具足圆成 Until All Become One for Computer Vision and Kyma

by

Tao Li

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Data-Driven Music Performance and Composition

Dissertation Committee:

Jeffrey Stolet, Chair

Jon Bellona, Core Member

David Crumb, Core Member

Colin Ives, Institutional Representative

University of Oregon

Spring 2026

© 2026 Tao Li
This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs (United States) License.



DISSERTATION ABSTRACT

Tao Li

Doctor of Philosophy in Data-Driven Music Performance and Composition

Title: 具足圆成 Until All Become One for Computer Vision and Kyma

My dissertation composition titled *具足圆成 Until All Become One* is a quadraphonic human-machine interactive data-driven instrument composition written for computer vision and Symbolic Sound's Kyma. This dissertation consists of a performance video of the piece, affiliated files that required to perform this piece, and the text document which explains, analyzes, and discusses several aspects of this composition. These aspects include: 1) the design and implementation of the data-driven instrument used in the piece, 2) musical challenges and opportunities, 3) data mapping strategies, 4) performance techniques, 5) intentionality of spatialization design, and 6) compositional structure.

CURRICULUM VITAE

NAME OF AUTHOR: Tao Li

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
DePaul University, Chicago
Central Conservatory of Music, Beijing

DEGREES AWARDED:

Doctor of Philosophy, Data-Driven Music Performance and Composition, 2026, University of Oregon
Doctor of Philosophy, Music Composition, 2020, University of Oregon
Master of Music, Music Composition, 2013, DePaul University
Bachelor of Arts, Music Composition, 2009, Central Conservatory of Music, Beijing

AREAS OF SPECIAL INTEREST:

Electroacoustic Composition
Human-machine Interactive Composition
Music Composition
Data-driven Instrument Performance
Ensemble Performance and Conducting

PROFESSIONAL EXPERIENCE:

Graduate Employee, School of Music and Dance, University of Oregon, Eugene, 2016-2017, 2022-2026

Director, WAVE Electroacoustic Music Festival, 2026

Assistant Faculty, Summer Academy for Computer Music, University of Oregon, Eugene, 2025

Conference Adjudicator, The Society for Electro-Acoustic Music in the United States National Conference, 2024-2026

GRANTS, AWARDS, AND HONORS:

Committee for Equity and Inclusion (CEI) Grant, *WAVE* Electroacoustic Music Festival,
University of Oregon School of Music and Dance, 2026

SEAMUS Micro CREATE Grants, *WAVE* Electroacoustic Music Festival, Society for Electro-Acoustic Music in the United States, 2025

Libby Larsen Prize, *无人堪寄 Wu Ren Kan Ji* for cello and fixed media, International Search for New Music Competition International Alliance for Women in Music, 2021

PUBLICATIONS:

Li, Tao. "Song of Zhu" on *The Dragon Dream*, Aaron Pergram, bassoon. Recorded 2025. Mark Masters 57901-MCD. Released 2026, compact disc.

Li, Tao. "Qin Xiao Yi" on *Duo Per Se*, Duo Per Se. Recorded 2021. Centaur B0BT25P8KR. Released 2023, compact disc.

Li, Tao. "Lament for Theodore Antoniou" on *Lament for Theodore Antoniou* Vol. 2. Iwona Glinka, flute. Recorded 2020. Phasma-Music 032. Released 2021, compact disc.

Li, Tao. "Text Painting as a Vehicle of Narrative in Contemporary Electroacoustic Composition." *Emille* 16, (2019): 37-42.

ACKNOWLEDGMENTS

I would like to begin by express my sincere gratitude to my dissertation advisor, mentor, teacher of both art and life, Dr. Jeffrey Stolet, who has guided me with his constant encouragement, thoughtful challenges, and unique humor throughout my entire time at the University of Oregon. The wisdom he shared with me on music practice, teaching, and life has shaped me into the musician, teacher, and person I am today. I know his wisdom will continue to guide and inspire me and countless others in the years to come.

I would also like to thank Dr. Akiko Hatakeyama, who empathizes with me, encourages me, and supports me both in and out of school. Observing her teaching has greatly shaped the way I think about teaching and pedagogy. She is not only a mentor but also a friend whom I know I can always talk to about anything without worrying about judgement. I have always admired Dr. Jon Bellona's great energy in the classroom and the genuine care he shows for his teaching and his students. I hope to follow his example by bringing that same passion for teaching and music into my own classroom. To Dr. David Crumb, I am honored to have had you on my dissertation committee for both of my degrees. Thank you for being there for me and for your continued support! To Professor Colin Ives, whose insights on art and music always shine a light on new ways of thinking about my work. I am deeply grateful for his encouragement of interdisciplinary work and collaboration.

