Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc/

An Innovative Methodology of the Algorithmic Composition Knowledge Base for the Chinese Calligraphy Painting Interaction: from Sonification to Musification

Chih-Fang Huang[a,*] and Jenny Ren[b]

[a] *Department of Health and Marketing, Kainan Univrsity, Taiwan*

[b] *Chunghwa Telecom Laboratories, Taiwan*

ARTICLE INFO

Article History:
Submitted 11.8.2021
Revised 3.17.2022
Second Revision 5.20.2022
Accepted 5.30.2022

Keywords:
Chinese Calligraphy Painting, sonification scheme, two-dimensional spatial image information, temporal music acoustics domain, visually impaired, mind and imagery

ABSTRACT

Chinese characters are remarkable for the form of art, as the Chinese 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, the Sonification scheme, Im2Ms, is proposed to extract the melody between the lines, i.e., the lines in strokes. 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 interaction. Therefore, the Sonification of 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 appreciation environment of Chinese Calligraphy Painting can be further developed.

1. Introduction

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 the form of Chinese Characters (i.e., calligraphy) through an alternative modality, hearing?” Fortunately, we found that both image and sound 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 two different kinds of medium (images and sounds) are being manipulated. (Figure 1) depicts the interaction of different display modalities of cross-disciplinary arts.

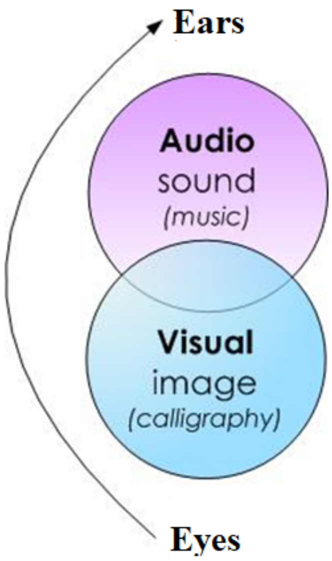


Figure 1: Three display modalities of cross-disciplinary arts in this study.

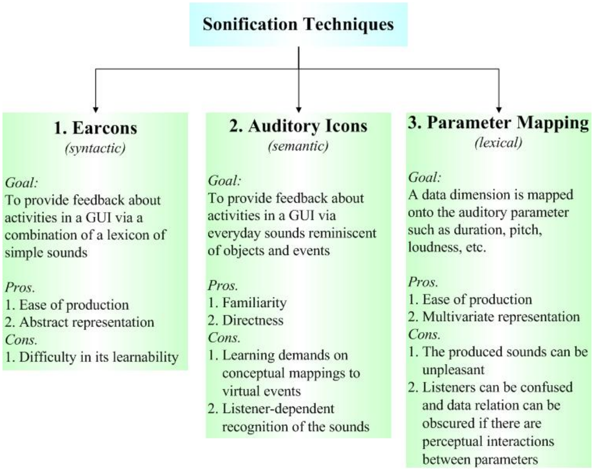
DOI reference number: 10.18293/JVLC2022-N1-015

1.1 Scenarios and Contributions

The following is the scenario in this study, which maps from image onto sound, or space to time: appreciating the Chinese Calligraphy Painting, taking the example of the “Cursive Script”. Although there has been research on mapping from image to sound, none of them is dedicated to Chinese Characters. The contributions of this study are as follows: Firstly, image is analytically transformed from 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.

1.2 Sonification

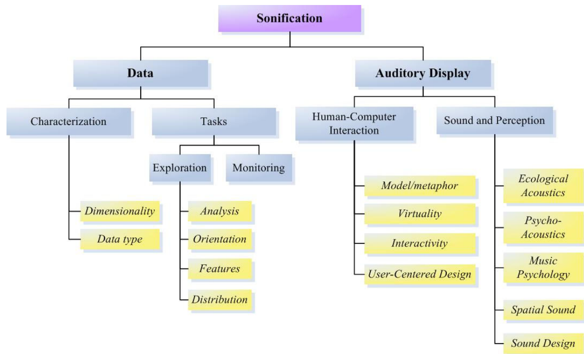
The word “Sonification” comprises the two Latin syllabus “sonus”, meaning sound, and the ending “fication”, forming nouns from verbs which are ending with ‘-fy’. Therefore, to “sonify” means to convey the information via sound. A Geiger detector can be seen as the very basic scientific example for Sonification, 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” [1]; “Sonification is the use of data to control a sound generator for the purpose of monitoring and analysis of the data” [2]; “Sonification is the transformation of data relations into perceived relations in an acoustic signal for the purpose of facilitating communication or interpretation” [3]; “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” [4]. (Figure 2) illustrates the existing Sonification techniques, which are already categorized into three types according to the mapping approach adopted: syntactic, semantic or lexical mapping [5], [6].



arranged from [5]

Figure 2: Three types of existing sonification techniques.

