will be different in different calligraphic styles. (3) The same sentence written in the same style by different persons will also generate the same set of notes but in different rhythms, since different people write differently even if they are writing in the same calligraphic style. Through our user studies, we have validated these projections. Hence at least some of the more obvious physical-artistic elements in Chinese Ink Brush Calligraphy are successfully transformed into music. We are now extending the study into other more subtle elements. The metaphoric transformation will then extend into more nuanced metaphors such as being calm, excited, strong and soft.

## 2. RELATED WORK

Computer-generated music is not new. A wide range of techniques have been used already, with varying degrees of success.

Probability calculus and statistics have been used in the field of automatic music generation early on. [6] Markov model [1] and the n-gram technique borrowed from natural language processing [9] were applied to generate music statistical model with success.

Generating music from Chinese ink brush strokes (or writing in general), on the other hand, is quite new, and there have been few related work. DrawSound [5] generates sound frequencies with the writing of ordinary paper and conductive brush on top of a multi-touch input surface. However, it uses a special brush which is not designed for traditional Chinese calligraphy. The Hé system [7] takes an approach somewhat closer to what we do in that it generates musical notes with ink brush painting. However, it is designed to accept arbitrary brush movements and does not take the stroke type or the character into account. Likewise, the MelodicBrush system [3],[4] generates musical notes in real-time and correlated to brush strokes made while writing Chinese characters. However, that system is somewhat constrained by the real-time nature of the system and thus the music generated is often considered fairly mechanical.

### 3. SYSTEM OVERVIEW

CalliMusic captures user input in the form of brush motion in traditional Chinese calligraphy writing, without affecting the way the calligraphy is written - where a person uses a real brush with real ink to paint on a piece of paper. After the user finishes writing, the system will analyze the whole writing and generate a piece of Chinese music automatically where the pitch and rhythm are determined by the written characters and writing intervals.

#### 3.1 Writing Platform

Chinese calligraphy is traditionally written on parchment, with a soft hairy brush and ground ink. It is an art form with a very rich and long history of cultural significance. Artists feel it can be used to express a wide range of moods and emotions ranging from formal seriousness, to muscularly powerful, to delicate articulation, ..., to unrestrained freedom. Anecdotally, it is even believed that one's personal character can be reflected through one's calligraphy. Hence one of our objectives is to capture as much as possible the "feeling", or affective elements in the calligraphy through the mechanics and rhythm of writing. Given that the paper or brush cannot be replaced by a touch surface or electronic brush, we follow previous work [4] in using a vision-based device to capture the motion of writing. A Microsoft Kinect depth camera placed in front of the user (Figure 1) captures the location and movement of the brush. Combining this information with the location of the paper, the types of strokes and writing time intervals can be obtained.

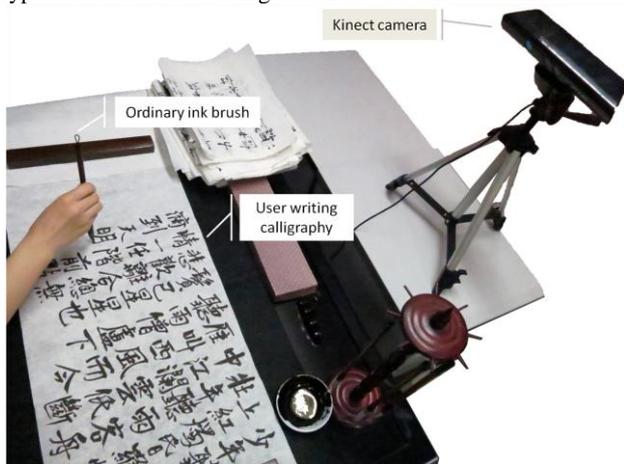


Figure 1 Writing platform: user writing on the paper, and a Kinect camera in front of the desk capturing the motion.

