國立交通大學

音樂研究所　音樂科技組

碩士論文

以演算法作曲為基礎的華文詩詞與書法之可聽化研究

A Study of Algorithmic Composition-Based Sonification on Chinese Classical Poetry and Chinese Calligraphy Painting

研究生：　任珍妮

指導教授：黃志方

中華民國 九十八 年 六 月

以演算法作曲為基礎的華文詩詞與書法之可聽化研究

A Study of Algorithmic Composition-Based Sonification on Chinese Classical Poetry and Chinese Calligraphy Painting

研究生：任珍妮　　Student: Jenny Ren

指導教授：黃志方　　Advisor: Chih-Fang Huang

國立交通大學

音樂研究所　音樂科技組

碩士論文

A Thesis Submitted to
the Institute of Music
College of Humanities and Social Sciences
National Chiao Tung University
in Partial Fulfillment of
the Requirements for the Degree of
Master of Art
(Music Technology)

Hsinchu, Taiwan

June 2009

中華民國 九十八 年 六 月

以演算法作曲為基礎的華文詩詞與書法之可聽化研究

研究生：任珍妮 指導教授：黃志方博士

國立交通大學音樂研究所

摘要

自古至今，中華文字在藝術上的精緻表現以詩詞與書法著稱。然而，對於視覺障礙或不熟悉華文的人們來說，體驗中華文字之美卻總是困難重重。本研究分別針對五言絕句與草書提出兩種可聽化的方法，目的在於萃取出詩詞字裡行間的韻律之美以及書法揮毫紙上的線條之美，將其映射為數字化資料，並以演算法作曲化數字為旋律。

本論文在詩詞可聽化的過程中，一方面透過文字平仄的格律分析，依據馬可夫鍊判斷節奏的連接、控制音符音量大小與音程變化的範圍；另一方面藉由語音聲韻的共振峰分析，以歐基里得距離挑選出最佳五音調式，並運用篩選理論篩選出符合最佳調式的音高。在書法可聽化的過程中，運用影像分析將二維空間領域的影像資訊轉換成時間領域與頻率領域的聲音資訊。

透過詩詞與書法的可聽化研究，除了提供視覺障礙或不熟悉華文的人們另一種途徑來欣賞中華文字的藝術表現之外，也使得詩詞與書法在文字與圖像上可以透過聲音的輔助讓鑑賞者能更即時地感知其中的意境，甚至建立出一套沉浸式華文詩詞與書法學習環境。

關鍵字：演算法作曲、華文詩詞與書法、可聽化、馬可夫鍊、共振峰

A Study of Algorithmic Composition-Based Sonification on Chinese Classical Poetry and Chinese Calligraphy Painting

Student: Jenny Ren Advisor: Dr. Chih-Fang Huang

Institute of Music
National Chiao Tung University

Abstract

Chinese characters are remarkable for their two forms of art — the classical poetry and the calligraphy painting. However, it is difficult for the visually impaired and people who are unfamiliar with Chinese to experience the beauty of the Chinese characters. In this study, two Sonification schemes, *Tx2Ms* and *Im2Ms*, are proposed to extract the melody between the lines, i.e., both the lines in verses and the lines in strokes.

In *Tx2Ms*, the movement of multi-dimensional musical elements such as durations, dynamics and interval relations are modeled by Markov Chain for stochastic algorithmic composition based on the poesy analysis. In addition, the best pentatonic mode for a specific poem is recommended according to the formants analysis. In *Im2Ms*, the two-dimensional spatial image information is transformed into the temporal music acoustics domain based on artistic conception and human perception among space, color and sound.

Therefore, the Sonification of Chinese Classical Poetry and Chinese Calligraphy

Painting not only provide a free access for the visually impaired and people who are unfamiliar with Chinese to appreciation but also enrich the state of mind and imagery in the delivery process. Thus, an immersive learning environment of Chinese Classical Poetry and Calligraphy Painting can be further developed.

Keywords: Algorithmic Composition, Sonification, Chinese Classical Poetry and Chinese Calligraphy Painting, Markov Chain, Formants

Acknowledgements

This thesis is dedicated to Dr. Phil Winsor, father of Computer Music in my life, who inspires and encourages me in the beautiful wonderland of Computer Music. I would like to acknowledge the contribution of my advisor, Dr. Chih-Fang Huang, for his continued advisement and kindly guidance during the past three years. Thank you for supporting my work and dreams! Without them, this work would not have been possible. I would also like to thank members of my committee: Dr. Shian-Shyong Tseng and Dr. Ming-Sian Bai for their useful feedback on this research.

Special thanks go to James Ma, Dr. Chao-Ming Tung, and Dr. Yu-Chung Tseng, from whom I have learned diverse and invaluable lessons about Electro-Acoustic Music Composition. Thanks also go to Dr. Lap-Kwan Kam, from whom I have learned the way of thinking. I am thankful to my friends in Music Technology, Eddie, Dennis, Karol, Gaspard, Vivian, John, Jose, and Hsing-Ji, who usually provid me with stimulus for better work.

Finally, I am deeply grateful to my wonderful family for their love, patience, and support: especially my beloved Daddy Douglas, Sister Anny, Aunt Lillian, and Grandparents, who always pray for me during the toughest time.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Motivation and Objectives

The birth of this study is driven by the idea: *"Is there any mechanism for assisting the visually impaired, and people who are unfamiliar with Chinese in experiencing the art in forms of Chinese characters (i.e., poetry and calligraphy) through an alternative modality, hearing?"* Fortunately, we found that text, image and sound, all these mediums can evoke emotional responses. Since each modality has its certain strengths and each combination of modalities may produce different synergistic results, sound can provide an additional and complementary perceptual channel. Besides, sound can be used to augment the visualization by permitting a user to visually concentrate on one field, while listening to the other. Consequently, the aim of this study is to explore and utilize the auditory display to strengthen the synesthesia and to supplement the visual interpretation of data based on the artistic interrelationship. In other words, the digital data in three different kinds of medium (text, images and sounds) are being manipulated. Fig. 1 depicts the interaction of different display modalities of cross-disciplinary arts.

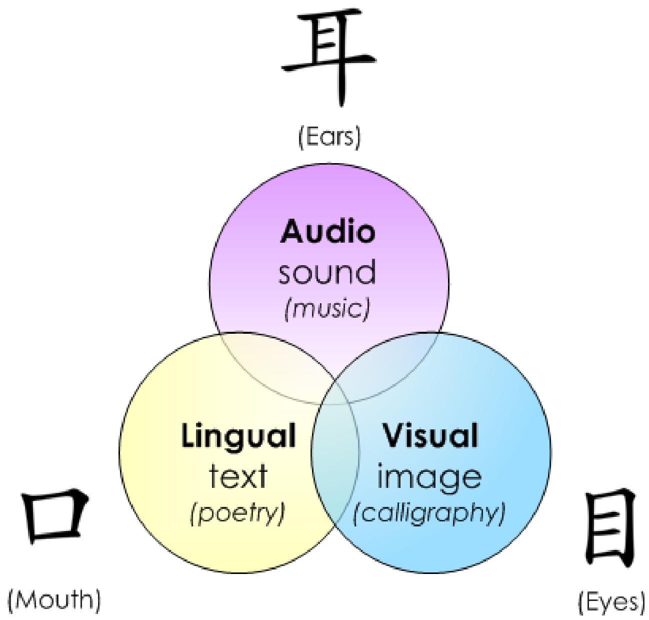


Fig. 1 Three display modalities of cross-disciplinary arts in this study

1.2 Scenarios and Contributions

The following are the two basic scenarios in this study: the former maps from text onto sound, or, language to music; the latter maps from image onto sound, or space to time.

1. Learning the Chinese Classical Poetry, taking the example of the "Five-Character Quatrain".
2. Appreciating the Chinese Calligraphy Painting, taking the example of the "Cursive Script".

Although there has been research on mapping from text or image to sound, none of them is dedicated to Chinese characters. The contributions of this study are as follows: Firstly, either text or images, is analytically transformed from lingual or graphic data view into abstract sonic space. Secondly, data is mapped to sound in a musical way. The visual data representation is algorithmically compiled into audio data representation with philosophical and aesthetic interrelation through compositional mind and process rather than arbitrarily or directly converted—a step forward from Sonification to Musification. Thirdly, an immersive learning environment with audio-visual aids is built since it supports concentration, provides engagement, increases perceived quality, and enhances learning creativity during the appreciation process.

1.3 Thesis Organization and Research Process

The remainder of this paper is organized as follows: Chapter 2 presents a profound survey on Sonification studies and existing tools for both text-to-audio and image-to-audio transformations. To understand the role of Chinese characters in classical poetry and calligraphy painting, this chapter also gives a brief sketch of the imagery and state of mind inside the Chinese characters. Next, we propose our approach of both text-to-music (Tx2Ms) and image-to-music (Im2Ms) mappings from structural music-level aspect rather than from direct audio-level aspect (Chapter 3). A prototype implementation and experimental results are presented in Chapter 4. Finally, we summarize the results of our work (Chapter 5).

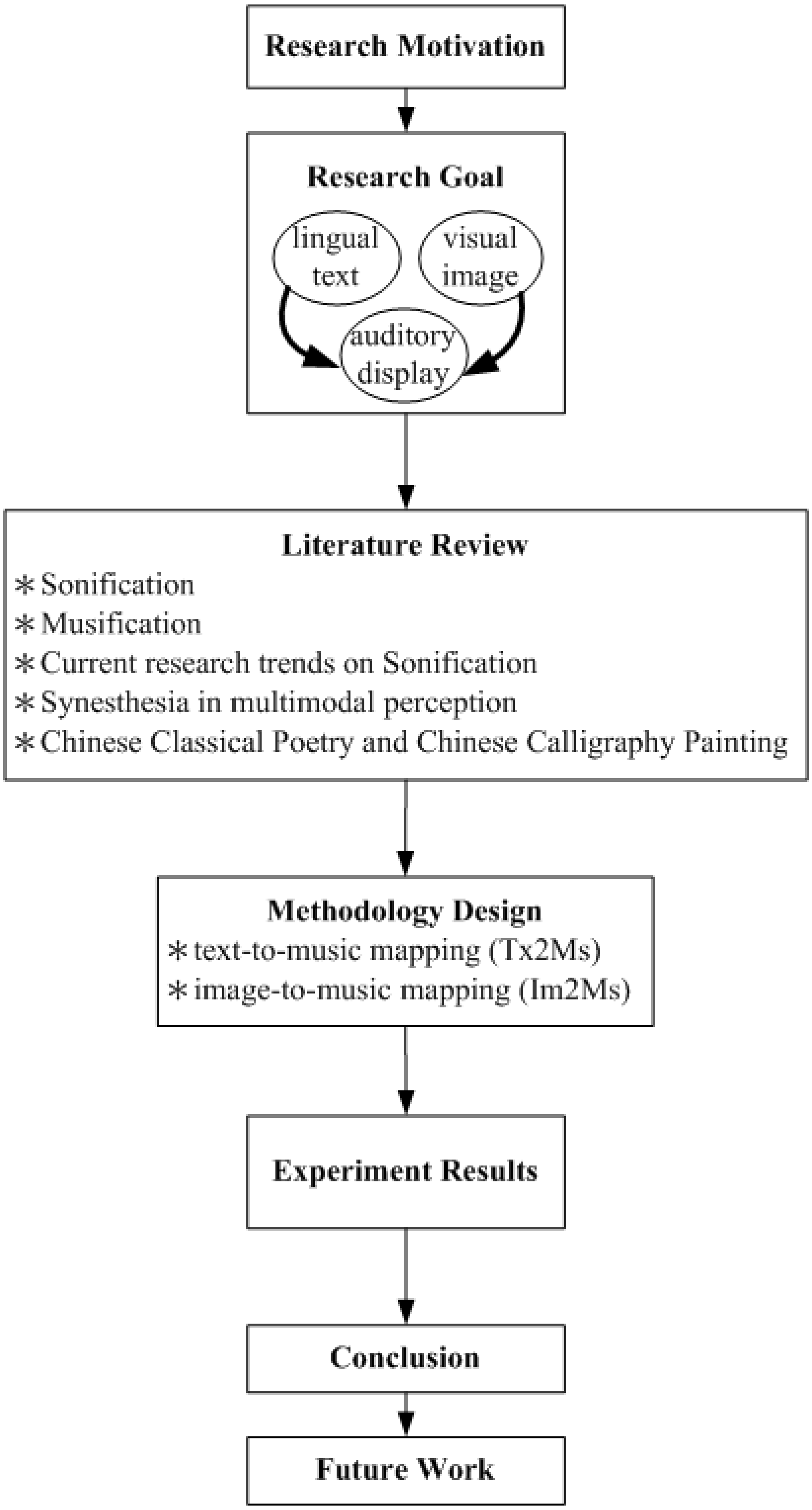


Fig. 2 Research process flow chart

Chapter 2 Related Works

As shown in Table 1, the framework of literature reviews in this study is based on the motivation and background knowledge, which is mainly focused on Sonification, Musification, current research trends on Sonification, and Synesthesia in audio-visual perception.

Table 1 Framework of literature reviews

Domains	*Aspects*
Sonification	Definition Techniques Elements Schema Effectiveness consideration
Musification	Definition Difference between Sonification and Musification
Current research trends on Sonification	Sonification in a diversity of usages Text-to-Sound Sonification Image-to-Sound Sonification
Synesthesia in multimodal perception	Synesthesia and multimodalities Case study on educational environment
Two major arts in the form of Chinese character	Chinese Classical Poetry Chinese Calligraphy Painting

2.1 Sonification

The word "Sonification" comprises the two Latin syllabus "sonus", meaning sound, and the ending "fication", forming nouns from verbs which are ending with '-fy'. Therefore, to "sonify" means to convey the information via sound. A Geiger detector can be seen as the very basic scientific example for Sonification, which conveys (i.e., sonifies) information about the level of radiation. A clock is even more basically an example for Sonification, which conveys the current time.

The word "Sonification" has already been defined in a majority of researches. "*Sonification is to communicate information through nonspeech sounds*" [6] (Ballas 1994, 79); "*Sonification is the use of data to control a sound generator for the purpose of monitoring and analysis of the data*" [17] (Kramer 1994b, 187); "*Sonification is the transformation of data relations into perceived relations in an acoustic signal for the purpose of facilitating communication or interpretation*" [18] (Kramer et al. 1999); "*Sonification is a mapping of numerically represented relations in some domain under study to relations in an acoustic domain for the purpose of interpreting, understanding, or communicating relations in the domain under study*" [24] (Scaletti 1994, 224).

Fig. 3 illustrates the existing Sonification techniques, which are already categorized into three types according to the mapping approach adopted: syntactic, semantic or lexical mapping [7] [8].