A special thanks to Carla and Kurt, the true heroes behind Kyma—a keyword that will echo throughout this document and, without doubt, throughout the rest of my life. Kyma is the sound synthesis environment that truly opened the door to electronic music for me. I cannot thank you enough for creating this universe, where infinite possibilities of sound float in every corner. I look forward to continuing to explore, to wonder, and to dream within this universe.

I want to thank my colleagues, friends, comrades, Drs. Sunhuimei Xia and Oliver Kwapis, who stood by me through both difficult and joyful moments during my time at the University of Oregon and have continued to support and help me pursuing my career goals.

Finally, I would like to express my love and gratitude to my family and friends. To my family—Yue Jin, Li Lianfu, and Chris—thank you for supporting me with unconditional patience and love, especially in my decision to pursue a second doctoral degree. To all of my friends and the many people beyond those I can name here who have inspired, supported, and guided me—thank you!

To all women in music!

TABLE OF CONTENTS

Chapter Page

I. DISSERTATION DOCUMENT OVERVIEW 13

Data-driven Instrument Overview 14

II. THE DESIGN AND IMPLEMENTATION OF THE DATA-DRIVEN INSTRUMENT USED IN THE PIECE 15

The Connection of Calligraphy and Data-driven Instruments 16

Overview of YOLO 17

The Training Processes 18

Creating the Training Data 18

Labeling the Dataset 20

Training YOLO with the Dataset 20

III. MUSICAL CHALLENGES AND OPPORTUNITIES 21

IV. DATA MAPPING STRATEGIES 26

V. PERFORMANCE TECHNIQUES 32

VI. INTENTIONALITY OF SPATIALIZATION DESIGN 34

First Panning Strategy 34

Second Panning Strategy 36

Third Panning Strategy 38

Fourth Panning Strategy 39

VII. COMPOSITIONAL STRUCTURE 41

Postlude 45

BIBLIOGRAPHY 47

LIST OF FIGURES

Figure Page

1. Basic Three-part Data-driven Instrumental Structure 14

2. Data Flow Diagram for *具足圆成 Until All Become One* 16

3. Frames From the Video Clips Taken in Studio 74 and Room 163 19

4. Data Stream Send by Python Script..... 23

5. X and Y Data Ranges and Orientation in Studio 74..... 25

6. Two of the Two-fold and One Four-fold to Divide the Space of Writing ... 27

7. The tripWires Triggering Mechanism 30

8. Bells that Hang on the Guanyin Pavilion..... 35

9. Third Bell Sound Spatial Design with Example of Left Panning Algorithm 36

10. Two Kinds of Two-fold Panning Space 38

11a. Delay Example Structure 38

11b. Delay Example Algorithm 38

12. Throat Singing Panning Algorithm 39

13. Value in Terms of the Distance to Audience 41

14. Stroke Orders of Character 成..... 45

SUPPLEMENTAL FILES

Video Performance of *具足圆成 Until All Become One*

Custom software and affiliated files required to perform *具足圆成 Until All Become One*

CHAPTER I

Dissertation Document Overview

This dissertation document consists of original software and associated files used to create the composition and the video documentation of the performance. In addition to the supporting files, this text document examines six key aspects of the work: 1) the design and implementation of the data-driven instrument used in the piece, 2) musical challenges and opportunities, 3) data mapping strategies, 4) performance techniques, 5) intentionality of spatialization design, and 6) compositional structure.

For my data-driven instrument dissertation, I composed an original electroacoustic composition titled *具足圆成 Until All Become One* for computer vision and Symbolic Sound's Kyma. *Until All Become On* is a multichannel real-time human-machine interactive work that uses a video camera and machine learning algorithms connected to the Kyma sound synthesis environment.[1] As a Chinese composer, my country and culture continue to be essential inspirations for my musical expression. The rich history of China and its resplendent art, literature, and way of life profoundly influence my musical style. For my dissertation in particular, the main inspiration is drawn from my encounter with one Chinese calligraphic masterpiece on a trip to Dule Temple that I made during the summer of 2023. Dule Temple (独乐寺) is a Buddhist temple located in a suburb of present-day Tianjin, China. The temple is of historical and architectural significance. Its oldest surviving buildings are two timber-frame structures that date back to the Liao

[1] Kyma is a language for creating, manipulating, and combining sounds. Carla Scaletti. *Kyma X Revealed!* (Symbolic Sound Corporation 2004), 19.

dynasty (916–1125) and are among the oldest surviving wooden buildings in China. Among the uniquely structured buildings, another treasure of this temple are the calligraphic masterpieces. At the center of the Guanyin Pavilion (观音阁), on the ground level, the center hanging calligraphy work shows "具足圆成" written by Xianfeng (咸丰), an Emperor of the Qing dynasty (1636–1912). The text of this work is borrowed as the title of my dissertation and roughly translates to "with enough of practice, one will achieve enlightenment."