#### 3.2 Transformation

In this project we explored a variety of different mappings both in melody and rhythm. Each of the mappings take in the user writing information as input parameters, and output note pitches for the melodic mapping and note lengths for the rhythmic one. A total of 3 melodic mappings and 3 rhythmic mappings are studied and presented.

#### 3.3 Music Generation

Generated melody and rhythm will be converted to GNU LilyPond file format [2], along with other information such as the characters and stroke types that contribute to a particular segment of music. From the file, a music score sheet can be automatically engraved that can be studied and be played by musicians using Chinese musical instruments. The powerful customization of the score sheet format allows us to study various mapping easily (Figure 2). More importantly, a MIDI file can also be generated from that LilyPond file, which can be

further synthesized using various synthesizers to produce realistic traditional Chinese music.

Our system converts the user's writing to music as it is written. However, since a best-fit optimization step is involved and the syntax of the written text is considered in the mapping, the generation is in semi-real-time - in that the user needs to reach a suitable "pause point" (e.g. the end of a sentence, or the end of a paragraph) before the music is generated and played.



Figure 2 Score sheet generated by LilyPond for the study of differences between mappings. Different colors indicate different mappings being used to generate that note.

### 4. MAPPING WRITING TO MUSIC

The three main components of a piece of music are pitch, rhythm and harmony. Traditional Chinese music did not have a strong concept of chord and not much in the form of harmony. Usually, one musical instrument or many musical instruments of the same type will take the lead and dominate a musical piece. Therefore, our system considers only the pitch and rhythm components. To gain a deeper understanding of the effect of each individual part, these two parts will be processed independently in the beginning of this project and integrated together subsequently.

#### 4.1 Traditional Chinese Music

Most traditional Chinese music uses a pentatonic system, i.e. a scale with five notes in an octave. Traditionally these five notes are named as 宫(Gong), 商(Shang), 角(Jue), 徵(Zhi) and 羽(Yu), which correspond to the notes Do, Re, Mi, Sol and La in Western music. Traditional Chinese music does not have harmony, so there are no simultaneous notes being played at any given time. In other words, all notes are played one by one in series.

In this work, we investigated 120 pieces of traditional Chinese music that were composed from 1027BC (Zhou Dynasty) to 1911AD (Qing Dynasty), ranging from court music to military music to folk music. We found that the rhythm of traditional Chinese music is very simple: tied notes or complex beat, such as triplets, were very rare in the corpus. We therefore postulate that simple beat representation is sufficient for notation. Also, most of the pieces are in basic time signature of 4/4 or 2/4, with a few exceptions of other complex time signatures such as 10/4. Since there is no change of time signature in the middle of any piece in our corpus, most pieces can be rearranged to 4/4 for the whole song.

#### 4.2 Generating the Pitch

In this project, we researched on three mappings for the generation of pitch. To afford human control, we first create a human-inspired mapping that is based on investigations from previous work [3]. Then we test a statistical mapping for a more naturally pleasing sound. Finally, a hybrid mapping is developed by complementing human control with statistical input. The result is quite promising.

### 4.2.1 Human-Inspired Mapping

Previous work [3] has investigated the likelihood of a metaphoric transference from Chinese calligraphic strokes to musical notes. The 39 types of strokes from Chinese calligraphy writing were categorized into 5 major classes as seen in one of the most commonly used Chinese input methods for computerization, 五笔(Wubi, literally meaning five strokes). The results indicated that there is a consistent and natural (although not universal) correlation between stroke type and pitch according to human perception. Hence, when a user writes a character with a sequence of strokes, a sequence of notes will be generated accordingly. For example, Figure 3 shows the music sequence that is generated when the character 木 (wood) is written.



**Figure 3 Musical sequence generated in response to writing the character 木 (wood).**

The advantage of this human-inspired mapping is that it takes into consideration the correspondence between the perception of a calligraphic stroke and the pitch of a note. For example, some strokes give the perception of stability, and these are perceived to correspond with stable note pitches.