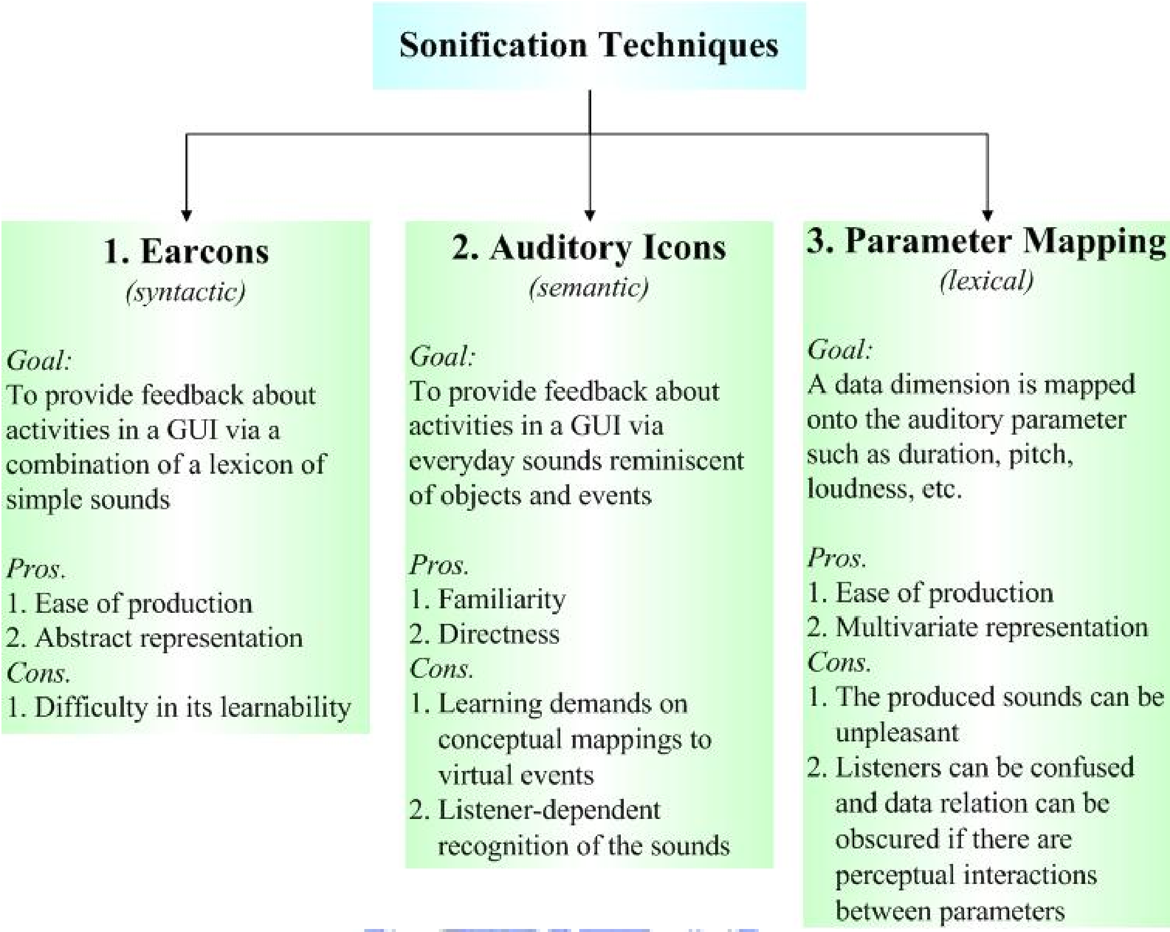


Fig. 3 Three types of existing Sonification techniques

(arranged from [7])

Besides, the fundamental elements of a Sonification are suggested in Fig. 4 from both data-centric and human-centric points of view, including the functionality to be identified, tasks to be performed, and several related disciplines to be worked with [23].

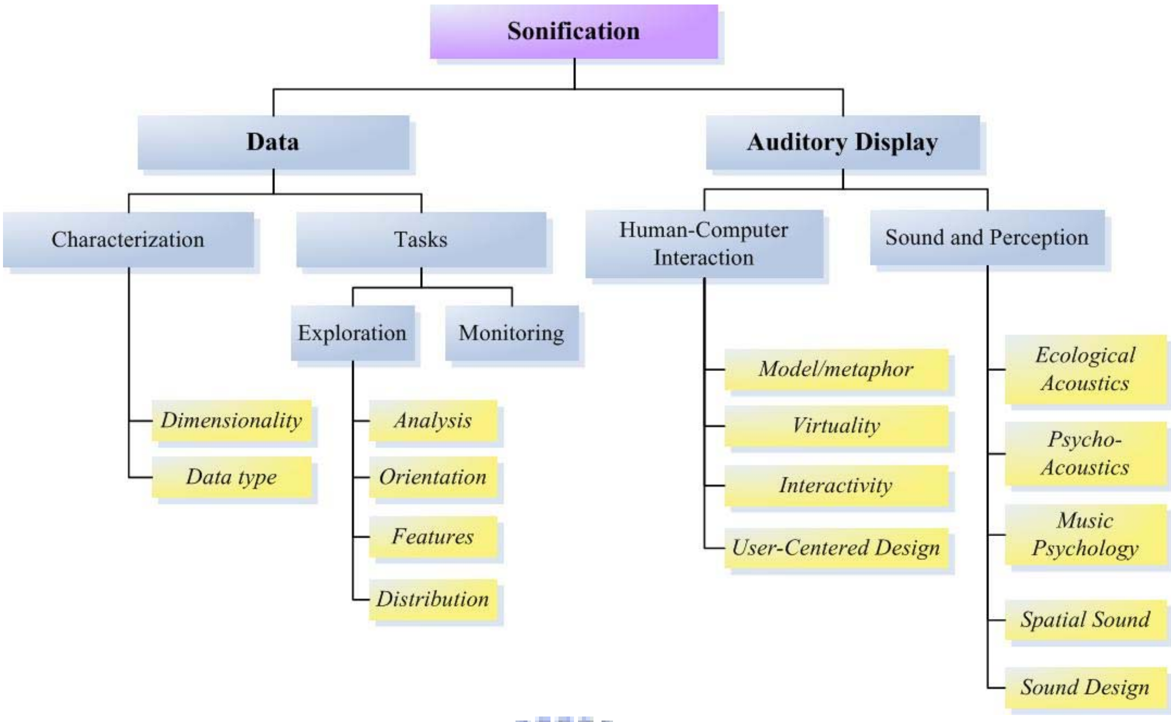


Fig. 4 Elements hierarchy of a Sonification display

(rewritten from [23])

Moreover, Fig. 5 illustrates the fundamental procedure about how to design a Sonification system, where the Communicative Medium is the core of Sonification [16].

Fig. 5 Schematic of an Auditory Display System

(rewritten from [16])

However, the essential goal of Sonification is to yield an auditory display that will be orderly and intuitively maximal in meaning (i.e., coherence) to the observer. Inevitably, the effectiveness is what most counts in designing a Sonification software or system (see Fig.

6).

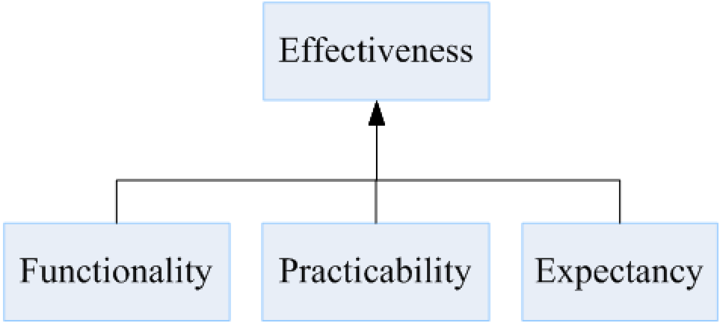


Fig. 6 How to make an effective Sonification?

- Functionality
  - The goal-oriented function of the system must be clearly defined.
- Practicability
  - If the sound is ugly, people won't use it!
  - The craft of composition is important to auditory display design (i.e., a composer's skill can contribute to making auditory displays more pleasant and sonically integrated and so contribute significantly to the acceptance of such displays).
- Expectancy
  - Evaluation (e.g., questionnaire) is needed.

2.2 Musification

Since data can be visualized by means of graphics and sonified by means of sounds, data can be musified by means of music as well. “*Musification is the musical representation of data*” [10] (Edlund 2004). However, the difference between Sonification and Musification lies in the fact — Music is “organized sounds” (coined by French composer, Edgard Varèse). Specifically speaking, data is no longer directly mapped onto audio signal level, but algorithmically complied onto musical structure level, which means, to follow some musical grammars or based on musical acoustics.

2.3 Current Research Trends in Sonification

This section reviews relevant research trends in Sonification from three aspects: a diversity of usages, text-to-sound, and image-to-sound.

2.3.1 Sonification in a diversity of usages

Sonification has been put into practice in a variety of areas, inclusive of medical usages, assistive technologies, or even data mining and information visualization. The idea of using Sonification in medical usage is to use sounds to diagnose illness; the idea of carrying out Sonification in assistive technologies is to make maps, diagrams and texts more accessible to the visually impaired through multimedia computer programs; the idea of applying a direct playback technique, called “Audification”, in data mining and information visualization is to assist in overviewing large data sets, event recognition, signal detection, model matching and education [7]. Besides, the method for rendering the complex scientific data into sounds via additive sound synthesis and further visualizing the sounds in Virtual-Reality environment has been proposed in [15], which is aimed to help scientists explore and analyze huge data sets in scientific computing.

2.3.2 Text-to-Sound Sonification

The program "Poem Generator" in Phil Winsor's book, "*Automated Music Composition,*" has already illustrated the conversion from the constituent letters to the pitch domain by mapping their individual ASCII values onto the pitch values in MIDI. The mapping mechanism is basically derived from the idea that each character has an inherent ASCII values as its digital information in every computer. The output results convey the structure of the letters in a phrase, where rests are allocated for blank spaces and pitches are assigned for different ASCII values of the letter [25].

2.3.3 Image-to-Sound Sonification

Kandinsky produced many paintings, which borrows motifs from traditional European music, based on the correspondence between the timbres of musical instruments and colors of visual image [14]. Contrary to Kandinsky's attempt to "see the music", there are researchers and artists who have been trying to "hear the image".

Iannis Xenakis' UPIC (Unité Polyagogique Informatique du CEMAMu) system may be one of the first digital graphics-to-sound schemes. Composers are allowed to draw lines, curves, and points as a time-frequency score on a large-size and high-resolution graphics tablet for input [19]. Later on, many of the ideas that drive image-to-sound software are inspired from Xenakis' research.

Unlike UPIC as a graphical metaphor of score, Coagula is an image synthesizer which uses pixel-based conversion, where x and y coordinates of an image are regarded as time and frequency axis, with a particular set of color-to-sound mappings. Red and Green control stereo panning, while Blue smears the sounds to noise content [11].

The vOICe (read the capitalized letters aloud individually to get “Oh, I see!”), or, “Seeing with Sounds”, is a system that makes inverted spectrograms in order to translate visual images into sounds, where the two-dimensional spatial brightness map of a visual image is 1-to-1 scanned and transformed into a two-dimensional map of oscillation amplitude as a function of frequency and time [20]. The mapping translates, for each pixel, vertical position into frequency, horizontal position into time-after-click, and brightness into oscillation amplitude — the more elevated position the pixel, the higher frequency the associated oscillator; the brighter the pixel, the louder the associated oscillator. The oscillator signals for a single column are then superimposed.

In Wang’s research, the image is converted from RGB to HSI system and then be mapped from Hue (0-360 degree) to pitch (MIDI: 0-127), from Intensity (0-1) to playback tempo (0-255), respectively [1]. In the research of Osmanovic, the image is mapped from its electromagnetic spectrum to tone frequency and from intensity to volume based on color properties and sound properties. The frequency of the tone is redoubled 40 times to compute the frequency of the color: $\textbf{\textit{tone}} \times 2^{40} = \textbf{\textit{color}}$ [22].

After all, a vast majority of the previous works focus on Image-to-Sound, or even Image-to-Music mechanisms rather than Text-to-Music mechanisms (as shown in Fig. 7).

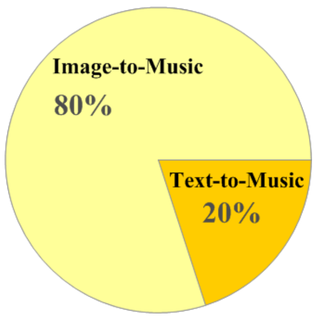


Fig. 7 Comparison of proportions of the previous works on Image-to-Music and Text-to-Music mappings

2.4 Synesthesia in Multimodal Perception

Synesthesia: from the Greek *syn*, meaning together, and *aisthesis*, meaning sensation, literally means experiencing together. The human perception or cognition is a multimodality process, combined of several sensations such as auditory modality, visual modality, and so forth. Take the three different modalities in Fig. 1 for example.

- Between lingual text and non-lingual audio sound:

  Rhythm plays a significant role in reading. Besides, the linguistic tone of voices makes the words expressed with the implications of pitch, loudness, and speed/tempo. For instance, a sentence at loud volume might convey the feel of anger, while a sentence at fast tempo might imply the feel of urgency.

- Between visual image and non-lingual audio sound:

  The spatial shaping constituents of a painting made up of dots, lines, shapes and colors can bring about temporal musical features. Prof. Pei-sui Ma, a watercolor painter, once mentioned that the length, width, position, slope, thickness, and density of lines can produce the correspondence to pitch contour, loudness, and tempo.

The experiment conducted by Chen [3] shows the benefits how audio-visual multimedia and its methods facilitate learning. The result proves that audio-visual aids in educational environment assist the students in learning more, learning faster, and memorizing for a longer time (as shown in Table 2 and Table 3).

Table 2 Learning with audio-visual media and methodology

Audio-Visual Media and Methods	*To Use*	*Not To Use*
A degree of understanding in a limited time	94	6
B time to spend until complete understanding	1	12

(rewritten from [3])

Table 3 The relation between memory maintenance and media usage

Usage Methods	*Memory Maintenance*	
	3 hours later	*3 days later*
Hearing Only (oral teaching)	70%	10%
Vision Only (observation method)	72%	20%
Both Hearing and Vision (audio-visual method)	85%	65%

(rewritten from [3])

2.5 Chinese Classical Poetry and Chinese Calligraphy Painting

Chinese characters have been classified into six categories by etymology: pictogram, ideogram, phonetic compounding, meaning aggregation, mixed word creation, and transliteration. Strictly speaking, the Chinese character is a logogram, primarily comprising pictograph and semasiograph, different from the phonogram, which represents phonemes (speech sounds) or combinations of phonemes. From the old days, the Chinese character is not only a kind of symbol to record the language but also a kind of art. The two most well-known artistic creations based on the Chinese characters are Chinese Classical Poetry and Chinese Calligraphy Painting. The former uses characters for syntactic expression and for semantic narration to deliver the beauty of speech, while the latter uses characters for graphemic structure to reveal the beauty of lines and shapes.

Jintishi, or, "modern-form poetry", is one set of the popular poetic forms among Chinese Classical Poetry. In these form, each couplet comprises a series of set tonal patterns using the four tones of the mid-ancient Chinese pronunciation. There are basically the level, rising, falling and entering tones in the classical Chinese intonation system. Furthermore, the key to the composition of Jintishi hinges on the intonation score of Ping/Ze opposition in traditional Chinese verse, where level tone belongs to Ping and the others belong to Ze. Overall, Jintishi is a specific form of Chinese Classical Poetry which carries consistent and well-defined rules for not only its prosody (i.e., regular meter, rhythm and intonation) but also the rhyming scheme.