Data-driven Instrument Overview

A data-driven instrument typically is comprised of three sections: 1) data acquisition through human performance; 2) a data mapping and transformation layer that processes the acquired data; and 3) a sound generating algorithm that renders the converted data as digital representations of sound.[2] A basic three-part data-driven instrument structure is shown below in Figure 1.

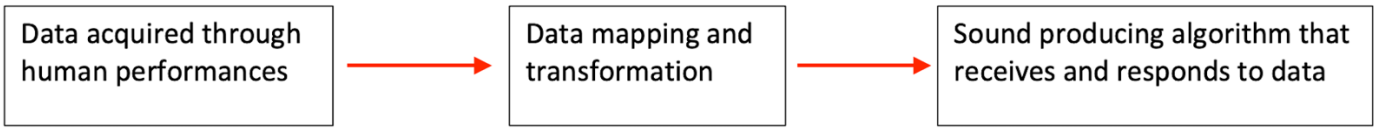


Figure 1. Basic Three-part Data-driven Instrumental Structure

[2] Jeffrey Stolet, *Do: Notes about Action in the Creation of Musical Performance with Data-driven Instruments* (Lulu, 2021), 24.

CHAPTER II

The Design and Implementation of the Data-driven Instrument Used in the Piece

Given this dissertation's focus on calligraphy as a central theme, the interface design deliberately incorporates a physical brush as an integral component of the instrument. Camera-based tracking captures the brush's spatial movement, enabling the performance to move beyond a conventional two-dimensional table surface and effectively to extend into a three-dimensional field perpendicular to the performer. This configuration offers two significant advantages: it permits audience members to experience brush strokes directly and it allows the performer to maintain a face-to-face engagement with viewers, avoiding a downward, table-focused posture or reliance on a screen projection of the performative actions.

The data-driven instrument used in *具足圓成 Until All Become One* consists of two main components: 1) a camera paired with YOLO object detection, which captures and outputs positional data of the tip of the brush from the brush prop's movement, and 2) Kyma, which receives the data, executes data mapping operations, and produces sonic responses (see Figure 2). Although, Figure 2 appears to only have two parts, the data mapping process is actually divided between YOLO and Kyma, so conceptually the standard three-part data-driven instrument model remains. The first data mapping layer occurs after YOLO detects the brush's tip's X and Y coordinates: a Python script normalizes these values to a 0-1 range and sends them to Kyma via the OSC protocol. The second data mapping layer takes place within Kyma at the parameter level, where the normalized data is further processed to generate the final sonic output.

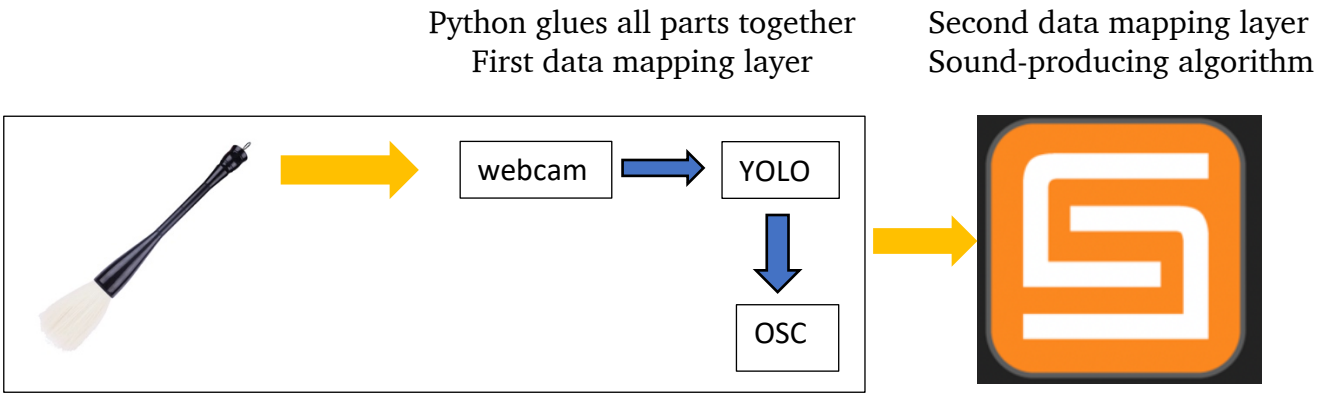


Figure 2. Data Flow Diagram for *具足圆成 Until All Become One*

The Connection of Calligraphy and Data-driven Instruments

Chinese calligraphy has been part of my artistic practice since a young age. Seeing the masterpiece in Dule Temple inspired me to compose a musical composition using this ancient art form that combines the beauty of a visual art with the practical function of a communication tool. Interestingly, the practice of calligraphy has many similarities to the practice of data-driven instruments.