Besides, the fundamental elements of a Sonification are suggested in (Figure 3) from both data-centric and human-centric points of view [7], including the functionality to be identified, tasks to be performed, and several related disciplines to be worked with.



rewritten from [7]

Figure 3: Elements hierarchy of a sonification display.

Moreover, (Figure 4) illustrates the fundamental procedure about how to design a Sonification system, where the Communicative Medium is the core of Sonification [8].

rewritten from [8]

Figure 4: Schematic of an Auditory Display System.

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 as shown in (Figure 5). For functionality, the goal-oriented function of the system must be clearly defined. For utility, if the sound is ugly, people won't use it. Consequently, 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). For expectancy, evaluation (e.g., questionnaire) is needed.

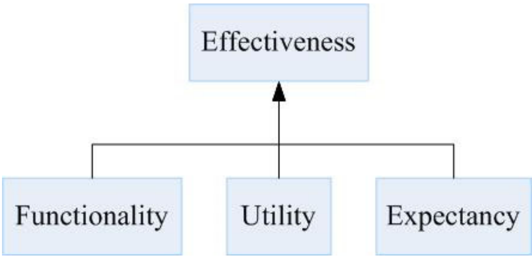


Figure 5: How to make an effective sonification?

1.3 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" [9]. However, the difference between Sonification and *Musification* lies in the fact — Music is "organized sounds", coined by French composer, Edgard Varèse [10]. 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.

1.4 Current Research Trends in Sonification

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

1.4.1 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 [11]. 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 [12], which is aimed to help scientists explore and analyze huge data sets in scientific computing.

1.4.2 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 [13]. 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 [14]. 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 [15]. 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 [16]. 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 [17], 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. 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: $\textbf{\textit{tone}} \times 2^{40} = \textbf{\textit{color}}$ [18].

1.4.3 Chinese Calligraphy Painting

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 1) 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 1: Evolution of four primary styles of Chinese calligraphy.

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

This artistic creation of Chinese Characters is rich in pictorial splendor, and deep in implicit imagery. 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.

2. Methodology

"The mapping problem" has been regarded as the essential issue of Sonification. In Chinese Calligraphy Painting, we found that the calligraphers signify their personality through the strokes and modeling of a character. In this study, the Sonification mechanisms of transforming Chinese Calligraphy Painting (image) into music is explored. First of all, the aesthetic features are extracted from Chinese Calligraphy Painting. Afterwards, the rules are applied to Im2Ms (image-to-music mapping of Chinese Calligraphy Painting) parameter-mapping mechanism. The conversion of image-to-music mapping is based on a relationship that exists among space, color and sound in human perception. (Figure 6) 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.

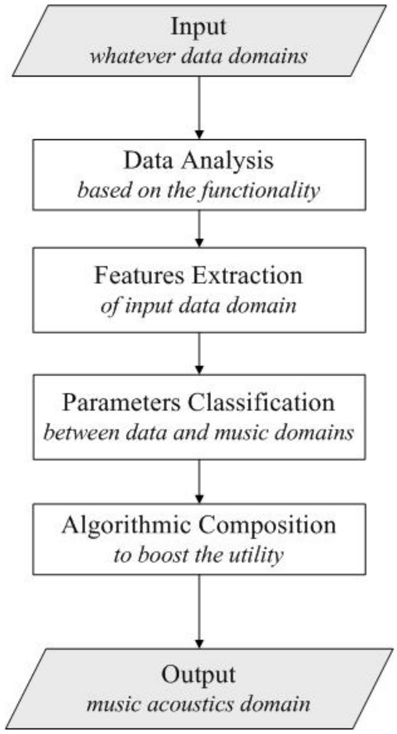


Figure 6: General paradigm of sonification.