Jintishi could be further categorized into three major forms based on the number of lines in each poem [2]. (All forms of Jintishi could be written in five or seven character lines.)

- Quatrain (with four lines in each poem): Some tonal patterns are followed.
- Regulated Verse (with eight lines in each poem): In addition to the tonal constraints, this form requires parallelism between the lines in the second (third and fourth lines) and third (fifth and sixth lines) couplets. The lines in these couplets have contrasting content, while the characters in each line are in the same grammatical relationship.
- Long poem in Regulated Verse (with over eight lines in each poem): This form extends the Regulated Verse to unlimited length by repeating the tonal pattern. The parallelism is required in each couplet except the first and last couplets.

However, the tonal rules received greater emphasis than parallelism. According to the tonal rules [2] [4] [5], we can infer four basic types of tonal patterns of the Five-Character Quatrain, where 1 represents Ping, and 0 represents Ze, respectively (Table 4).

Table 4 Four basic types of tonal patterns of Five-Character Quatrain

Type	*Name*	*Ping Ze Structure*	*Ping Ze ID*
I	First-line rhyming and the second syllable being Ping	PPZZP ZZZPP ZZPPZ PPZZP	11001 00011 00110 11001
II	First-line without rhyming and the second syllable being Ping	PPPZZ ZZZPP ZZPPZ PPZZP	11100 00011 00110 11001
III	First-line rhyming and the second syllable being Ze	ZZZPP PPZZP PPPZZ ZZZPP	00011 11001 11100 00011
IV	First-line without rhyming and the second syllable being Ze	ZZPPZ PPZZP PPPZZ ZZZPP	00110 11001 11100 00011

Chinese Calligraphy Painting is highly ranked as an important art form in East Asia, referring to the beautiful handwriting of Chinese characters. Seal Script, Clerical Script, Cursive Script and Regular Script are the primary styles in the evolution of Chinese Calligraphy. Table 5 displays the same Chinese character, which means "thousand" in Chinese, in four different styles. Among all, Cursive Script is the most expressive and individual style, which draws the musical rhythm and speed in two dimensional space on the Shuan Paper.

Table 5 Evolution of four primary styles of Chinese Calligraphy

Image	*In Chinese*		*In English*
	篆書	chuan-shu	Seal Script
	隸書	li-shu	Clerical Script
	草書	tsao-shu	Cursive Script
	楷書	kai-shu	Regular Script

These two artistic creations of Chinese characters are rich in poetic and pictorial splendor, and deep in implicit imagery. The so-called state of mind of an artistic work is composed of subjectively emotional feeling and objectively existential image in harmony and unity. The state of mind is the territory of imagery expansion and the land with flows of emotions. On the one hand, the refined verses with phonemic orderliness give birth to the pleasant sounds of the recitation, making the poetry easy to read and to remember, which might have been lingering in the audience's heads for days. On the other hand, the thickness, length, strength, speed and shape of the characters with the transition (stop and change) of the brush strokes convey the disposition of the calligrapher and enchantment of the calligraphy itself.

Chapter 3 Methodology

“The mapping problem” has been regarded as the essential issue of Sonification. In two representative artistic creations which stem from Chinese character, we found that the poets are interested in the patterns of prosody while the calligraphers signify their personality through the strokes and modeling of a character.

In this study, the Sonification mechanisms of transforming Chinese Classical Poetry (text) and Chinese Calligraphy Painting (image) into music are explored respectively. First of all, the aesthetic features are extracted from Chinese Classical Poetry and Chinese Calligraphy Painting. Afterwards, the rules are applied to individual parameter-mapping mechanisms. The conversion of text-to-music mapping is based on the pronunciation properties and the syntax characteristics in Chinese Classical Poetry. The conversion of image-to-music mapping is based on a relationship that exists among space, color and sound in human perception. Fig. 8 illustrates a general paradigm of our Sonification in this study.

A limited number of features and corresponding sonic attributes are taken into account so as to keep the resultant sounds as simple as possible and easy to decode because the listeners always wish to hear what the data is doing. However, the sound will be still rich in itself.

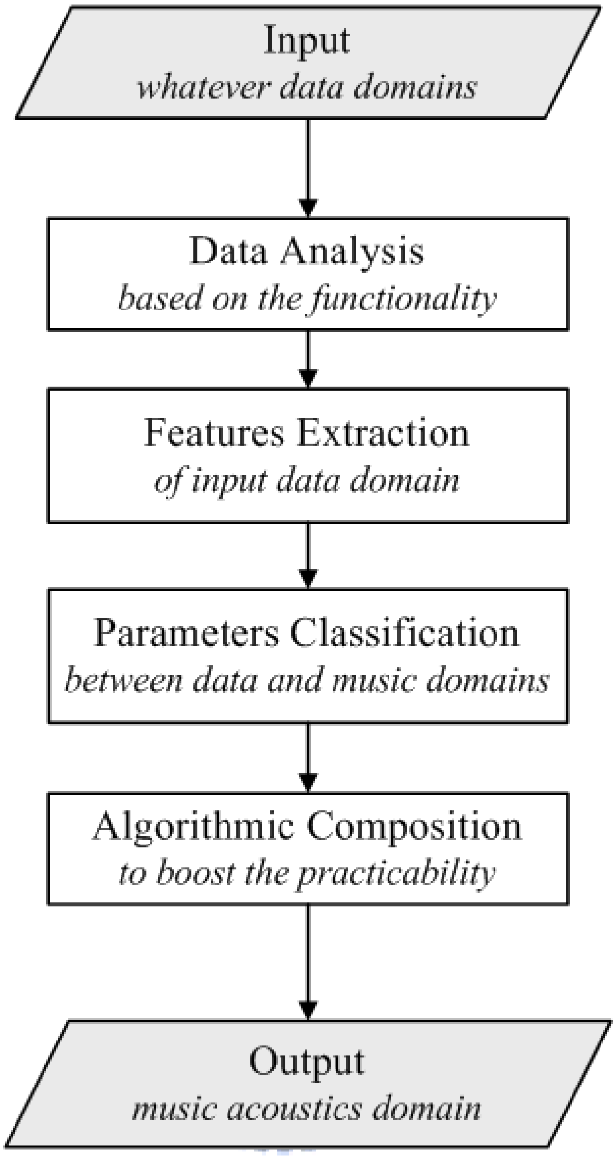


Fig. 8 General paradigm of Sonification

3.1 Tx2Ms (text-to-music mapping of Chinese Classical Poetry)

Throughout this section, a mechanism of mapping poetry data onto appropriate sound features is provided, followed by an overview of the system architecture.

3.1.1 Mapping Recipe of Tx2Ms

The four tones in mid-ancient Chinese pronunciation are dichotomized into the only two categories in the classical Chinese intonation with the following characteristics (as shown in Table 6).

Table 6 Features of classical Chinese intonation system

Two Categories	*Four Tones*	*Characteristics*
Ping (level tone)	Level Tone	Long, without any inflection
Ze (deflected tones)	Rising Tone	Moving up
	Falling Tone	Moving down
	Entering Tone	Short

In addition to the intonation of the prosody, there are more features which could be taken into consideration, such as the semantic mood or style of the poetry (see Table 7).

Table 7 Parameters classification and mapping of Chinese Classical Poetry

Parameters Classification between two domains		***Music***					
		Rhythm	*Interval Size*	*Sonority*	*Dynamics*	*Tempo*	*Mode*
Poetry	*Prosody Intonation*	O	O	O	O	X	O
	Poetic Mood	X	X	X	O	O	X

For instance, the parameter of intonation (i.e., Ping or Ze) in text domain is reasonably classified and mapped to multiple parameters in music domain, where interval size refers to horizontal adjacent pitches and sonority refers vertical simultaneous pitches, respectively. Table 8 shows the mapping between poetic attributes and music parameters. (The characters “C”, “J”, and “B” represent Chinese Pentatonic Mode, Japanese Hirajoshi Five-Tone Mode, and Balinese Gamelan Five-Tone Pelog Mode respectively.)

Table 8 Mappings from poetry information to musical parameters

<table>
<tr><th colspan="2" rowspan="2">Parameters Mapping</th><th colspan="6">Music</th></tr>
<tr><th>Rhythm</th><th>Interval Size</th><th>Sonority</th><th>Dynamics</th><th>Tempo</th><th>Mode</th></tr>
<tr><td rowspan="3">Prosody Intonation</td><td>Ping</td><td>Sparse</td><td>small</td><td>harmonic</td><td>soft</td><td colspan="2" rowspan="2">NULL</td></tr>
<tr><td>Ze</td><td>Dense</td><td>large</td><td>inharmonic</td><td>loud</td></tr>
<tr><td>Tone</td><td colspan="5">NULL</td><td>C/J/B</td></tr>
<tr><td rowspan="3">Poetic Mood</td><td>brightness</td><td colspan="3" rowspan="3">NULL</td><td>loud</td><td>fast</td><td rowspan="3">NULL</td></tr>
<tr><td>darkness</td><td>soft</td><td>slow</td></tr>
<tr><td>neutrality or exoticism</td><td>free</td><td>free</td></tr>
</table>

3.1.2 Preliminaries of Tx2Ms

a. Use Markov Chain in transition table construction

There are several divisions of techniques in algorithmic composition, inclusive of stochastic, rule-based flow control, grammar, chaotic and artificial intelligence. Markov Chain is one of the stochastic processes in probability theory. The Markov models have been widespread used in many other fields, like Wireless Communication and Bioinformatics.

Most of all, the principle of Markov Property is to memorize the current state. Thus, the conditional probability of future states of the process depends only on the current state, i.e. it is conditionally independent of the past states, and the path of the process, given the present state. For instance, the probability of the (N+1)th state only correlates to the current Nth state, having nothing to do with other previous states. Accordingly, the following shows the Markov Property, also known as Markovian.

$$\Pr\left(S_{n+1} \middle| S_n, S_{n-1}, \ldots, S_1, S_0\right) = \Pr\left(S_{n+1} \middle| S_n\right) \tag{1}$$

Further, the Markov Chain is described as a sequence of random variables S_1, S_2, S_3…, S_n, with Markov Process, where each S_i is one of the possible values from a state space S. Take two lower level musical elements for example. The state space of Pitch Class is {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11} while the state space of Rhythm could be {1/1, 1/2, 1/4, 1/8, 1/16, 1/32}. The following illustrates the process with Markov Property in Markov Chain.

$$S(t_0) \rightarrow S(t_1) \rightarrow S(t_2) \rightarrow \cdots \rightarrow S(t_n) \rightarrow S(t_{n+1}) \tag{2}$$

A Markov Chain could be represented either by a Directed Graph or a Transition Matrix. A Directed Graph consists of a set of states and a set of transitions with associated probabilities. A Transition Matrix of an N+1-dimensional probability table represents an Nth-order Markov Chain, which tells us the likelihood of an event's occurrence, given the previous N states [21] [25]. Fig. 9 and Table 9 show the 1^{st}-order Markov Chain in terms of Directed Graph as well as Transition Matrix where C, E and G refer to the name of the tone (Do, Mi and So), respectively.

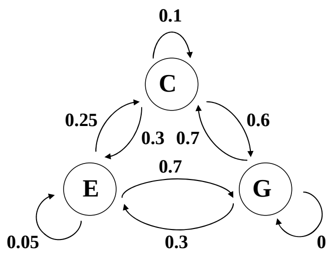


Fig. 9 Directed Graph

Table 9 Transition Matrix

Next / *Current*	**C**	**E**	**G**
C	0.1	0.3	0.6
E	0.25	0.05	0.7
G	0.7	0.3	0

The Transition Matrix (or the stochastic matrix) P is the transition probability distribution, with *(i,j)'*th element of P equals to

$$p_{ij} = \Pr(S_{n+1} = j | S_n = i) \tag{3}$$

A vast majority of the uses of Markov Chain in the algorithmic composition is to analyze and model the existing compositions. For example, some researches have already analyzed the improvisation and chord progression by means of Markov Model [9] [12] [13]. The Markov Models used in these studies are mainly regarded as an analyzer or a model. In this study, we suppose the Markov Model contribute to the stochastic algorithmic composition (see Fig. 10). The function of Markov Model is to facilitate meaningful mapping between poesy data and musical elements. With the advent of highly-relevant music, the emotional perception could be greatly improved during Chinese Classical Poetry appreciation.

Fig. 10 The Markov Model is developed for stochastic algorithmic composition

In the Preprocessing Phase I of Tx2Ms, the interchanging Ping/Ze in each phrase is decomposed and aggregated for further recomposing of the rhythm sequence of original complete poem. Based on the rhythm sequence, a rhythmic transition table of 1^{st}-order Markov Chain is computed for further algorithmic composition design. Fig. 11 takes the famous Five-Character Quatrain, *"Love Seed"*, for example.