Firstly, in calligraphy, one has to carefully choose the right tools for writing. The tools include brush, ink, ink stone, and paper—a series of tools that determine the end result of the work. The size, weight, and materials of the brush; the consistency of ink prepared; and the material of the paper can also each drastically change the presentation of the resulting work. These requirements and dependencies correspond directly to the design and preparation of a data-driven instrument: choosing the right interface; deciding how the data is to be mapped; and discovering how these data can be used to control sound-producing algorithms to create the desired aesthetic result.

Secondly, within the creation process, the order of strokes, when strokes start

and end, and how characters are organized in the calligraphic practice resembles the same order of organization and operation as in data-driven instruments. Finally, both calligraphy and data-driven instruments require practicing repetitive steps in order to present a polished aesthetic work. The conceptual and practical similarities between calligraphy and data-driven instruments form the core ideas of this dissertation. The meaning of the title, although originally from the Buddhist scripture quoted in the Dule Temple calligraphy, perfectly suits this endeavor in pursuing the practice and execution of music in the past, present, and future.

Overview of YOLO

YOLO—which stands for You Only Look Once—is an advanced machine learning-based real-time algorithm for object detection in image and video data. Pre-defined YOLO models can be fine-tuned to detect particular custom objects, in my case, the brush tip of the prop I use to "write" the calligraphy.[3] Images of my movements waving the brush in the air and mimicking strokes of the characters are captured by the external camera then sent to YOLO for object recognition so that I can use the movements of the brush as a control signal.

To fine-tune the original YOLO model for the brush prop and techniques employed within the composition, I used *supervised learning* procedures based on short videos of me waving the performance brush with different types of strokes in various

[3] YOLO models can be downloaded and used directly via the Ultralytics Python package (https://github.com/ultralytics/ultralytics). The models automatically download when used the first time in Python. Alternatively, download can be accomplished through Hugging Face (https://huggingface.co/models?other=yolo&p=3&sort=downloads). For this dissertation composition I used version 26 (YOLO26).

environments. Then, I manually labeled the location of the brush tip that I want to detect within selected frames of the video. Lastly, I fine-tuned the default YOLO models using these labeled video frames to create a customized object detection model that only sensed and located the brush prop's tip. After my custom model was trained to recognize the brush tip, during real-time performances it was able to accurately output data that specifies the X and Y position of the brush tip in each frame.

The customizable object detection capability from YOLO provided several advantages for use in my piece. First, ready-made commercial products often have limitations on what objects can be detected. Often these shape-recognizing models detect predetermined objects such as hands, facial expressions, road signs, etc., but not such things as a brush tip. Second, ready-made models often have messy training data and sloppy labeling that may lead to hallucination issues.[4] By using a custom model, I can "quality control" the model because I create the training data for fine-tuning. In particular, during the labeling process, I fine-tuned how detailed the labeling should be in order to reduce the possibility of technological hallucinations within my instrument.

The Training Process

Creating the Training Data

The training process included three sequential stages: dataset creation, data labeling, and model training using the customized dataset. I created a custom dataset by

[4] Hallucinations are false, nonsensical, or misleading outputs produced by a learning model that does not reflect the model's intended output. Hallucinations have many causes including bad and/or incomplete training data. See Wecheng He, et al., *Mitigating Hallucinations in YOLO-based Object Detection Models: A Revisit to Out-of-Distribution Detection* (Arxiv.org, 2025), 2.

recording short videos (each about a minute long) of myself performing brush movements that mirrored the gestural vocabulary I would employ in the final composition. To maximize dataset diversity, I intentionally recorded in three spatially and contextually distinct environments, about four or five videos in each environment. The first recording session took place in my home, where I performed in black clothing against an off-white wall with a window shade visible in the background. The second session occurred in Studio 74 (at the University of Oregon), my primary composition and rehearsal space, where I wore the same black outfit with a purple curtain as the background. The third session was conducted in SOMD Room 163, the designated performance venue, which approximates a standard performance environment. For this final session, I maintained consistent black clothing, calibrated the lighting to match intended performance conditions, and incorporated three different background colors to capture a broader range of performance scenarios (See Figure 3).



Figure 3. Frames From the Video Clips Taken in Studio 74 (left) and Room 163 (right)

Once the video clips were recorded, I employed a Max patch to strategically select and extract frames from each clip. The frame selection process prioritized capturing both conventional and unconventional brush tip configurations, since the

brush's appearance changed substantially based on the velocity and direction of the brush's movement. From each video, I extracted approximately 80 to 100 frames that represented a diverse set of brush tip morphologies.