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

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 (Figure 7).

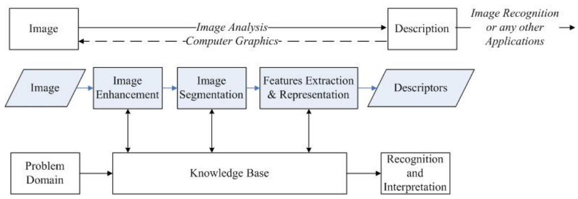


Figure 7: Typical process in image analysis.

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 (Figure 8).

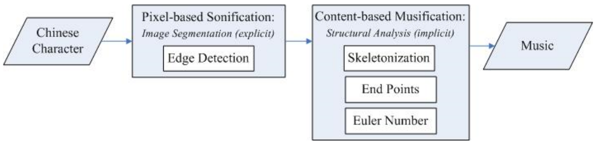


Figure 8: Image analysis of a Chinese character.

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 2) shows the mapping between image and music parameters.

Table 2: Mappings from image information to musical parameters.

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

2.2 Preliminaries of Im2Ms

2.2.1 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. For 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:

$$\nabla \cong \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} \tag{1}$$

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:

$$\vec{g}(x,y) \cong \nabla f(x,y) = \left(\frac{\partial}{\partial x}\vec{\imath} + \frac{\partial}{\partial y}\vec{\jmath}\right) f(x,y) = f_x\vec{\imath} + f_y\vec{\jmath} \tag{2}$$

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

$$|\nabla f(x,y)| = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2} \tag{3}$$

For Laplacian-based edge detection methods, the Laplace operator is defined as the dot product (inner product) of two gradient vector operators:

$$\Delta \cong \nabla^2 \cong \nabla \cdot \nabla = \left(\frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j}\right) \cdot \left(\frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j}\right) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{4}$$

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

$$\Delta f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{5}$$

2.2.2 Features Extraction

Based on the topologically and morphologically structural attributes of a Chinese Character, two descriptors, as shown in (Table 3), 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 (Figure 9):

$$EP(x,y) = 1 \tag{6}$$

$$EP(x+1,y) + EP(x+1,y+1) + EP(x,y+1) + EP(x-1,y+1) + EP(x-1,y) + EP(x-1,y-1) + EP(x,y-1) + EP(x+1,y-1) = 1 \tag{7}$$

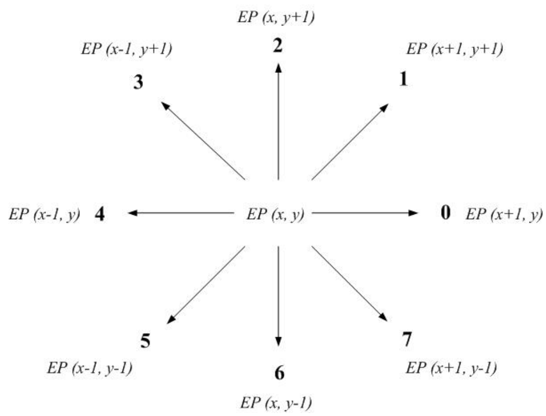
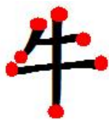


Figure 9: End Point detection by examining the 8-connected chain code elements.

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:

$$E = C - H \tag{8}$$

Table 3: Two structural descriptors of a Chinese caracter.

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	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.

	= 0	

2.2.3 System Architecture

(Figure 10) and (Figure 11) illustrate the system flow chart and the system architecture of Im2Ms, namely the image-to-music conversion of Chinese Calligraphy Painting, where Fig. 11 takes a Chinese Cursive Script, which means "thousand", for example.