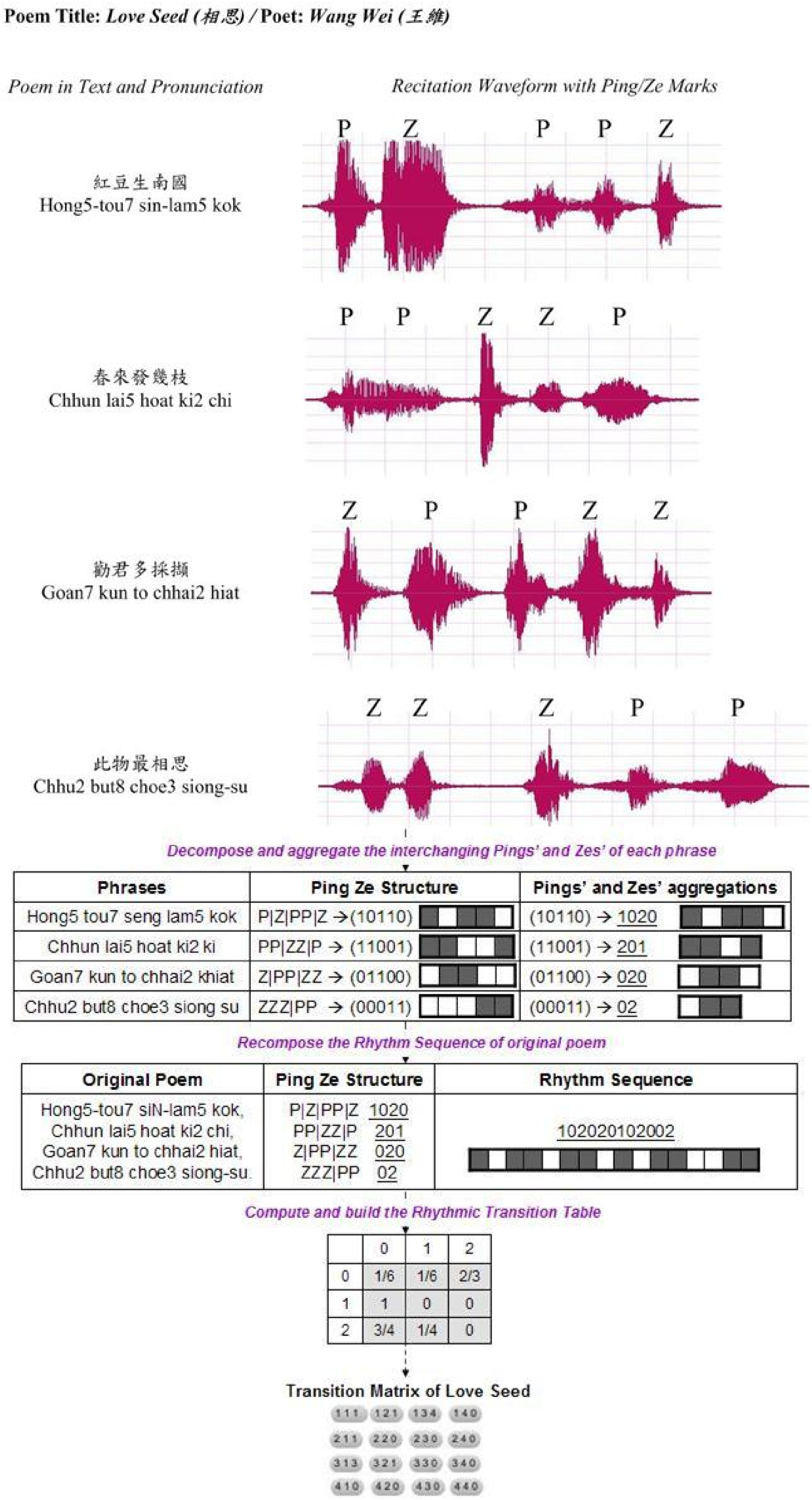


Fig. 11 An example of computing the rhythmic transition table from intonation

b. Apply Sieve Theory in pentatonic mode generation

The pentatonic or five note mode occurs in most of the ancient folk music in Asia. The prevalence of pentatonic modes in Chinese, Japanese, and Javanese music makes pentatonic modes have an Asian character for a long time. In particular, the pentatonic mode typifies the Chinese-style music since the traditional Chinese music is primitively based on pentatonic mode. Besides, A Pentatonic Mode, or, a Five Tone Mode, is a mode with five notes per octave. Table 10 and Fig. 12 present three typical pentatonic modes in the Tx2Ms.

Table 10 Three pentatonic modes used in Tx2Ms

Pentatonic Mode Name	*Pitch Name*	*Pitch Class (PC)*
Chinese Pentatonic	(C, D, E, G, A)	PC_C: (0, 2, 4, 7, 9)
Japanese Hirajoshi Five-Tone	(C, D, E♭, G, A♭)	PC_J: (0, 2, 3, 7, 8)
Balinese Gamelan Five-Tone Pelog	(C, D♭, E♭, G, A♭)	PC_B: (0, 1, 3, 7, 8)

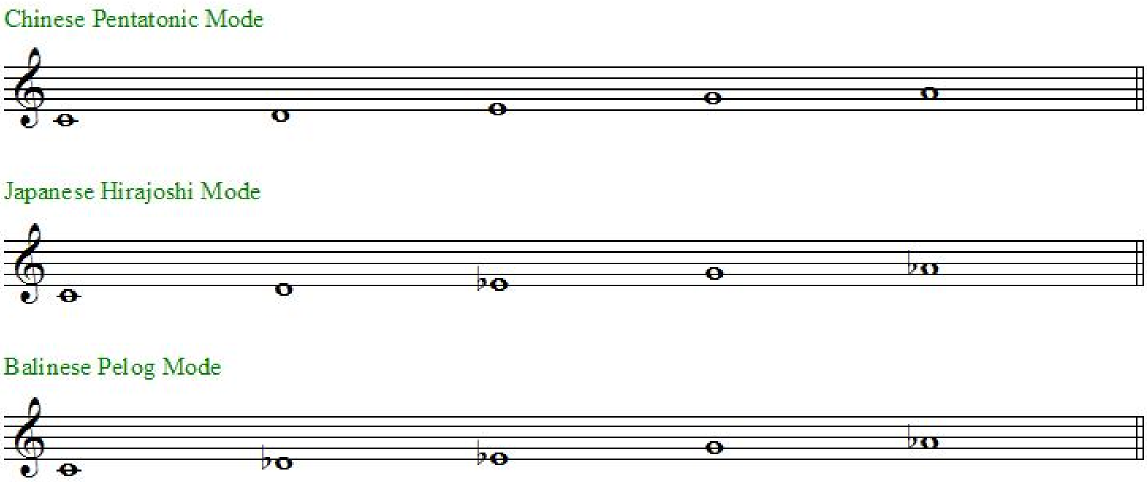


Fig. 12 The notation of the three typical pentatonic modes

Sieve Theory is utilized here to generate pitches within a specific mode once the mode is recommended by *Tx2Ms*. Pitch Class uses "modulo 12". By using "mod 12", any integer number above 12 should be reduced to a number from 0 to 11. This modulo operator can be visualized using a clock face (Fig. 13):

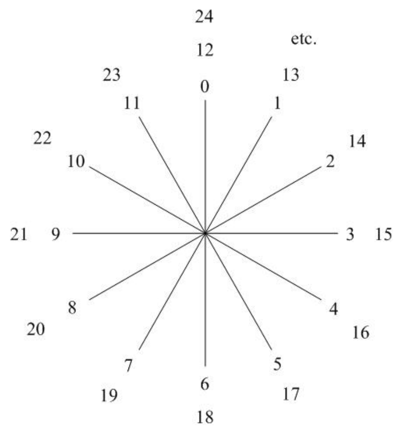


Fig. 13 Pitch Class of modulo 12

The function is described as below, where RP means Random Pitch (i.e., a random number integer) and $0 <= RP <= 127$; RC refers to Residue Class (i.e., the set of integers filtered), and is specified $RC = \{a, \ldots, b\}$, where a is the minimum, and b is the maximum.

$$RC = RP \bmod 12 \tag{4}$$

- RC set of Chinese Pentatonic Mode (RC_C): *RP (mod 12) == {0, 2, 4, 7, 9}*
- RC set of Japanese Hirajoshi Mode (RC_J): *RP (mod 12) == {0, 2, 3, 7, 8}*
- RC set of Balinese Pelog Mode (RC_B): *RP (mod 12) == {0, 1, 3, 7, 8}*

For simplification, RC_C, RC_J, and RC_B are all named Pitch Class (PC) in ascending order of individual modes as PC_C, PC_J, and PC_B, respectively.

c. Speech-to-Mode Conversion based on Mapping Formants into Pitch Class

The formants of the top-5 significant and reliable phonetic segments are extracted to estimate the maximum likelihood of Pitch Class among the three predefined pentatonic modes. Firstly, the two rhyming words and other three longest sounds are selected from all phonetic segments of the poem recitation. Secondly, formants in the vowels of the five words are analyzed with *Praat*, a free software for acoustic analysis (by Paul Boersma and David Weenink, Institute of Phonetic Sciences, University of Amsterdam). Then, each formant is converted into its approximate pitch based on the following equation to map a pitch's fundamental frequency *f* (measured in hertz) to a real number *p*

$$p = 69 + 12\log_2(f / 440) \quad (5)$$

Afterwards, the Pitch Class of the real number p is further derived with modulo of 12

$$PC = p \bmod 12 \quad (6)$$

The five derived Pitch Class are arranged in ascending order. Thirdly, Pitch Class, $PC(p_1, p_2, \ldots, p_n)$, is transformed to Interval Class, $IC(i_1, i_2, \ldots, i_n)$, where

$$i_k = (p_{k+1} - p_k + 12)\bmod 12; \;\; if \;\; k+1 > n \;\; then \;\; p_{k+1} = p_1. \quad (7)$$

The dissimilarities (distance) between the Interval Class of the poem and the Interval Class of the three predefined pentatonic modes are compared by Euclidean Distance. The Euclidean Distance between points $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$, in Euclidean *n*-space, is defined as:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (8)$$

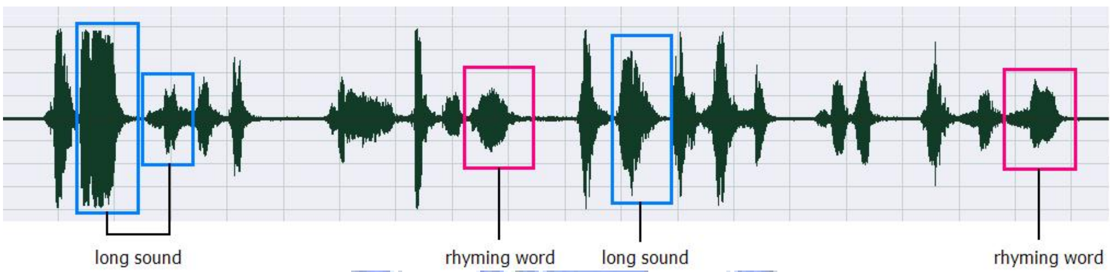
Finally, the Pitch Class with minimum Euclidean Distance is then selected as the best suitable mode for the particular poem. See the following two examples.

1st Example: *"Love Seed"* by Wang Wei

step1. Select the top-5 phonetic segments

Table 11 Significant phonetic segments extraction in "Love Seed"

Phonetic Segments of the Poem "Love Seed" in Waveform[1] Display



Text	*Pronunciation[2]*	*Length (s)*	*Significance*
紅	hong5	0.32	
豆	**tou7**	**0.47**	**long sound**
生	**seng**	**0.45**	**long sound**
南	lam5	0.27	
國	kok	0.23	
春	chhun	0.27	
來	lai5	0.38	
發	hoat	0.28	
幾	ki2	0.25	
枝	**ki**	**0.47**	**rhyming word**
願	goan7	0.34	

[1] The digital speech file comes from National Digital Archives Program, TAIWAN (recited by 洪澤南); *URL: http://dlm.ntu.edu.tw/Education/94Web/7/index.html*

[2] The pronunciation, most recited in "Southern Min" by the Roman Pinyin, derives from 羅鳳珠-中華典籍網路資料中心-唐詩三百首; *URL: http://cls.admin.yzu.edu.tw/300/Home.htm*

君	**Kun**	**0.50**	**long sound**
多	To	0.27	
采	chhai2	0.33	
擷	Khiat	0.30	
此	chhu2	0.33	
物	but8	0.27	
最	choe3	0.36	
相	Siong	0.42	
思	**Su**	**0.55**	**rhyming word**

step2. Convert formants of vowels into Pitch Class

Table 12 Significant formants in "Love Seed"

Text	*Pronunciation*	*F1*	*F2*	*F3*	*Approximate Pitch*	*PC*
豆	tou7	**581**	705	2955	D	2
生	sin	**364**	2101	2825	F#	6
枝	chi	**437**	2479	3196	A	9
君	kun	**396**	947	2777	G	7
思	su	**295**	2211	3139	D	2

step3. Calculate the Interval Class dissimilarity

Transform Pitch Classes PC_X (2, 2, 6, 7, 9), PC_C (0, 2, 4, 7, 9), PC_J (0, 2, 3, 7, 8) and PC_B (0, 1, 3, 7, 8) to Interval Classes IC_X (0, 4, 1, 2, 5), IC_C (2, 2, 3, 2, 3), IC_J (2, 1, 4, 1, 4) and IC_B (1, 2, 4, 1, 4). Compute the Euclidean Distance between IC_X (0, 4, 1, 2, 5) and IC_C (2, 2, 3, 2, 3), IC_J (2, 1, 4, 1, 4), and IC_B (1, 2, 4, 1, 4) as E_C, E_J, and E_B, respectively.

$$E_C = \sqrt{(0-2)^2 + (4-2)^2 + (1-3)^2 + (2-2)^2 + (5-3)^2} = 4$$

$$E_J = \sqrt{(0-2)^2 + (4-1)^2 + (1-4)^2 + (2-1)^2 + (5-4)^2} = 2\sqrt{6}$$

$$E_B = \sqrt{(0-1)^2 + (4-2)^2 + (1-4)^2 + (2-1)^2 + (5-4)^2} = 4$$

The similarity result is: Chinese Pentatonic ≈ Balinese Pelog > Japanese Hirajoshi. Thus, both PC_C and PC_B are recommended for the poem "Love Seed".

2

2nd ***Example:*** "*Quiet Night Thoughts*" by Li Po

step1. Select the top-5 phonetic segments

Table 13 Significant phonetic segments extraction in "Quiet Night Thoughts"

Phonetic Segments of the Poem "Quiet Night Thoughts" in Waveform Display

Text	*Pronunciation*	*Length (s)*	*Significance*
床	chhong5	0.33	
前	chian5	0.57	
明	beng5	0.47	
月	goat8	0.24	
光	**kong**	**0.58**	**rhyming word**
疑	gi5	0.32	
是	**si7**	**0.60**	**long sound**
地	ti7	0.43	
上	siong7	0.34	
霜	**song**	**0.51**	**rhyming word**
舉	ku2	0.22	
頭	thou5	0.52	
望	bong7	0.47	
明	beng5	0.40	
月	goat8	0.40	
低	te	0.28	
頭	**thou5**	**0.62**	**long sound**

思	su	0.56	
故	kou3	0.34	
鄉	**hiong**	**0.60**	**rhyming word**

step2. Convert formants of vowels into Pitch Class

Table 14 Significant formants in "Quiet Night Thoughts"

Text	*Pronunciation*	*F1*	*F2*	*F3*	*Approximate Pitch*	*PC*
光	kong	**622**	1781	2593	D#	3
是	si7	**437**	1034	2665	A	9
霜	song	**541**	825	3025	C#	1
頭	thou5	**347**	879	2707	F	5
鄉	hiong	**489**	1181	2429	B	11

step3. Calculate the Pitch Class dissimilarity

Transform Pitch Classes PC_X (1, 3, 5, 9, 11), PC_C (0, 2, 4, 7, 9), PC_J (0, 2, 3, 7, 8) and PC_B (0, 1, 3, 7, 8) to Interval Classes IC_X (2, 2, 4, 2, 2), IC_C (2, 2, 3, 2, 3), IC_J (2, 1, 4, 1, 4) and IC_B (1, 2, 4, 1, 4). Compute the Euclidean Distance between IC_X (0, 4, 1, 2, 5) and IC_C (2, 2, 3, 2, 3), IC_J (2, 1, 4, 1, 4), and IC_B (1, 2, 4, 1, 4) as E_C, E_J, and E_B, respectively.

$$E_C = \sqrt{(2-2)^2 + (2-2)^2 + (4-3)^2 + (2-2)^2 + (2-3)^2} = \sqrt{2}$$

$$E_J = \sqrt{(2-2)^2 + (2-1)^2 + (4-4)^2 + (2-1)^2 + (2-4)^2} = \sqrt{6}$$

$$E_B = \sqrt{(2-1)^2 + (2-2)^2 + (4-4)^2 + (2-1)^2 + (2-4)^2} = \sqrt{6}$$

The similarity result is: Chinese Pentatonic > Japanese Hirajoshi $\approx$ Balinese Pelog. Thus, PC_C is best recommended for the poem "Quiet Night Thoughts".

d. Utilize harmonic series in sonority construction

Sonority is the resultant of two or more musical sounds combined simultaneously, different from melody, which comprises of two or more successive pitches as a temporal sequence. The harmonic series is utilized as a reference for sonority construction because even an untrained user can intuitively and conveniently detect by ear whether the sonority sounds harmonic or inharmonic. Taking the frequency of C1 (32.703 Hz) as an arbitrary fundamental frequency, frequencies of tones sharing the same relationships as harmonics in this harmonic series over the fundamental is calculated in a true overtone series as shown in Fig. 14. The indications added to individual notes imply each tone’s deviation in cents based on 1200 cents per octave standard.