Labeling the Dataset

The initial training dataset comprised approximately 400 frames. Prior to model training, I performed manual annotation by establishing a single object category labeled "brush" and drawing bounding boxes around the brush tip in each frame. Upon completion of the labeling process, the annotated dataset was exported and downloaded as a packaged file.

Training YOLO with the Dataset

I trained the YOLO26 model—the newest model of the YOLO family—on the custom dataset, beginning with an initial experiment using 20 epochs, where each epoch represents a single complete iteration through the entire dataset. Testing of this initial model revealed inadequate performance: the model failed to reliably locate the brush tip and proved unusable for performance purposes. Subsequently, I retrained the model for approximately 210 epochs, which produced substantially improved results. The retrained model demonstrated stable performance when I performed in black clothing; however, I identified hallucination artifacts when wearing more colored clothing. Occasionally, when the brush was absent from the frame, the model exhibited false positive detections even when I wore black clothing. This detection of false positives stemmed from a labeling artifact—during the initial annotation process, portions of my body, including my hands, face, and arms, were inevitably captured in

the dataset due to the natural grip and positioning required to manipulate the brush. To mitigate the issue, I augmented the dataset with unlabeled negative examples: frames showing me standing in the performance space without the brush and with no annotations. The inclusion of these negative training examples enabled the model to learn the distinction between frames requiring detection and those where no detection should occur.

Finally, my fine-tuned model did not suffer from overfitting as demonstrated by its observed behavior during the process of composition, practice, and performance.[5] In particular, it generalized sufficiently well and was able to reliably detect the brush tip even as the exact camera angle, lighting, and body positions changed. This is likely due to the inherent simplicity of the learning problem (only one type of object to detect in a limited type of environment), accurate labeling, insertion of negative examples, stopping training after a fixed number of epochs (before any overfitting begins to occur) rather than waiting for the algorithm to terminate, and the data augmentation techniques applied automatically by the YOLO training algorithm.

CHAPTER III

Musical Challenges and Opportunities

My fine-tuned YOLO model outputs continuous positional data for the brush tip when the object is detected in the frame, consisting of a stream of X and Y coordinates. These two data streams are produced at a 30 frames per second rate. Position estimates

[5] Overfitting occurs when a machine learning model poorly generalizes from its training data and interprets noise and other oddities of the training datasets into its understanding of the desired relationship between input and output.

are computed by averaging the two detected points along each axis. I considered adding a binary on/off trigger—activating whenever the brush appears in the frame—but opted against it because of the risk of false triggers. This decision emerged from practical observation: across different lighting environments and clothing variations, YOLO occasionally produces false detections arising from light-colored surfaces, reflective objects, or portions of my body (particularly my hand). The latter is difficult to avoid because of the physical intimacy between hand and brush means that training annotations inevitably capture portions of the hand within the brush bounding box. These false positives typically vanish once the actual brush enters the frame and begins moving, allowing YOLO to recover a correct localization. However, incorporating an on/off trigger would risk translating these momentary artifacts into sustained false triggers. Continuous positional data, by contrast, provides a more nuanced and stable foundation for mapping strategies and it allows me to create multiple trigger types without having to deal with on/off messages.

The Python script that captures video frames and runs YOLO's object detection algorithm sends the positional data stream directly to Kyma through the OSC protocol (Figure 4). In Kyma, the incoming messages may be routed in one of two ways: either through the Ethernet connection already used by the Kyma software application or through the WIFI connection established by the Kyma hardware. In either case, the OSC messages are received by the Pacamara. Each network path is identified by a unique IP address, allowing the data source to select the path to the receiver. These two pathways

each offer particular opportunities and challenges that effect performance of how the data is delivered to the sound synthesis algorithms.