However, there is a major shortcoming to this mapping. Although the user can directly control the generation of pitches, the generated music may not sound appealing if the music is generated strictly from a piece of written prose, or even poetry, since the writer cannot completely control the writing at the level of individual strokes – the sequence of strokes is dictated by the characters chosen. Even if the writer is allowed to write strokes without regard to whether the strokes form actual characters, probably only very adept musicians can generate pleasing music in this manner.

Another, relatively minor, shortcoming is that the whole song will stay in one octave if no further parameter is involved. To address this issue, we extend previous work in allowing the musical sequence to traverse octave boundaries if it would minimize the number of notes traversed between the current note and the following note. Constraints are applied to avoid notes deviating too far away from the octave of middle C that will not please the human ear.

### 4.2.2 Statistical Mapping

To produce more Chinese-sounding music, we develop a statistical model trained on a corpus of traditional Chinese music to recreate a musical pitch piece that carries the same probabilistic characteristics of traditional Chinese music. Our corpus consists of 120 pieces of well-known or typical traditional Chinese music that were composed between 1027BC and 1911AD that are preprocessed by transposing the scale to C Major, fixing special markings on the notes and normalizing the time signature.

Our statistical model uses the conventional trigram model. For each stroke  $c$ , a note with pitch  $p$  is picked to maximize

$$P(p_c|p_{c-1}p_{c-2})$$

The choice of trigrams rather than longer sequences was made to capture the characteristics of the music without memorizing actual portions. Smoothing by backing off to an interpolated bigram-unigram combination is used when the sequence of the

previous two notes cannot be found as the first two notes of a trigram in the corpus. [8]

The issue of this mapping is that the generated music bears no obviously discernable relationship to the writing style or written strokes, save for the number of notes generated. It is simply automatic music generation based on a statistical model. By itself, it cannot achieve our objectives.



**Figure 4 A sample score generated by statistical pitch mapping with rhythm extracted from a classical song.**

### 4.2.3 Hybrid Mapping

Direct human control affords human control but does not guarantee pleasing music. Statistical mapping generates pleasing music but does not allow human control. Hence either human-inspired mapping or statistical mapping alone cannot provide what we need. But combined they can. Hence, a back-tracking algorithm has been developed to maximize the likeliness of n-gram probability while still respecting the written stroke types. The algorithm goes through each possible path of note sequences presented in the corpus, remembering the most likely path as a back-track path for each run. The result is that the user can, to some degree, control the generation of pitch sequences, while at the same time the computer helps her to generate pleasing sequences of pitches.

In statistical mapping, the generation of pitch is a probability of previous pitches. In human-inspired mapping, the relationship between stroke and pitch is also a probability. So combining these probabilities together we get the following probability to be used to generate pitch.

For each stroke  $c$  of type  $t$ , a note with pitch  $p$  is generated to maximize:

$$\text{Max}(P(p_c|t_c)P(p_c|p_{c-1}p_{c-2}))$$

where  $P(p_c|t_c)$  is the most likely pitch that is correlated with the stroke type according to human perception, and  $P(p_c|p_{c-1}p_{c-2})$  gives the most likely pitch according to our trigram model.

Our model thus takes into account both human perception as well as musical theory, as evidenced by the patterns that are commonly found in real musical pieces.



**Figure 5 Score generated by hybrid melodic mapping. The text (怒) on top is the character being written while the texts on bottom correspondes to the stroke types of that character. Note that different pitches may be generated even with same stroke type Na, but the same pitch for the first and third Zhe.**

## 4.3 Generating the Rhythm

Two studies of rhythmic mapping were carried out. First we used a statistical mapping to generate rhythm. For human-inspired mapping, we tried to capture the writing mechanics presented in calligraphy, along with the syntax of the content being written.