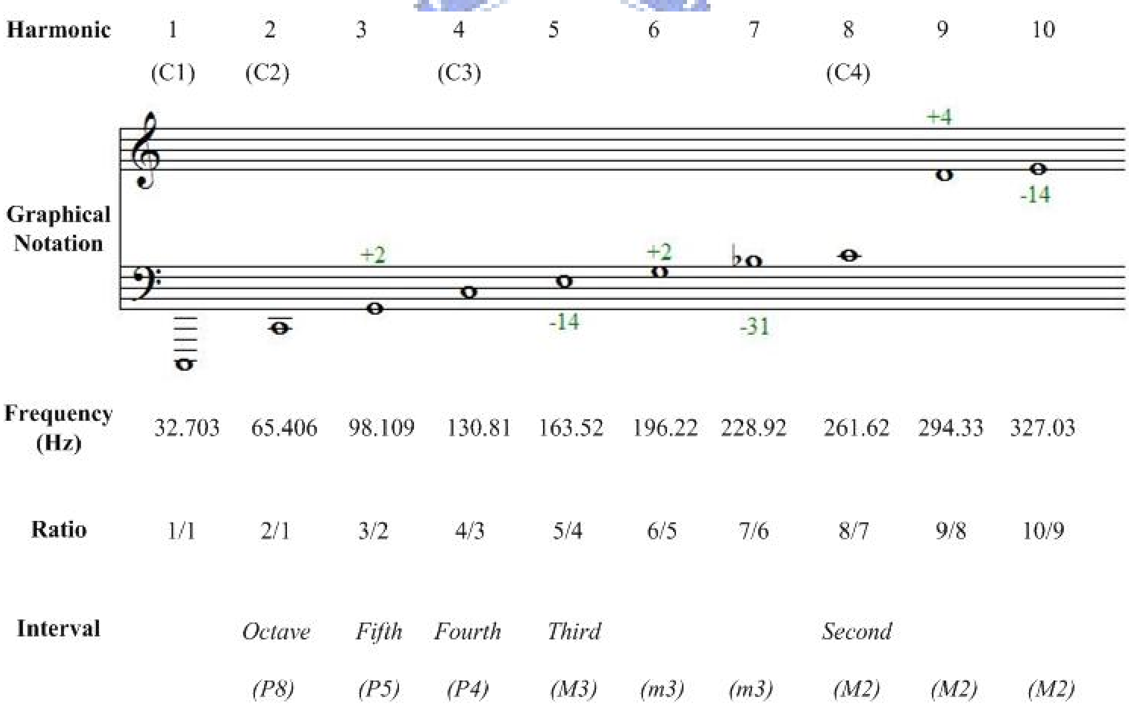


Fig. 14 The harmonic series over the fundamental frequency C1 (32.703 Hz)

The higher up the harmonic series, the more dissonant the sonority becomes. If the prosody intonation equals Ping then the sonority consists of intervals between lower successive harmonic series such as the P8, P5, and P4 (i.e., Perfect 8^{th}, 5^{th}, and 4^{th}); if the prosody intonation equals Ze then the sonority consists of intervals between higher successive harmonic series such as the M3, and m3 (i.e., Major, and minor 3rd).

3.1.3 System Architecture

Fig. 15 and Fig. 16 illustrate the system flow chart and the system architecture of Tx2Ms, namely the text-to-music conversion of Chinese Classical Poetry.

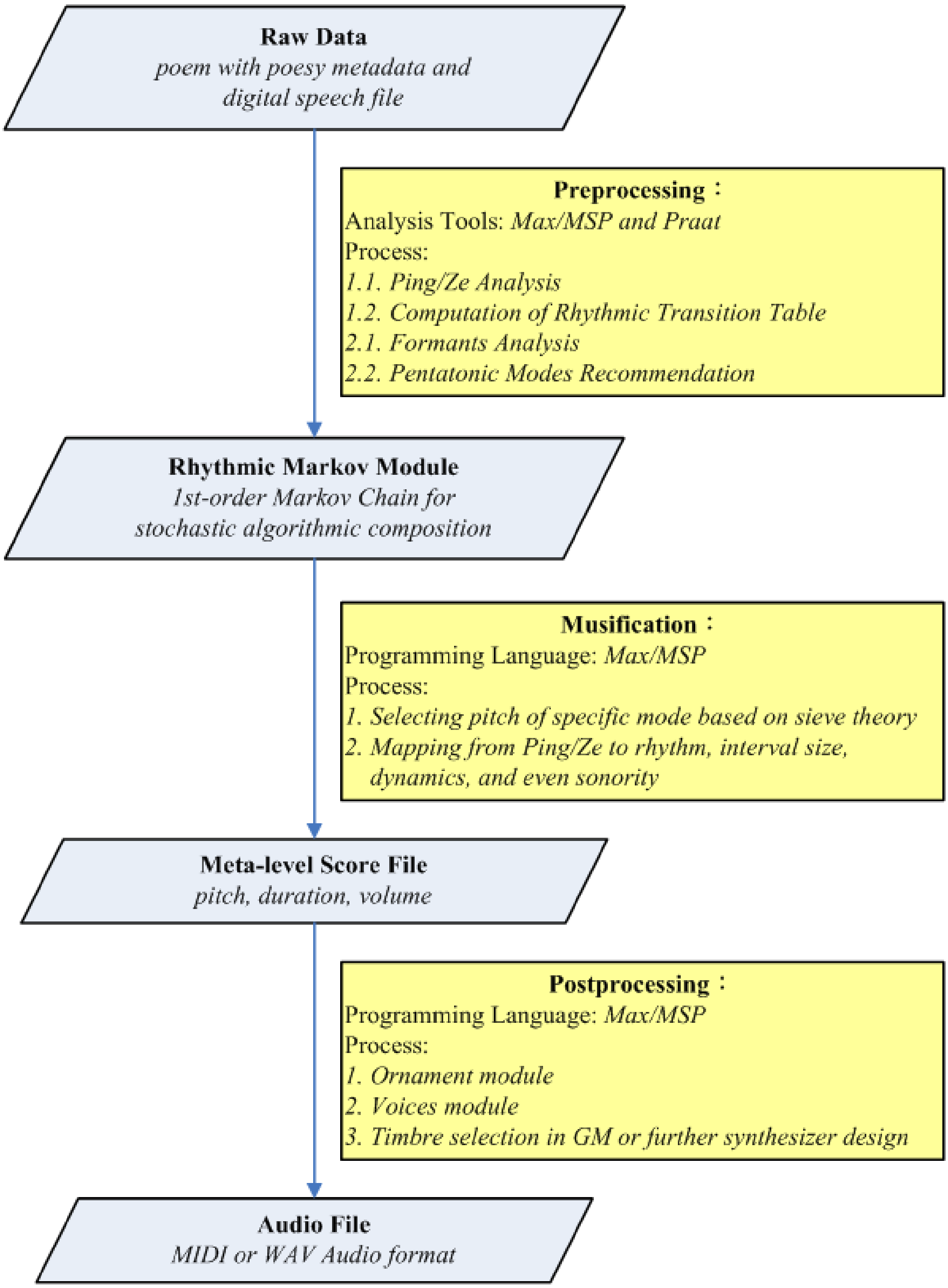


Fig. 15 System flow chart of Tx2Ms

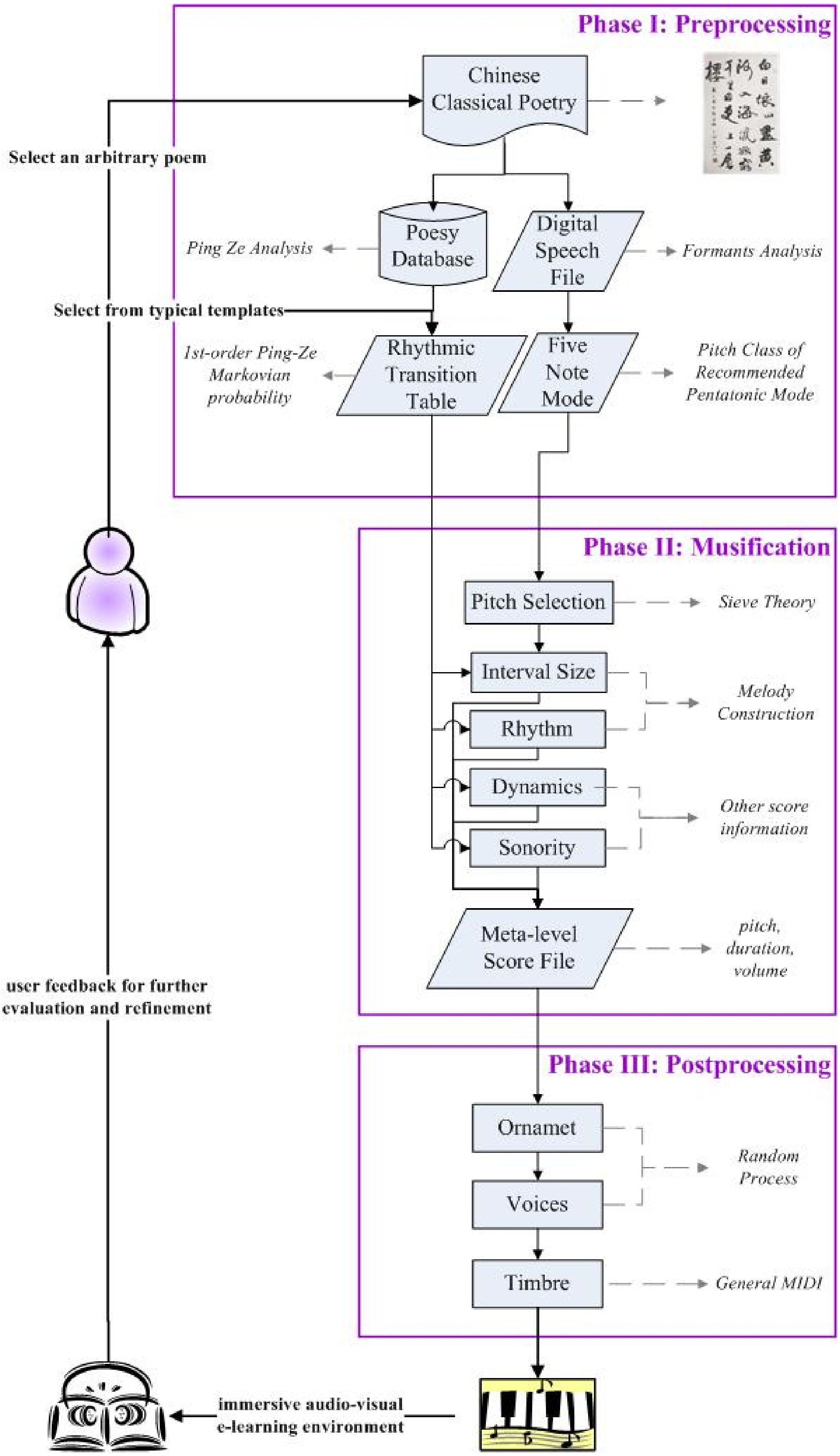


Fig. 16 System architecture of Tx2Ms

In Phase II, the best suitable pentatonic mode is recommended according to the formants of prosody intonation. The mapping rules in the Musification procedure are as shown in Fig. 17.

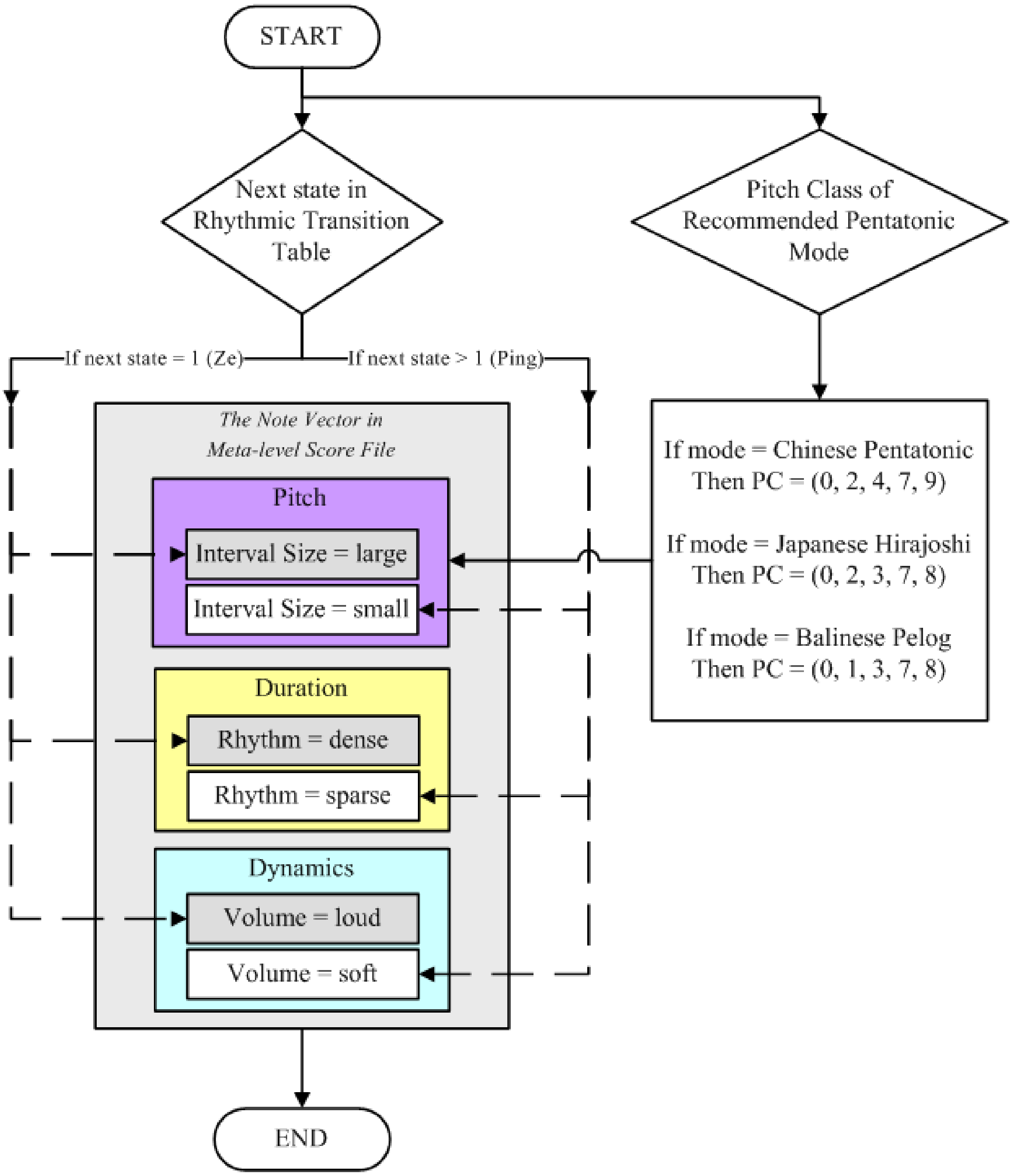


Fig. 17 The Musification Phase in Tx2Ms

3.2 Im2Ms (image-to-music mapping of Chinese Calligraphy Painting)

Throughout this section, we provide a mechanism of mapping calligraphy data onto appropriate sound features along with an overview of the system architecture.