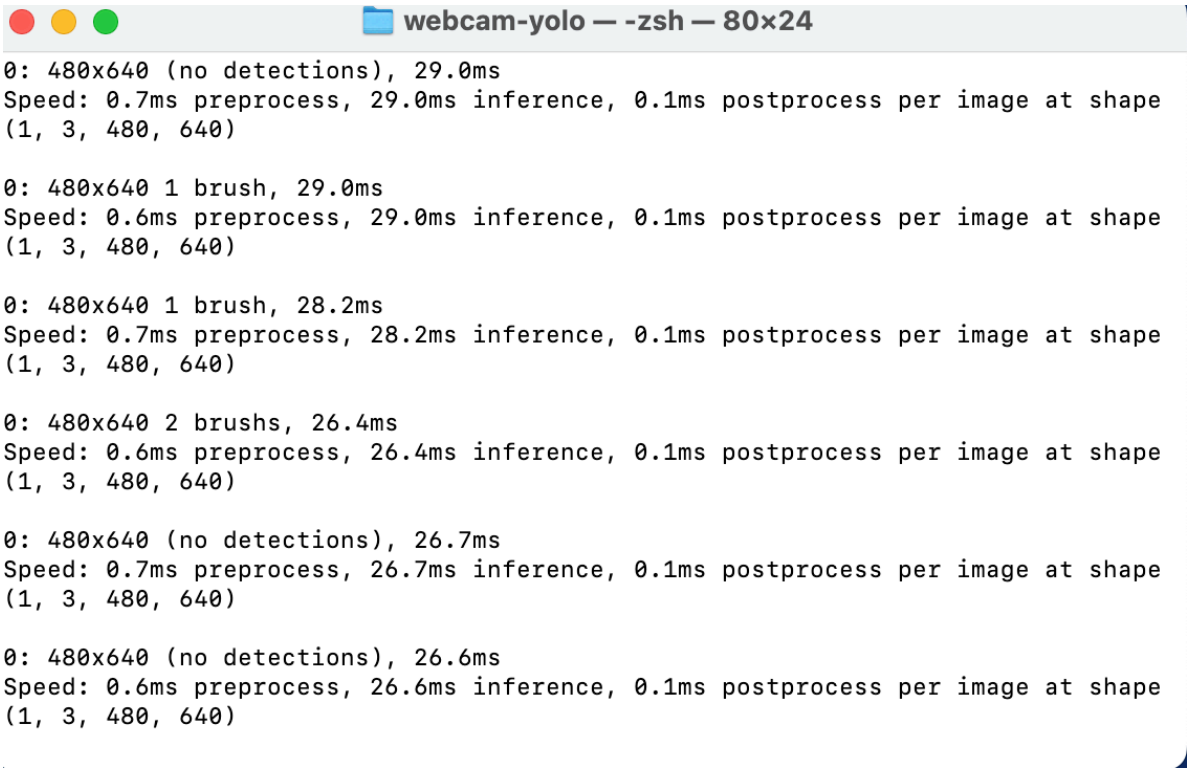


Figure 4. Data Stream Send by Python Script

Option 1, send over Ethernet. The direct hardware pathway offers a crucial performative advantage: the Ethernet connection between Pacamara and the computer provides a stable, optimal latency, low-noise transmission environment compared to WIFI radio frequencies. This choice reflects a deliberate prioritization of temporal precision in data transmission—a foundational requirement for responsive, embodied performance. A notable characteristic of this architecture is the dynamic nature of the hardware's IP address; each reboot prompts the operating system to reassign the address. While this might seem a limitation, it functions as an integral feature of the instrument design: before each performance or rehearsal session, I must verify that the Python

code's IP address aligns with the current hardware assignment, a small ritual that ensures synchronization between the digital and physical components of the instrument.

Option 2, send over WIFI. The WIFI pathway offers a contrasting set of strengths and weaknesses. The static nature of WIFI IP addressing eliminates the need for pre-session verification, but more importantly, wireless transmission enables the spatial freedom in interface placement. While this advantage is not directly applicable to the current instrument setup—YOLO's dependence on the host computer limits physical mobility—it proved valuable in my previous work, *Scattered Resonance* for iPad and Kyma, where OSC messages traveled from iPad to Kyma via WIFI. This wireless architecture permitted large freedom in terms of where the interface is placed, even when there is a door in between the iPad and the computer that is running Kyma, offering substantial flexibility across different performance contexts. However, this flexibility comes at a cost: radio signal interference introduces inherent latency and potential connectivity disruptions that do not occur with direct Ethernet transmission.

Figure 5 illustrates the two continuous data streams and their coordinate mapping: the X axis extends from 0 to 1 across the horizontal plane (left to right), while the Y axis extends from 0 to 1 along the vertical plane (top to bottom). This orientation represents a departure from the conventions established by iPad and Wacom tablet interfaces, which invert the Y-axis polarity—ascending from 0 at the bottom to 1 at the top. Each orientation carries its own internal logic and integrates naturally with its host interface. While Kyma provides the technical means to invert YOLO's Y-axis orientation through Capytalk's negate function, I chose to preserve the native directionality. This

decision revealed an unexpected resonance: the top-to-bottom directionality aligns seamlessly with the fundamental logic of Chinese calligraphy, where writing progresses vertically downward.[6]