**Table 3 Example score table for the character 青.**

Character 青 (green)	Stroke s								
	0	一	一	丨	一	ノ	フ	一	一
$w_{1,s}$	/	93	94	203	219	125	578	109	93
$g_{1,s}$	172	172	156	125	172	172	188	141	141
$w_{1,s}/t_1$	/	3.52%	3.56%	7.69%	8.30%	4.73%	21.89%	4.13%	3.52%
$w_{1,s}/e_1$	/	6.14%	6.21%	13.41%	14.46%	8.26%	38.18%	7.20%	6.14%
$g_{c,s-1} + w_{c,s} + g_{c,s}/t_1$	/	16.55%	15.98%	18.33%	19.55%	17.77%	35.53%	16.59%	14.20%

$$\therefore R_{c,s}(31.25\%) = 0.1 \times 1.5 = 0.15$$

$$Z_{1,1}(31.25\%) = E_{1,1}(31.25\%) + I_{1,1}(31.25\%) + R_{1,1}(q) = 0.2204 + 0.15 + 0.15 = 0.5204$$

**Heuristic Fit Algorithm:**

Input: Current character  $c$ , Number of stroke  $s$ ,  
Total Note Length  $l$ , Stroke writing time  $w_{c,s}$ ,  
Total writing time  $e_c$

1. Set note lengths  $\vec{p}$  to  $[q_1 \ q_2 \ \dots \ q_s]$  where  $q_i = \lfloor \frac{w_{c,s}}{e_i} \rfloor$
2. Set *difference* to  $l - \sum_{i=0}^s q_i$
3. While *difference* is greater than 0
4. Repeat for  $i = 1$  to  $s$
5. Set  $p_i$  to  $q_i + \text{min\_note\_length}$
6. Set  $score_i$  to  $Z_{c,i}(p_i)$
7. Find the smallest  $score$  and set  $j$  to be its index
8. Set  $q_j$  in  $\vec{p}$  to  $q_j + \text{min\_note\_length}$
9.  $\vec{p}$  is output as the sub-optimized note lengths

**Algorithm 1 Heuristic Fit Algorithm.**

**4.3.2.4 Example**

Referring to Table 3, suppose we want to calculate the score for character written as  $c = 1$ ,  $n_1 = 8, t_1 = 2640, e_1 = 1514, b_1 = 1.5$ .

$$E_{1,1}(q) = \left| \frac{w_{1,1}}{e_1} \times b_1 - q \right| = \left| \frac{93}{2640} \times 1.5 - q \right| = |0.0921 - q|$$

For  $q = 6.25\% = 0.0625$ ,  
 $E_{1,1}(6.25\%) = |0.0921 - 0.0625| = 0.0296$

Since  $q$  can be represented as one sixteenth note,  $I_{1,1}(6.25\%) = 0$ .

Also  $\frac{w_{1,1}}{t_1} \times b_1 = 0.0528$  and  $\frac{g_{1,0} + w_{1,1} + g_{1,1}}{t_1} \times b_1 = 0.2483$ , and  $0.0528 < 0.0625 < 0.2483$ ,

$$\begin{aligned} \therefore R_{c,s}(6.25\%) &= 0 \\ Z_{1,1}(6.25\%) &= E_{1,1}(6.25\%) + I_{1,1}(6.25\%) + R_{1,1}(q) \\ &= 0.0296 + 0 + 0 = 0.0296 \end{aligned}$$

For  $q = 31.25\% = 0.3125$ ,  
 $E_{1,1}(31.25\%) = |0.0921 - 0.3125| = 0.2204$

Since  $q$  can be represented at best as one quarter note plus one sixteenth,

$$\therefore I_{1,1}(31.25\%) = 0.1 \times 1.5 = 0.15$$

Also  $\frac{w_{1,1}}{t_1} \times b_1 = 0.0528$  and  $\frac{g_{1,0} + w_{1,1} + g_{1,1}}{t_1} \times b_1 = 0.2483$ , and  $0.0528 < 0.3125 > 0.2483$ ,

**4.3.2.5 Sentence-level Optimization**

Apart from character-level optimization, our system takes into consideration the syntactic function of the characters. Since Chinese is written without spaces between characters, word segmentation is carried out to group characters into words. Using an automatic Chinese Part-Of-Speech (POS) parser and tagger, each word is then grouped and tagged with its POS  $G_c$ , e.g. ADJ, VERB or NOUN.