3.2.1 Mapping Recipe of Im2Ms

In most practical applications, a raw data image is hardly observed any useful information. However, the preprocessing of an image contributes to features extraction in image analysis (as shown in Fig. 18).

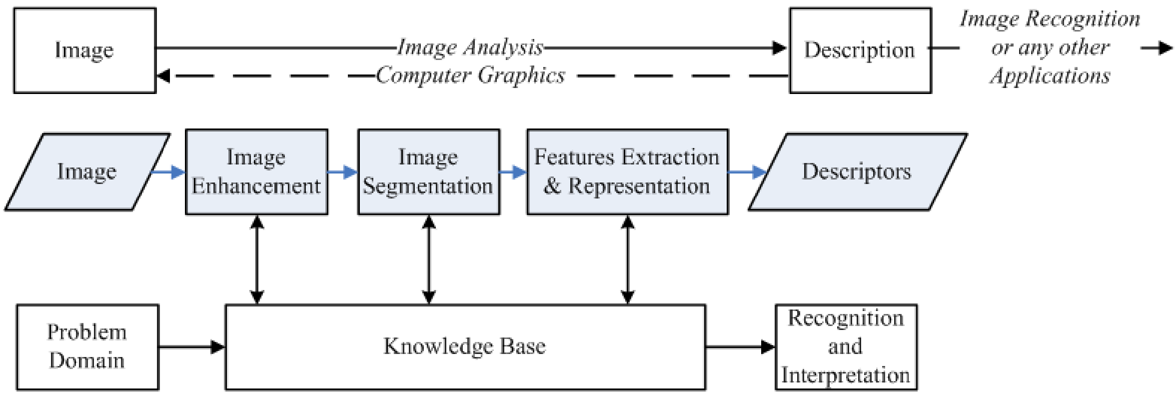


Fig. 18 Typical process in image analysis

Since a Chinese character is regarded as an image in this study, we are supposed to explore the abstract elements which comprise a picture from a Chinese character. Every image has its external frame and internal content, where the former means the explicit shape characteristics (i.e., contour information, gesture of lines) and the later means the implicit structural features (i.e., skeleton information, end points).

Basically we convert a Chinese character to sound according to its shaping frame of pixels. Furthermore, the sound is characterized by its structural component of content implied in the Chinese character (as shown in Fig. 19).

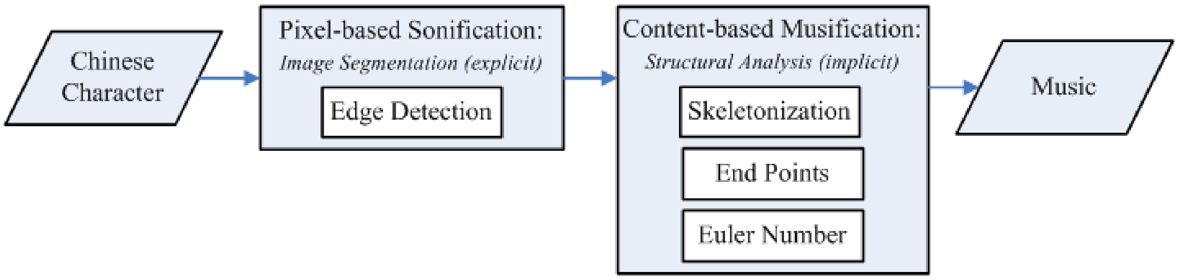


Fig. 19 Image analysis of a Chinese character

From the micro perspective, the smallest unit of an image in computer vision has information about its RGB (true-color), intensity (gray-level), position (X-Y), etc. According to the traditional spectrograph display of sounds, the two-dimensional axes are one for frequency and the other for time. Besides, concerning the human behavioral habits of writing a Chinese character (where most strokes are basically from top left towards bottom right corner), two-way scanning methods (i.e., from left to right & from top to bottom) are both adopted. Moreover, from the top-to-down perspective, each image as a whole can be analyzed by its contour, shape, etc. Table 15 shows the mapping between image and music parameters.

Table 15 Mappings from image information to musical parameters

Parameters Mapping			***Music***			
			Pitch	*Dynamics*	*Tempo*	*Timbre*
Image	*Position*	*Right, Top*	High	NULL		

Edge		*Left, Bottom*	Low			
	Intensity	*Dark*	NULL	Loud	NULL	
		Bright		Soft		
Image Structure	*End Point*	*More*	NULL		Slower	NULL
		Less			Faster	
	Euler Number	*Non Positive*	NULL			Smooth
		Positive				Sharp

3.2.2 Preliminaries of Im2Ms

a. Segmentation

Human vision is very good at edge detection so that edge detection is one of the most essential tasks in image analysis. Edge detection is extensively used in image segmentation since edges characterize object boundaries. Representing an image by its edges has the advantage that the amount of original image data is significantly reduced and useless information is filtered out, while preserving most of the important structural properties in an image. In typical image, edges are places with strong intensity contrast. An edge is a jump in intensity from one pixel to the next, where drastic change occurs in gray level over a small spatial distance (e.g. surface color or illumination discontinuity). Hence, edges correspond to high spatial frequency components in the image signal. The majority of different methods to perform edge detection can be grouped into two categories, gradient and Laplacian. The gradient-based method detects the edges by looking for the maximum and minimum in the first derivative of the image. The Laplacian-based method finds the edges by searching for zero crossings in the second derivative of the image.

- Gradient-based Edge Detection Methods:

The gradient vector represents: (1) the direction in the n-D space along which the function increases most rapidly, and (2) the rate of the increment. Here we only consider 2D field:

$$\nabla \cong \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} \tag{9}$$

where i and j are unit vectors in the x and y directions respectively. The generalization of a 2-D function *f(x, y)* is the gradient

$$\vec{g}(x,y) \cong \nabla f(x,y) = \left(\frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j}\right) f(x,y) = f_x\vec{i} + f_y\vec{j} \tag{10}$$

The magnitude of *g(x, y)* is first computed, and is then compared to a threshold to find candidate edge points.

$$\left|\nabla f(x,y)\right| = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2} \tag{11}$$

- Laplacian-based Edge Detection Methods:

The Laplace operator is defined as the dot product (inner product) of two gradient vector operators:

$$\Delta \cong \nabla^2 \cong \nabla \cdot \nabla = \left(\frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} \right) \cdot \left(\frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} \right) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{12}$$

The generalization of a 2-D function *f(x, y)* is the gradient

$$\Delta f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{13}$$

b. Features Extraction

Based on the topologically and morphologically structural attributes of a Chinese character, two descriptors (as shown in Table 16) are utilized to produce the mapping rules of implicit structural content along with the mapping criteria of explicit frame of pixels. One is the morphological attribute — End Point, where each End Point *EP(x, y)* must satisfy the following two conditions within the eight-connected chain code (Fig. 20):

1. $EP(x, y) = 1$ (14)
2. $EP(x+1, y) + EP(x+1, y+1) + EP(x, y+1) + EP(x-1, y+1) + EP(x-1, y)$
$+ EP(x-1, y-1) + EP(x, y-1) + EP(x+1, y-1) = 1$ (15)

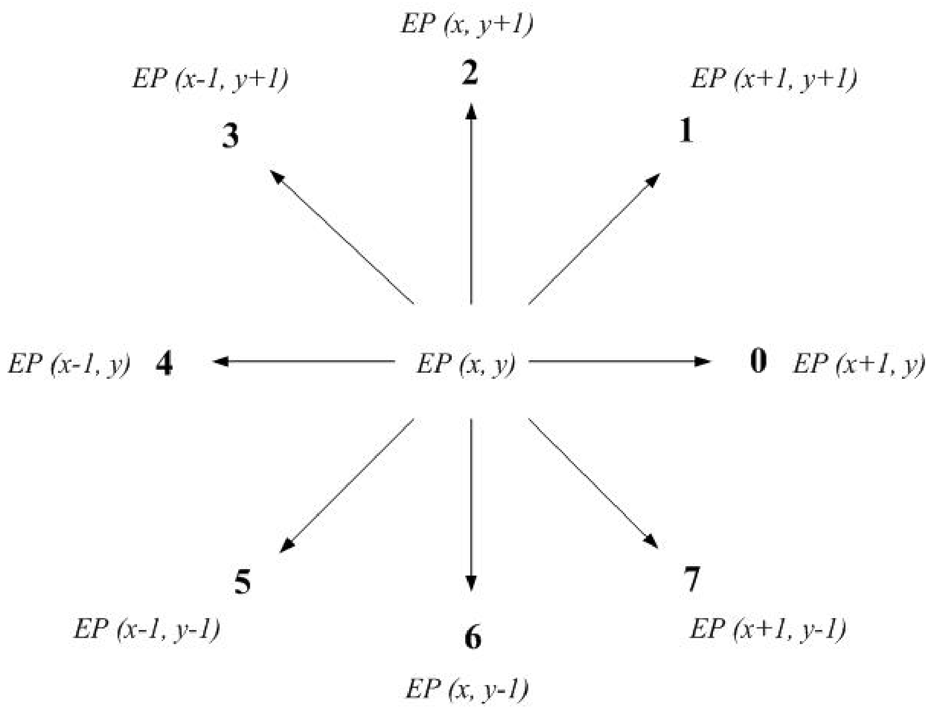
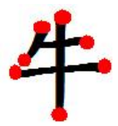


Fig. 20 End Point detection by examining the 8-connected chain code elements

The other is the topological attribute — Euler Number, which means the total number of objects in the image (i.e., C) minus the total number of holes in those objects (i.e., H), defined as:

$$E = C - H \tag{16}$$

Table 16 Two structural descriptors of a Chinese character

Structural Features	*Illustration*	*Metaphor*
End Point	 End Point = 7	Since the number of End Points implies the number of strokes of a Chinese character, the more the end points, the more complicated and higher fragmentation degree a Chinese character; the less the end

		points, the smoother a Chinese character.
Loop and Euler Number	 Loop = 3 Euler Number = -2  Loop = 2 Euler Number = -1  Loop = 1 Euler Number = 0	Since the number of Euler Number implies the number of closed regions in a Chinese character, the more the Euler Number, the higher closed degree a Chinese character. Moreover, the more End Points plus the Euler Number, the more changes of breaths during the writing and thus the slower tempo to accomplish a Chinese character.

3.2.3 System Architecture

Fig. 21 and Fig. 22 illustrate the system flow chart and the system architecture of Im2Ms, namely the image-to-music conversion of Chinese Calligraphy Painting, where Fig. 22 takes a Chinese Cursive Script, 「千」, which means “thousand”, for example.