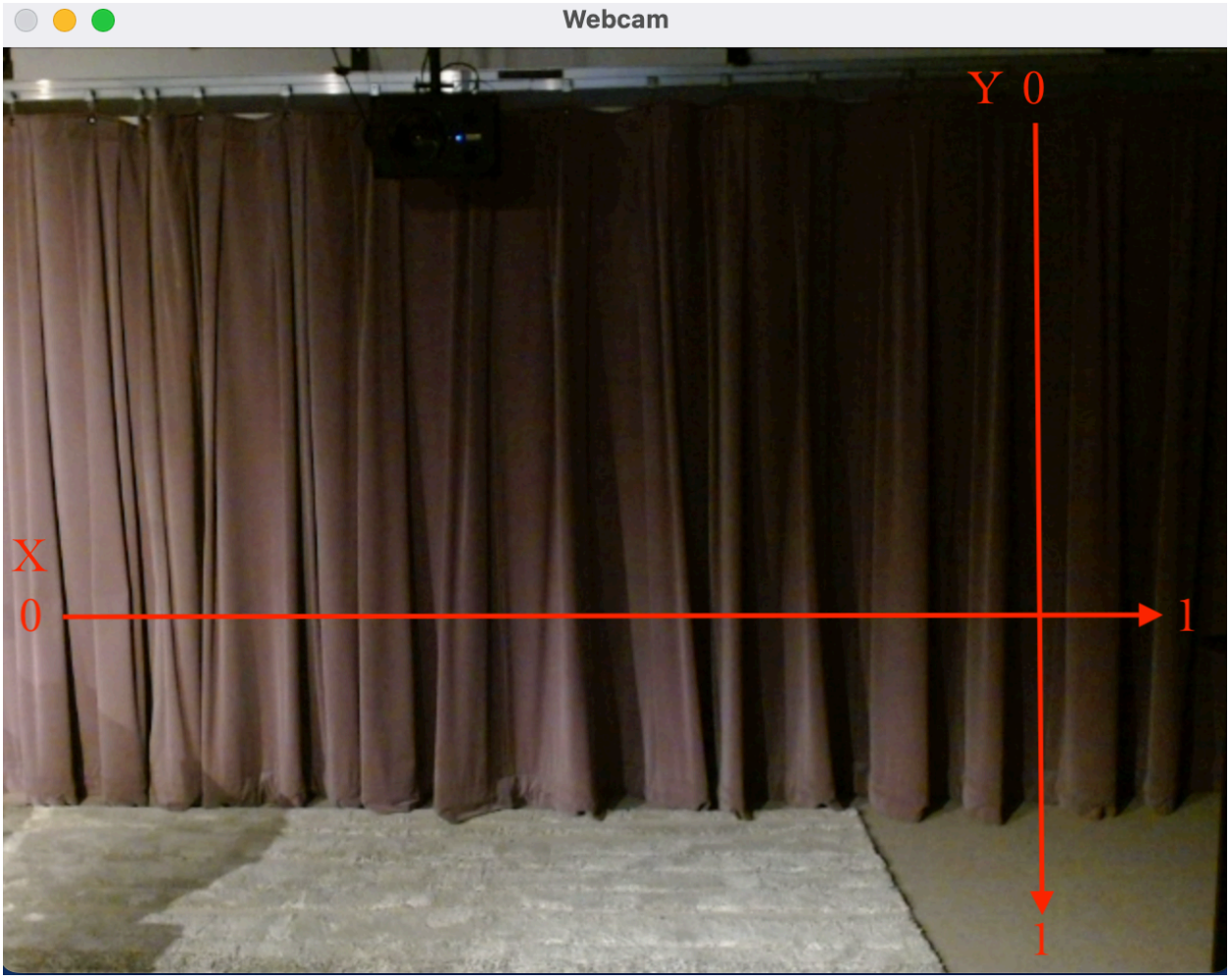


Figure 5. X and Y Data Ranges and Orientation in Studio 74

[6] Yee Chiang, *Chinese Calligraphy: An Introduction to Its Aesthetic and Technique* (Cambridge: Harvard University Press, 1974), 38.

CHAPTER IV

Data Mapping Strategies

Before turning to the data mapping strategies employed in this piece, I wish to foreground the underlying logic and spatial philosophy of Chinese calligraphy—a foundation that shapes the mapping, spatialization, and sonic architecture of the composition. As Yee Chiang observes, "The most marked difference between a Western language and the Chinese is that the former is written horizontally from left to right and the latter vertically from right to left… but to write horizontally with a Chinese brush is not convenient because every Chinese character is constructed in a square and is always written from the top downwards."[7] This distinction is not merely typographic; it reveals a fundamentally different orientation toward the page as a spatial field.

The contemporary landscape of Chinese writing reflects a productive tension between tradition and modernity. Horizontal, left-to-right text has been widely adopted as the standard convention; yet the classical vertical orientation, proceeding from top to bottom and read from right to left, endures as the governing spatial paradigm within calligraphic practice. It is not uncommon today to encounter publications that honor both conventions, the modern and the classical, coexisting on the shelf. Within the discipline of calligraphic art, however, the traditional spatial logic remains largely intact. Central to this logic is the invisible architecture of the character itself. Chiang further notes that "the Chinese character is always written in an imaginary square which, for the

[7] Ibid.

purpose of analysis, it is convenient to divide into two or four parts."[8] This imaginary square—at once a constraint and a canvas—governs proportion, balance, and rhythm, encoding a spatial intelligence that has quietly informed the mapping strategies explored throughout this work (See Figure 6).