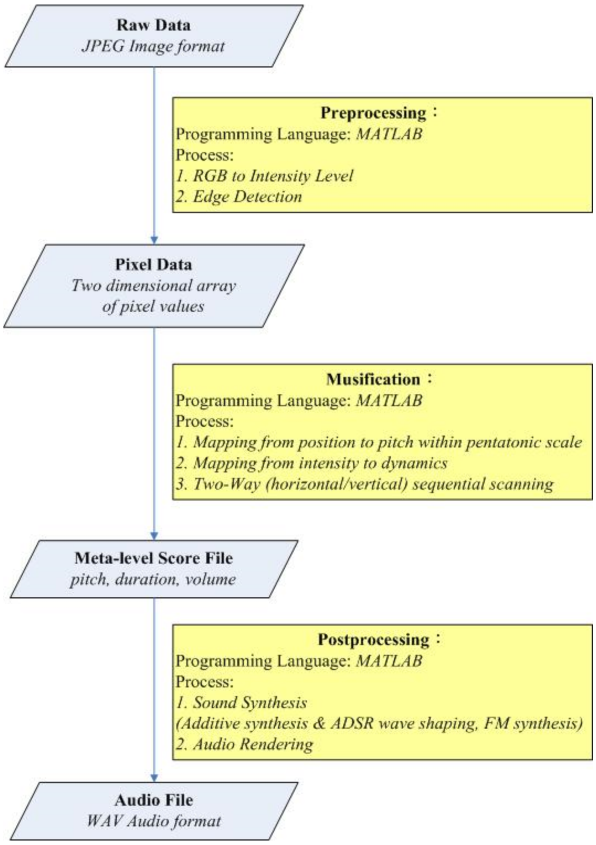


Figure 10: System flow chart of Im2Ms.

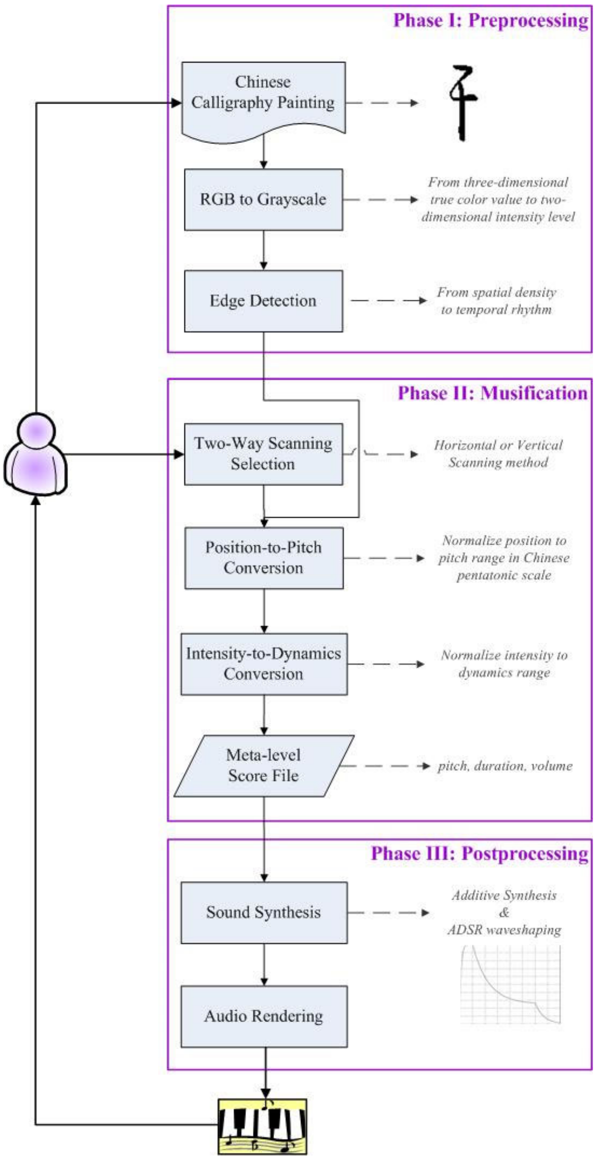


Figure 11: **System architecture of Im2Ms.**

3. Experiment Results

The Sonification scheme is implemented in MATLAB for rapid prototyping and exploration, where MATLAB is an environment with powerful mathematical computing especially for digital signal and image processing. (Figure 12) and (Figure 13) illustrate the output of Im2Ms in both waveform and spectrogram displays of the exemplified cursive script, as shown in (Figure 11), by horizontal and vertical scanning method respectively.