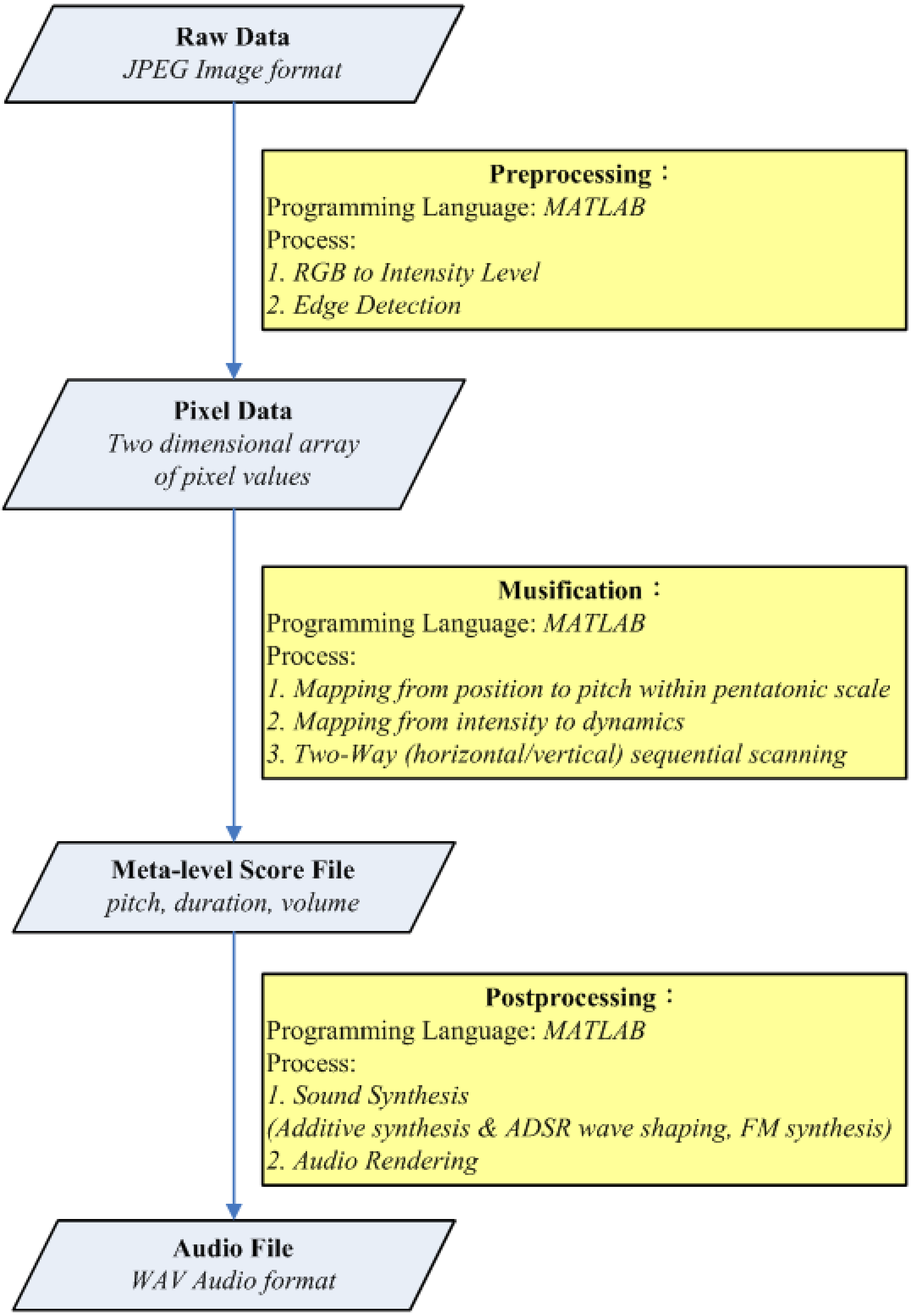


Fig. 21 System flow chart of Im2Ms

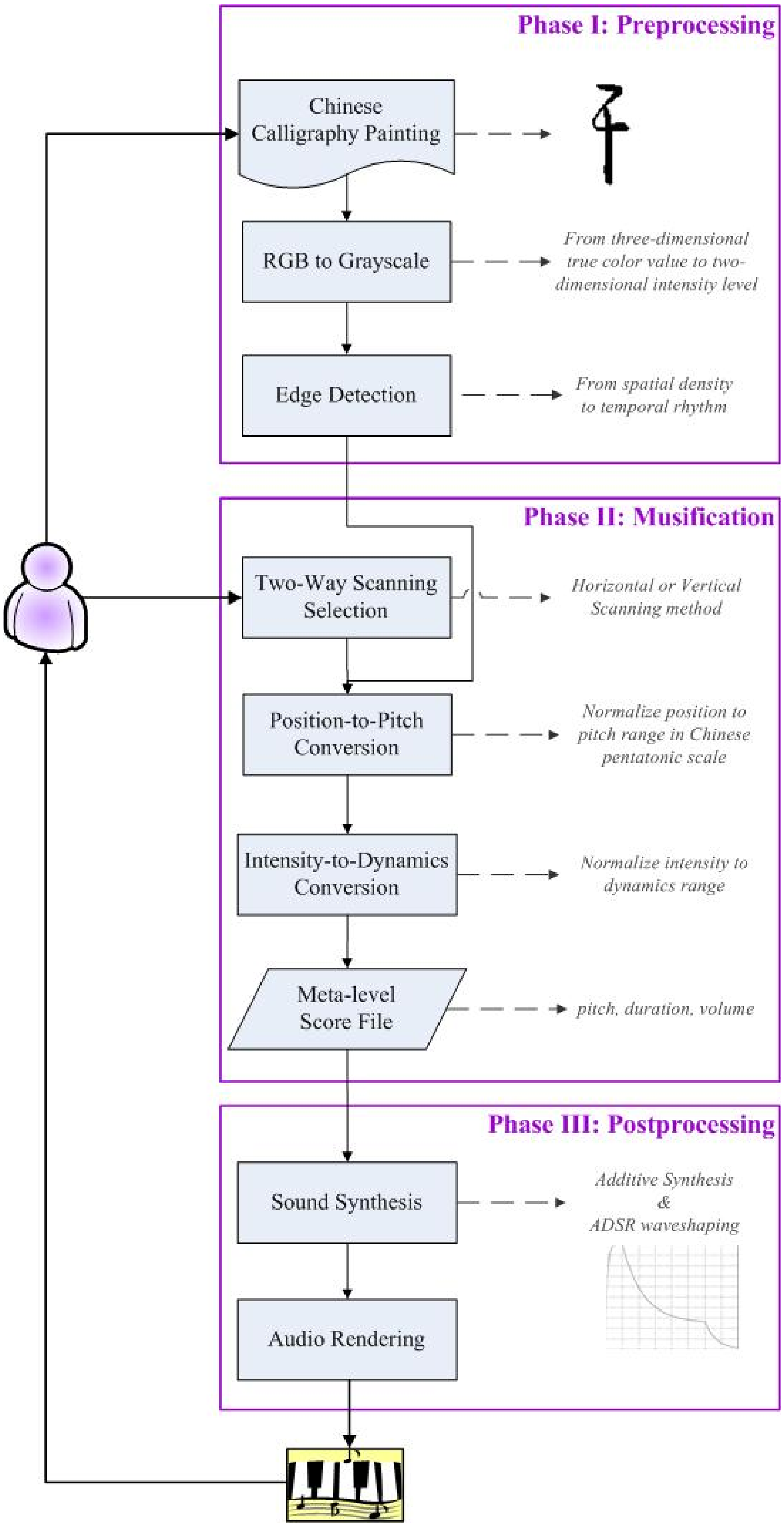


Fig. 22 System architecture of Im2Ms

Chapter 4 Experiment Results

The Sonification schemes are implemented two ways for rapid prototyping and exploration: Tx2Ms in Max/MSP, a graphical development environment for interactive computer music and multimedia, originally written by Miller Puckette and currently developed and maintained by Cyling’74, and Im2Ms in MATLAB, an environment with powerful mathematical computing especially for digital signal and image processing.

4.1 Tx2Ms (text-to-music mapping of Chinese Classical Poetry)

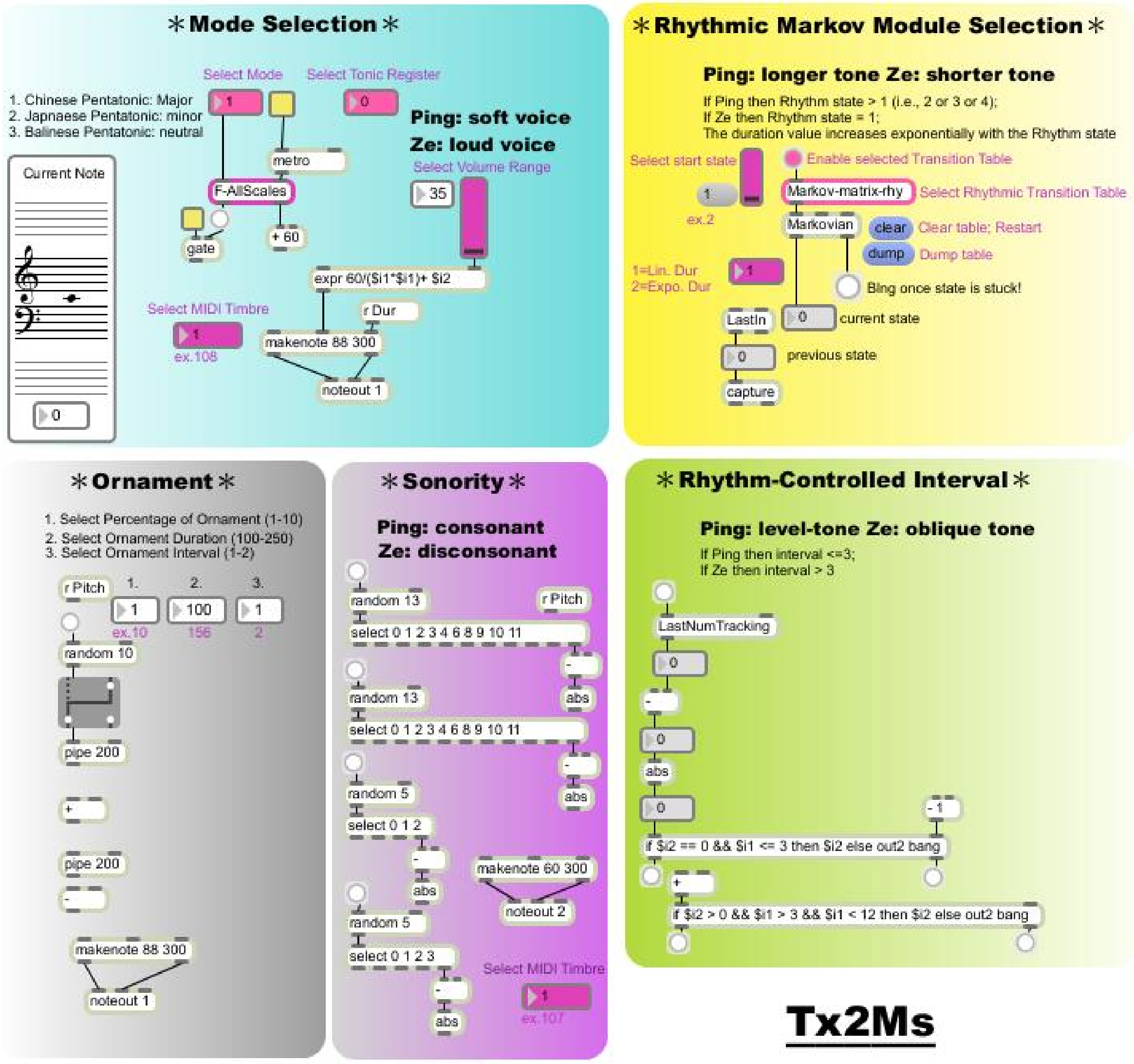


Fig. 23 Max/MSP implementation of Tx2Ms

The implementation details of Tx2Ms are shown in Fig. 23, demonstrating different user-controlled level modules:

- Mode Module: To select the recommended or user-preferred Pentatonic Mode, with Tonic and volume range selection.

- Markovian Rhythm Module: To select Ping-Ze based Rhythmic Transition Table templates for specific poem.
- Rhythm-Controlled Interval Module: the interval size is controlled by the transitional state result from Markovian Rhythm Module.
- Sonority Module: the sonority components are also controlled by the transitional state result from Markovian Rhythm Module.
- Ornament Module: user-controlled parameters for Random Process

Since Table 4 has illustrated the general Ping Ze for Five-Character Quatrain, Table 17 shows the Rhythm Sequence of each type of Five-Character Quatrain for building up the Transition Matrix.

Table 17 Rhythm sequences of four Five-Character-Quatrain types

TYPE	*Ping Ze ID*	*Ping Ze Aggregation*	*Rhythm Sequence*
TYPE I	11001 00011 00110 11001	201 02 020 201	*20102020201*
TYPE II	11100 00011 00110 11001	30 02 020 201	*302020201*
TYPE III	00011 11001 11100 00011	02 201 30 02	*040402*
TYPE IV	00110 11001	020 201	*02020402*

	11100 00011	<u>30</u> <u>02</u>	

Furthermore, Fig. 24 displays four basic templates for rhythmic matrices transferred from Ping Ze. Each message box has three numeric values, where the first and the second mean the current state and the next state, and the last value is the weighting of transition from current state to next state.

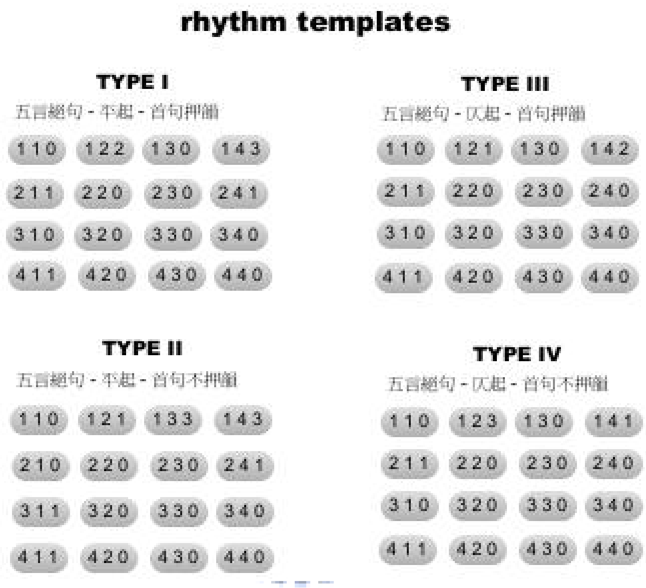


Fig. 24 Transition Matrix — four basic types of rhythm for the Five-Character-Quatrain in Classical Chinese Poetry

Fig. 25 demonstrates a Ping/Ze score of part of the state sequence output (1, 3, 1, 3, 2, 1, 2, 1, 3, 1, 2, 1, 3, 1, 1, 3, 1, 3, 1, 3, 1, 3, 1, 1, 3, 2, 1, 3, 2, 1, 3) from Rhythmic Transition Table with a sixteenth note quantization, which is exemplified from the poem in Fig. 11.

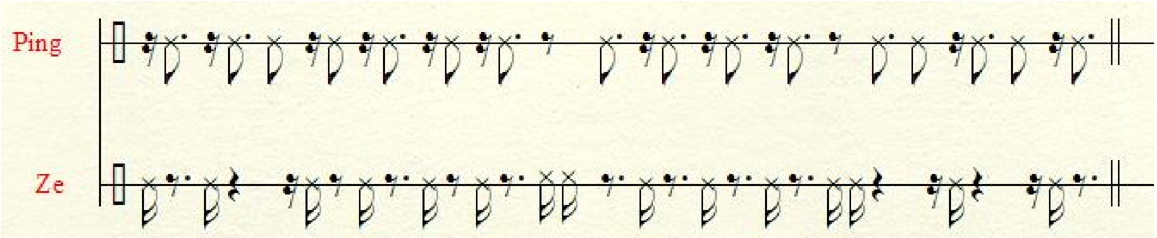


Fig. 25 Ping/Ze score of the poem “Love Seed” in Fig. 11

4.2 Im2Ms (image-to-music mapping of Chinese Calligraphy Painting)

Fig. 26 and Fig. 27 illustrate the output of Im2Ms in both waveform and spectrogram displays of the exemplified cursive script (shown in Fig. 22) by horizontal and vertical scanning method respectively.

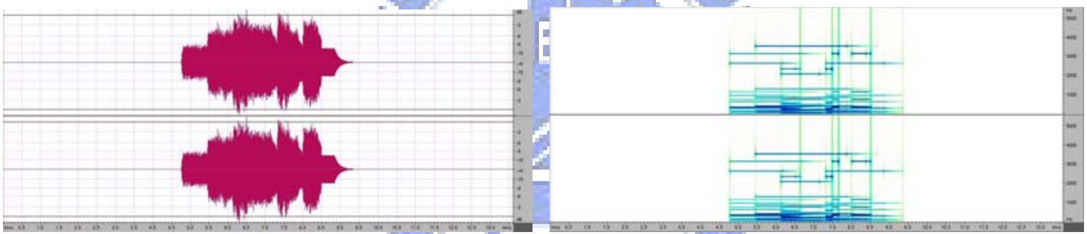


Fig. 26 Waveform and spectrogram of output in Fig. 22 by horizontal scanning

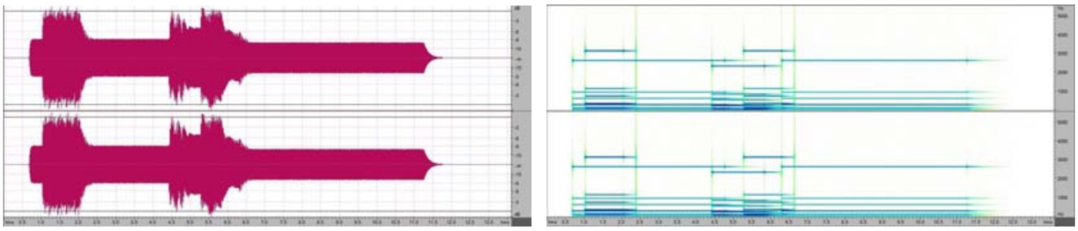


Fig. 27 Waveform and spectrogram of output in Fig. 22 by vertical scanning

As shown in Fig. 28 and Fig. 29, Im2Ms achieve its responsiveness during the Musification process by observing pitch distributions from two-way scanning mechanism. It does meet the goal in expectancy of an effective Sonification as mentioned before in Chapter 2.