Sentence Confine Penalty  $A_c(q)$  is defined as

$$A_c(q) = \begin{cases} \omega_c, & \text{Beginning of a word and} \\ & (G_{c-1} = \text{ADJ and } G_c = \text{NOUN}) \\ & (G_{c-1} = \text{ADV and } G_c = \text{VERB}) \\ \omega_f, & \text{Beginning of a word and} \\ & (G_{c-1} = \text{VERB and } G_c = \text{NOUN}) \\ \omega_b, & \text{Beginning of a word but no in previous cases} \\ 0, & \text{Not the beginning of a word} \\ \omega_n, & \text{otherwise} \end{cases}$$

Where we set  $\omega_c = 0.1, \omega_f = 0.15, \omega_b = 0.2, \omega_n = 1$ .

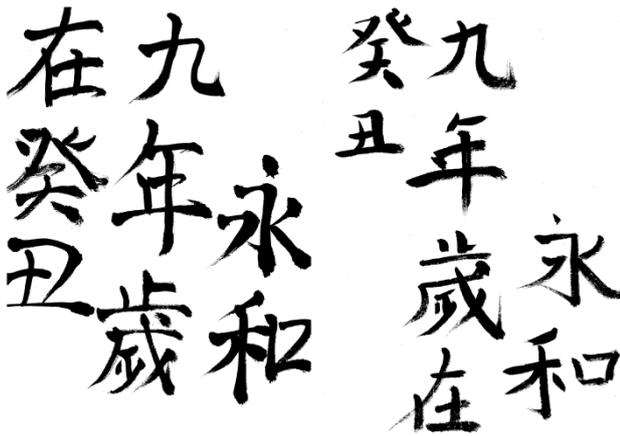
So instead of making the notes fit into the proposed note duration strictly for each character, we consider all possible lengths and try to concatenate these characters with the least overall Sentence Confine Penalty. A similar heuristic will be applied to get the sentence-level sub-optimized best possible fit note lengths.

**5. PERFORMANCE EVALUATION**

**Table 4 A summary of satisfaction level of all mapping combinations.**

Satisfaction		Pitch Mapping			
		Fixed	Statistical	Human Inspired	Hybrid
Rhythm Mapping	Fixed	(N/A)	Okay	Okay	Good
	Statistical	Okay	Good	Okay	Good
	Human Inspired	Quite good	Good	Good	Very good

We tested all the combinations of each melodic mapping with each rhythmic mapping, to determine the combinations that give the best results. To study how a melodic mapping performs independently from the rhythmic mapping, and vice versa, a fixed rhythm or melody part extracted from an original music piece is used. Table 4 is a summary of satisfaction levels of all mapping combinations reported by 4 users with basic to expert musical training. Not all mappings generated music pleasing to users. Statistical mappings provided a reasonably good result. The proposed hybrid pitch mapping and human inspired rhythm mapping received the best level of satisfaction, as expected.



**Figure 7** Two pieces of writing (portion only) by two subjects in the same writing style. On the left, (1) a more accomplished calligrapher; on the right, (2) a novice. (not really an expert).

The projections made when we designed CalliMusic have also been validated in the user test. We focus on the hybrid pitch mapping and human-inspired rhythm mappings since these give the user control over the music generation. (1) The same user writing the same sentence in the same style generated essentially the same piece of music, especially for expert calligraphers whose writing are more stable than those of novices. (2) The same person who writes the same sentence in different style - in this case we tested 楷書(KaiShu, regular) and 行書(XingShu, semi-cursive) styles - generated essentially the same set of pitches but different sets of rhythms. Only in some rare cases among expert users where their XingShu style simplified and changed the sequence of stroke, the set of generated pitches were different. (3) Different people writing the same sentence in the same style generated essentially the same set of pitches but different sets of rhythm. We observed that novice users tend to think more before actually writing on the paper, so the music rhythms they generated are slower and more regular than those by expert users.