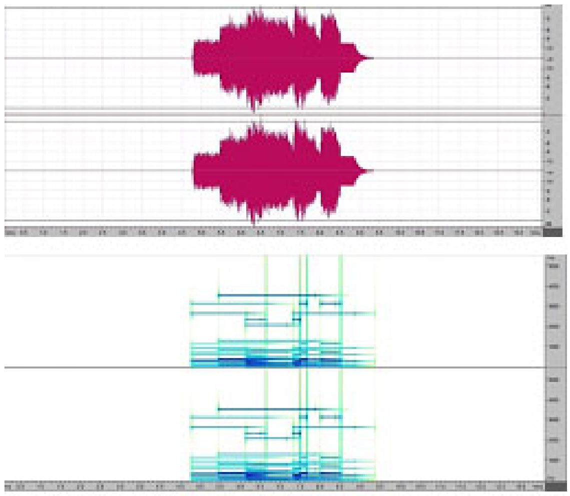


Figure 12: **Waveform and spectrogram of output in Fig. 11 by horizontal scanning.**

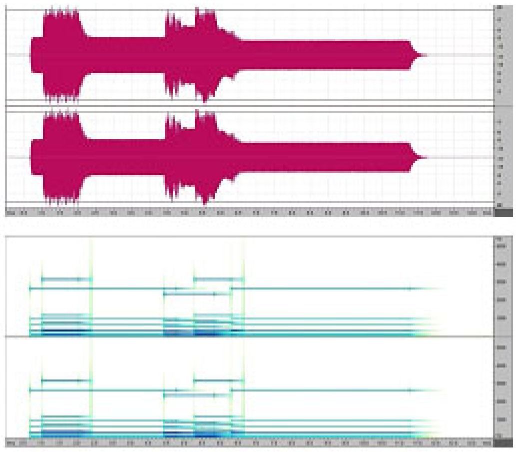


Figure 13: **Waveform and spectrogram of output in Fig. 11 by vertical scanning.**

As shown in (Figure 14) and (Figure 15), 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 the mentioned.

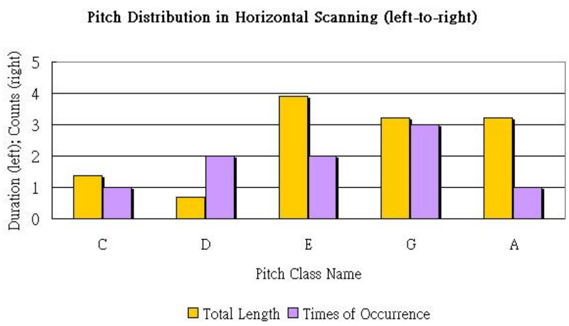


Figure 14: **Pitch Distribution of output in Fig. 11 by horizontal scanning.**

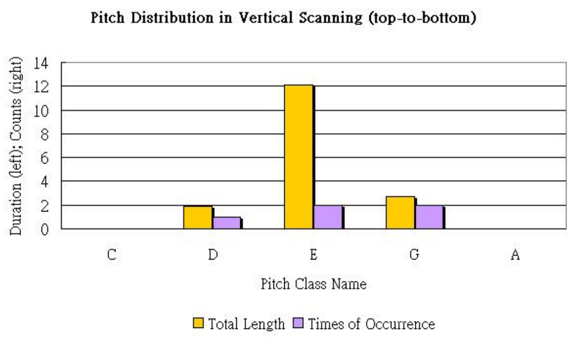


Figure 15: Pitch Distribution of output in Fig. 11 by vertical scanning.

4. Conclusion and Future Work

Currently, the adaptive Musification prototype designed for Chinese Calligraphy Painting is proposed. General conclusion is that the sounds produced convey the information about the imagery state of mind and the qualitative nature of the data. For Image-to-Music conversion, on the one hand, 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. On the other hand, 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 with interaction. Many interesting applications could be realized based on this study, such as an Audible Digital Album (let the photos sing!). 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 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, such as the FM (Frequency Modulatoin) Synthesis for the image sonification [19] and [20], 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, as shown in (Figure 16).