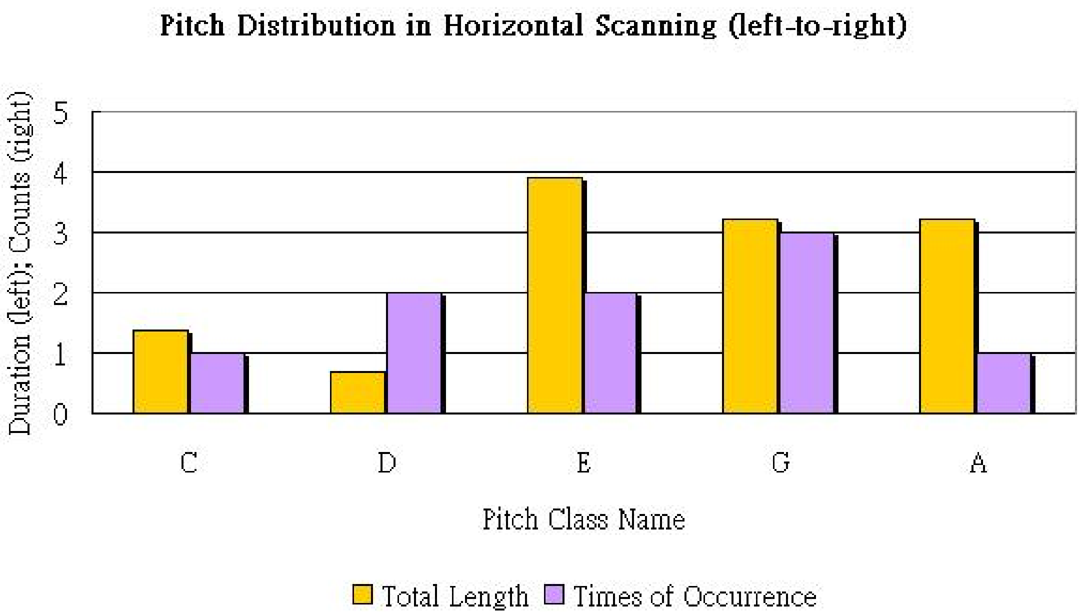


Fig. 28 Pitch Distribution of output in Fig. 22 by horizontal scanning

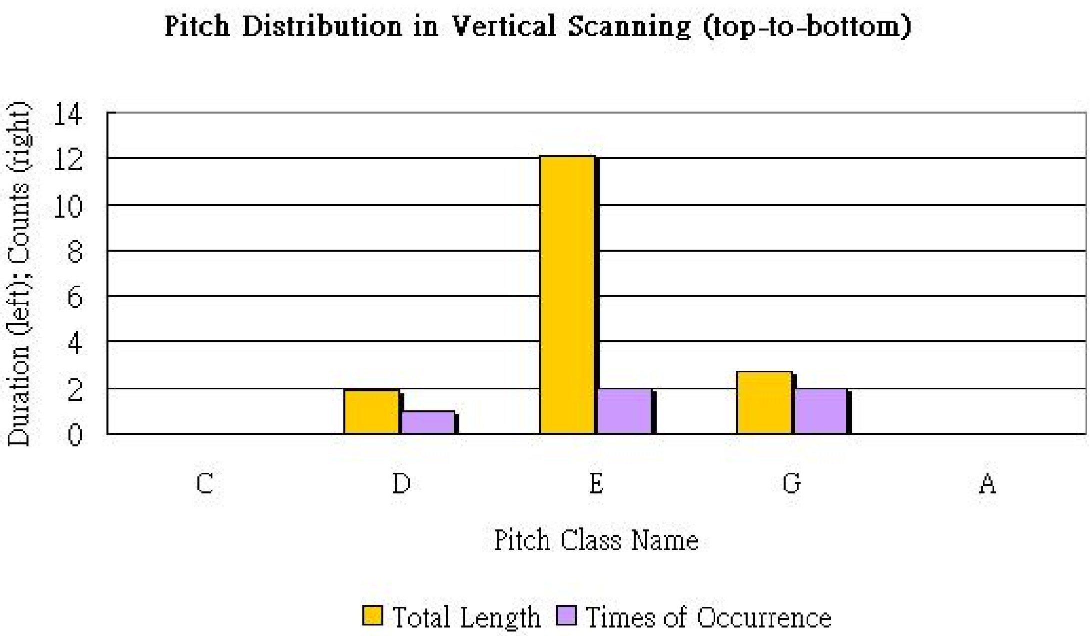


Fig. 29 Pitch Distribution of output in Fig. 22 by vertical scanning

Chapter 5 Conclusion and Future Work

Currently, the adaptive Musification prototypes designed for Chinese Classical Poetry and Chinese Calligraphy Painting are proposed. General conclusion is that the sounds produced in each experiment convey the information about the imagery state of mind and the qualitative nature of the data.

For Text-to-Music conversion:

- Not only the arrangement of text but also the pronunciation properties and the syntactic characteristics of the poem are conveyed in the music output.

For Image-to-Music conversion:

- The position-to-pitch mapping is more intuitively responsive to original visual data and easy for gestalt formation than color-to-pitch applied in the two related works. However, color could be mapped into timbre instead.

- Notwithstanding the two parameters are taken into account (i.e., position and intensity), the two-way scanning results in an extra musical effect — the sonority. To sum up, the texture of the image in both horizontally and vertically sequential scanning reflects on the sonority of the music.

Many interesting applications could be realized based on this study, such as an Audible Digital Album (let the photos sing!), or an Immersive Chinese Classical Poetry E-Learning

Platform. Nevertheless, the actual resolution obtainable with human perception of these sound representations remains to be evaluated, and the algorithmic composition throughout the Musification process need more improvements. The involvement of expertise in poetry composition (Chinese Classical Poetry Analysis), image processing (Chinese character Recognition), music composition, and even psychology is critical for its success.

Although this study has systematically investigated the logical and reasonable mappings from the degrees of freedom in the data to the parameters controlling the algorithmic composition or sound synthesis process, there are still few limitations of this study. The most obvious one is the lack of strokes sequential information in the Im2Ms. The sequential strokes of a Chinese character play a significant role in this kind of specific image as an important feature itself. Consequently, there might be an alternative demand for a real-time and interactive Musification for Chinese Calligraphy Painting. Mouse, write pad, or other related input devices could be used to obtain more image information, such as the sequence of the character, instead of simply horizontal and vertical scanning. Take the following idea for example. Since the writing sequence is based on the "arrow", the writing segments are then retrieved for sections of music, with "rest" based on the timing between the end of the last segment and the beginning of the next segment. Simply speaking, the scanning sequence is no longer the pure left-to-right or top-to-bottom, but the real-time writing strokes recorded sequentially. In this way, the image content could be mapped into music, where the vertical axis variance determines the pitch in Pentatonic Scale up or down and the horizontal axis variance determines the timbre (see Fig. 30).

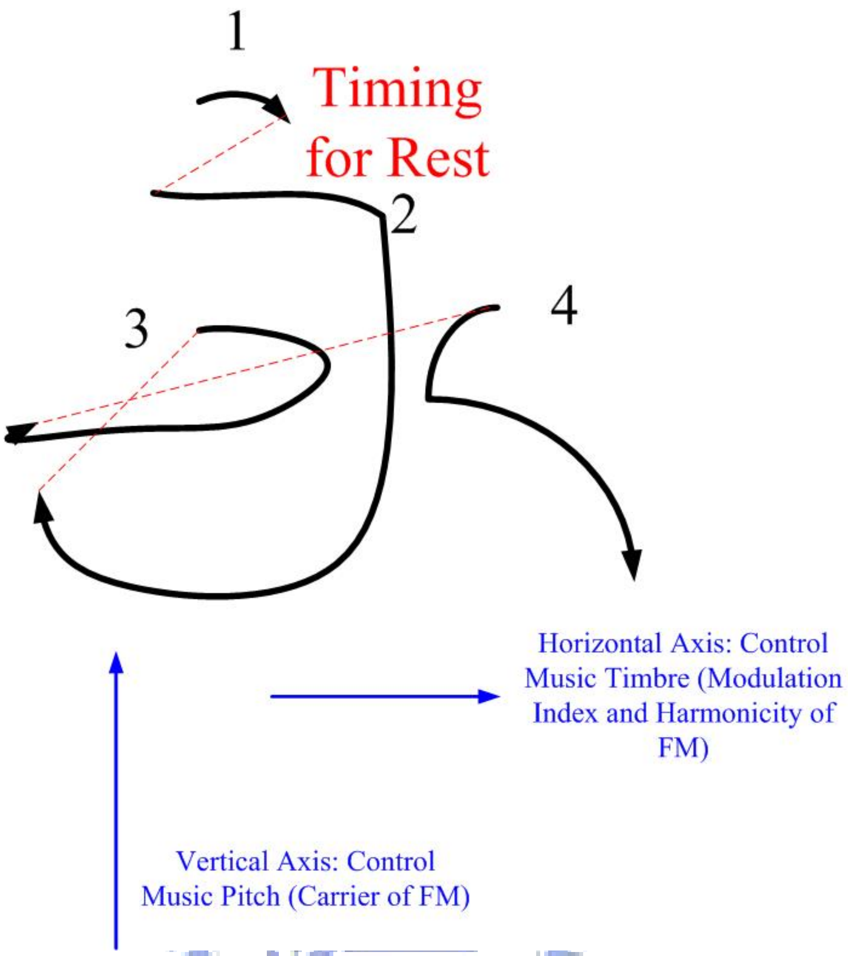


Fig. 30 An exemplified Calligraphy Musification algorithm with FM

References

[1] 王威欽。2005。運用影像分析實現電腦音樂創作之研究。台北市：國立台北藝術大學科技藝術研究所碩士論文。

[2] 許清雲。1997。《近體詩創作理論》。台北市：洪葉文化。

[3] 陳淑英。1987。《視聽教育與教育工學》。台北市：文景。

[4] 陳新雄。2004。《詩詞作法入門》。台北市：五南。

[5] 楊哲青。2004。詩作風格知識庫之研究－以蘇軾近體詩為例。新竹市：國立交通大學理學院專班網路學習組碩士論文。

[6] Ballas, J. A. 1994. Delivery of Information Through Sound. In *Auditory Display: Sonification, Audification and Auditory Interfaces*, ed. Gregory Kramer, 79–94. Reading, MA: Addison-Wesley.

[7] Barrass, S., and Kramer, G. 1999. Using sonification. *Multimedia Systems* 7, no. 1 (January), http://www.springerlink.com/content/xd19ftjpfe3pb30l/fulltext.pdf (accessed May 9, 2009).

[8] Blattner, M. M., Papp III, A. L., and Glinert, E. P. 1994. Sonic Enhancement of Two-Dimensional Graphics Displays. In *Auditory Display: Sonification, Audification and Auditory Interfaces*, ed. Gregory Kramer, 447–470. Reading, MA: Addison-Wesley.

[9] Clement, B. J. 1998. Learning Harmonic Progression Using Markov Models. http://www-lrn.cs.umass.edu/lab-lunch/papers/clement98learning.pdf (accessed May 10, 2009).

[10] Edlund, J. 2004. The Virtues of the Musifier: A Matter of View. InterAmus Music Systems. http://www.interamus.com/techTalk/musificationAndView.html (accessed May 9, 2009).

[11] Ekman, R. 2003. Coagula – Industrial Strength Color–Note Organ. http://hem.passagen.se/rasmuse/Coagula.htm (accessed May 9, 2009).

[12] Farbood, M., and Schoner, B. 2001. Analysis and Synthesis of Palestrina-Style Counterpoint Using Markov Chains. Paper presented at the International Computer Music Conference, September 18-22, in Havana, Cuba.

[13] Franz, D. M. 1998. *Markov Chains as Tools for Jazz Improvisation Analysis.* Master thesis, Virginia Polytechnic Institute and State University.

[14] Kandinsky, W. 1977. *Concerning the Spiritual in Art.* Trans. M.T.H. Sadler. New York: Dover Publications.

[15] Kaper, H. G., Wiebel, E., and Tipei, S. 1999. Data Sonification and Sound Visualization. *Computing in Science and Engineering* 1, no. 4 (July/August), http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=774840&isnumber=16814 (accessed May 9, 2009).

[16] Kramer, G. 1994a. An Introduction to Auditory Display. In *Auditory Display: Sonification, Audification and Auditory Interfaces*, ed. Gregory Kramer, 1–77. Reading, MA: Addison-Wesley.

[17] Kramer, G. 1994b. Some Organizing Principles for Representing Data with Sound. In *Auditory Display: Sonification, Audification and Auditory Interfaces*, ed. Gregory Kramer, 185–221. Reading, MA:

Addison-Wesley.
[18] Kramer, G. et al. 1999. Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. http://www.icad.org/websiteV2.0/References/nsf.html (accessed May 9, 2009).
[19] Marino, G., Serra, M.-H., and Raczinski, J.-M. 1993. The UPIC System: Origins and Innovations. *Perspectives of New Music* 31, no. 1 (Winter), http://www.jstor.org/pss/833053 (accessed May 10, 2009).
[20] Meijer, P. B. L. 1992. An Experimental System for Auditory Image Representations*.* *IEEE Transactions on Biomedical Engineering* 39, no. 2 (February), http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=121642&isnumber=3463 (accessed May 9, 2009).
[21] Moore, F. R. 1990. Chapter Five: Composing. In *Elements of Computer Music*. Englewood Cliffs, NJ: Prentice-Hall.
[22] Osmanovic, N., Hrustemovic, N., and Myler, H. R. 2003. A Testbed for Auralization of Graphic Art. Paper presented at the IEEE Region V 2003 Annual Technical Conference, April 11, in New Orleans, Louisiana.
[23] Saue, S. 2000. A Model for Interaction in Exploratory Sonification Displays. Paper presented at the International Conference on Auditory Display, April 2-5, in Atlanta, Georgia.
[24] Scaletti, C. 1994. Sound Synthesis Algorithms for Auditory Data Representations. In *Auditory Display: Sonification, Audification and Auditory Interfaces*, ed. Gregory Kramer, 223–251. Reading, MA: Addison-Wesley.
[25] Winsor, P. 1992. *Automated Music Composition*. Denton, TX: University of North Texas Press.

Appendix I

The following are translations of the two poems (Five-Character Quatrain) exemplified in this paper, excerpted from the book “*中英對照讀唐詩宋詞*”, written by 施穎洲.

相思

Poem: Love Seed

王維（唐）

Poet: Wang Wei (Tang Dynasty)

紅豆生南國，

Red beans grow in the southern land.

春來發幾枝？

How many shoots are there in spring?

願君多采擷，

Pray gather them till full your hand.

此物最相思。

Recalling love best is this thing.

靜夜思

Poem: Quiet Night Thoughts

李白（唐）

Poet: Li Po (Tang) Dynasty

床前明月光，

Before my bed a moonlight land,

疑是地上霜。

I thought frost had come on the sand.

舉頭望明月，

Head raised, I gaze at the bright moon;

低頭思故鄉。

Head bowed, I think of my homeland.

Appendix II

The following is photocopy of four Chinese characters in cursive script, which means sun, moon, river, and mountain in left-to-right order, excerpted from *Master of Cursive Calligraphy: the Art of Yu You-ren's Calligraphy*, 115.