蘭亭序

Subject 1

Subject 2

**Figure 8** Two according pieces of music (portion only) generated by these subjects using hybrid rhythmic mapping and human-inspired melodic mapping. Note that the work of subject 1 shows more rhythmic features than that of subject 2.

To illustrate the above-mentioned projection 3, two writings of the same sentence by two different subjects in the same style are shown in Figure 7. The subjects were asked to write a well-known classical calligraphy work Lantingji Xu by Wang Xizhi, written in semi-cursive style and composed in year 353. They both copied from the same work placed in front of their desk, so the stroke size and order were all the same. As a result, the

sets of pitch generated were the same. However, subject 1, the more accomplished calligrapher, wrote his work faster and with more complex rhythm throughout the writing of each stroke, showing more confidence. More nuanced details can be found in the starting and ending of every stroke. On the other hand, subject 2, the novice, wrote it slower and with more constant speed for each stroke, pausing and thinking for a longer time between strokes. Therefore the rhythms of their generated music were different in that the music of the first subject is more lively and pleasing to ear, while the music of the second one is more formal and boring. (Figure 8)

## 6. CONCLUSION AND FUTURE WORK

A new musical instrument that can transform the user performance in Chinese calligraphy into music is developed. Even a user who does not know music can generate reasonably pleasing music through different proposed mappings, as long as she can write Chinese calligraphy. Just like many other sophisticated art forms, it takes a lot to become an expert Chinese calligrapher. But it is relatively easy to learn to use the Chinese ink brush to write basic Chinese calligraphy. Hence CalliMusic is accessible to a wide range of users, from novices to experts.

Some parameters used in the mappings are set through heuristic processes. More complete experiments are planned in order to improve the mapping quality through parameter tuning.

This project focuses on the physical and syntactical aspects of Chinese calligraphy and music. Taking the semantics into account can result in better mappings in terms of meaning and emotional expression. The techniques are, of course, applicable to other forms of writing, and other types of music. Applications in performance arts, therapeutic and rehabilitative domains are also being explored.

## 7. REFERENCES

- [1] Chai, W. & Vercoe, B. Folk music classification using hidden Markov models. *Proceedings of International Conference on Artificial Intelligence*, 2001.
- [2] GNU LilyPond. <http://www.lilypond.org>
- [3] Huang, M. X., Tang, W., Lo, K. W., Lau, C., Ngai, G. & Chan, S. MelodicBrush: a cross-modal link between ancient and digital art forms. *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, ACM Press, 2012, 995-998.
- [4] Huang, M. X., Tang, W. W., Lo, K. W. K., Lau, C. K., Ngai, G. & Chan, S. MelodicBrush: a novel system for cross-modal digital art creation linking calligraphy and music. *Proceedings of the Designing Interactive Systems Conference*, ACM Press, 2012, 418-427.
- [5] Jo, K. DrawSound: a drawing instrument for sound performance. *Proceedings of the 2nd international conference on Tangible and embedded interaction*, ACM Press, 2008, 59-62.
- [6] Jones, K. Compositional Applications of Stochastic Processes. *Computer Music Journal*, MIT Press, 1981, 5(2), 45-61.
- [7] Kang, L. & Chien, H.-Y. Hé: Calligraphy as a Musical Interface. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2010, 352-355.
- [8] Miranda, E. R. Evolving Cellular Automata Music: From Sound Synthesis to Composition. *Proceedings of Workshop on Artificial Life Models for Musical Applications*, Prague University of Economics, 2001.
- [9] Ponsford, D., Wiggins, G. & Mellish, C. Statistical Learning of Harmonic Movement. *Journal of New Music Research*, 1999, 28, 150-177.