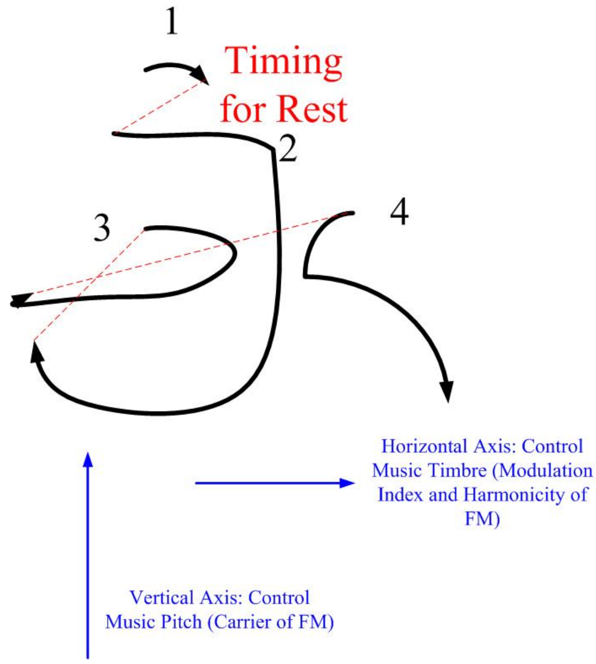


Figure 16: An exemplified Calligraphy Musification algorithm with FM.

Acknowledgement

The authors are appreciative of the support from the Ministry of Science and Technology project of Taiwan: MOST 108-2511-H-424 -001 -MY3.

References

[1] Kramer, G. (2000). "Auditory display: sonification, audification and auditory interfaces". Addison-Wesley Longman Publishing Co., Inc..

[2] Kramer, G. (1994). "Some organizing principles for representing data with sound". Auditory Display-Sonification, Audification, and Auditory Interfaces, 185-221.

[3] Kramer, G. et al. "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, 2020).

[4] Kramer, G. (2000). “Auditory display: sonification, audification and auditory interfaces”. Addison-Wesley Longman Publishing Co., Inc..

[5] Barrass, S., Kramer, G. (1999). “Using sonification. Multimedia systems”, 7(1), pp. 23-31.

[6] Blattner, M. M. (1994). “Sonic enhancement of two-dimensional graphics displays”. Auditory display-Sonification, audification and auditory interfaces, 447-470.

[7] Saue, S. (2000). “A model for interaction in exploratory sonification displays”. Georgia Institute of Technology.

[8] Kramer, G. (1994). “An introduction to auditory display. Auditory display-Sonification, audification and auditory interfaces”, pp. 1-77.

[9] Visi, F., et al. (2014). "Unfolding| Clusters: A Music and Visual Media Model of ALS Pathophysiology". Proceedings of SoniHED Conference: Sonification of Health and Environmental Data. York, United Kingdom.

[10] Richardson, R. (2005). “Edgard Varèse and the Visual Avant-Garde: A Comparative Study of Intégrales and Works of Art by Marcel Duchamp”. Doctoral dissertation, University of Cincinnati, USA.

[11] Barrass, S., Kramer, G. (1999). “Using sonification. Multimedia systems”. 7(1), pp. 23-31.

[12] Kaper, H. G., Wiebel, E., Tipei, S. (1999). “Data sonification and sound visualization”. Computing in science & engineering, 1(4), pp. 48-58.

[13] Kandinsky, W. (2012). “Concerning the spiritual in art. Courier Corporation”. Dover Publications, USA.

[14] Marino, G., Serra, M. H., Raczinski, J. M. (1993). “The UPIC system: Origins and innovations”. Perspectives of New Music, pp. 258-269.

[15] Ekman, R. (2003). “Coagula-Industrial Strength Colour Note Organ-Light Version 1.666”.

[16] Meijer, P. B. (1992). “An experimental system for auditory image representations”. IEEE transactions on biomedical engineering, 39(2), pp. 112-121.

[17] Wang, W. C. (2005) “A Study of Computer Composition on Using Image Analysis”. Master thesis, National Taipei University of the Arts, Taiwan.

[18] Osmanovic, N., Hrustemovic, N., Myler, H. R. (2003). “A testbed for auralization of graphic art”. In Annual Technical Conference IEEE Region 5, pp. 45-49, IEEE.

[19] Yeo, W. S., Berger, J. (2006). “Application of raster scanning method to image sonification, sound visualization, sound analysis and synthesis”. In Proc of the 9th Int. Conference on Digital Audio Effects, pp. 18-20.

[20] Kew, H., Stables, R. (2020). “Sonification of Spectroscopic analysis of food data using FM Synthesis”. In Audio Engineering Society Convention 149. Audio Engineering Society.