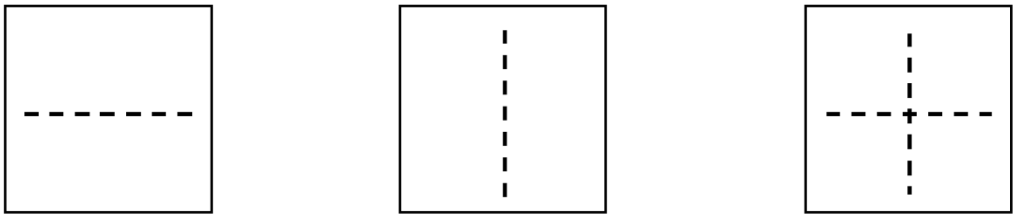


Figure 6. Two of the Two-fold and One Four-fold to Divide the Space of Writing

The treatment of space in Chinese art extends well beyond calligraphy into a broader aesthetic and philosophical principle. Known as 留白 or 布白, negative space is, in Chiang's formulation, "intimately connected with the Chinese philosophy of art."[9] This is not mere absence, but a charged and generative void. As Chiang elaborates, "in calligraphy the 'empty' portions of the imaginary character-squares are full of invisible muscles joining stroke with stroke. These spaces are not, in the truest sense, 'blank' at all, but constitute unstated expressions of sky, land or water."[10]

Silence, in this framework, is never empty—it is dense with implication; and it is precisely this quality that I seek to embody in this composition, where the spaces

[8] Ibid, 167.
[9] Ibid, 169.
[10] Ibid.

between sounds carry the unspoken emotions that resist direct musical expression. This philosophy of negative space permeates my composition at every scale: from the spatial design of the performance environment, to the structure of the data-mapping strategies, and to the compositional decisions embedded in the musical material. Negative space, in each of these domains, is treated not as absence but as a site of imaginative and emotional resonance.

To trigger sonic events and to provide ongoing control of sonic parameters I rely heavily on three mechanisms. The first of these is the breaching of thresholds that are numerically defined. Through the mechanism of threshold breaching, continuous data can be conditioned to simulate discrete button-like triggers. The piece opens with a single horizontal stroke that ascends into a vertical lift—two gestures, two thresholds, activating two distinct sonic events. This is achieved by using Capytalk's "greater than" and "less than" logical operations (gt: and lt:, respectively). During the horizontal stroke, when !YoloX is greater than 0.5, the first sound is triggered; in the subsequent lifting gesture, when !YoloY is less than 0.4, a gong strike formally opens the piece. All WaitUntils constructs throughout the work operate as triggers derived from either a single threshold condition or a compound X–Y intersection. Towards the end of the first section, a sequence of string plucking timbres—immediately followed by a drum stroke—is mapped to the calligraphic strokes 撇 (丿 pie) and 捺 (㇏na). These strokes are selected with deliberate symbolic intention: they constitute the concluding strokes of the character 具 (ju), thereby signaling the approaching completion of the current character

and the imminent transition to the next section of the composition. The threshold logic divides the coordinate space into lateral zones analogous to the two-fold square frame of traditional calligraphy practice. When conditions !YoloY is greater than 0.65 and !YoloX is less than 0.45 are simultaneously satisfied, the string plucking is triggered; when !YoloY greater than 0.65 and !YoloX is greater than 0.55 coincide, the drum is triggered—each condition anchoring the performer's gesture to a distinct spatial and sonic territory.

A second method I use to enable the triggering of events involves the placing of imaginary wires (Capytalk tripWires) across the space's dimension. Events are triggered each time the tip of the brush crosses over one of the imaginary wires. Unlike the threshold, which assesses whether the data streams eclipse (progressing above or below) a given value, the tripWires mechanism is concerned not with states but with crossings—it is the act of moving the tip of the brush beyond a designated point. This distinction grants the tripWires mechanism a particular spatial expressivity. The performance space becomes a field that the composer may partition freely—into two zones or twenty, into broad territories or narrow corridors—each invisible boundary lying in wait for the performer's movement of the brush to animate it. In the opening section of the piece, I assign a single wire (tripWires: 0.5) of the X-axis, the wire lies in the middle of the imaginary canvas. A rim shot sound is triggered every time the brush travels across this boundary in either lateral direction. The structural logic directly mirrors the two-fold calligraphic practice square: the wire reinstates its central division as a live, responsive

boundary—one where the brush crosses points in a continuum and enacts the spatial grammar of the calligraphic frame (See Figure 7).

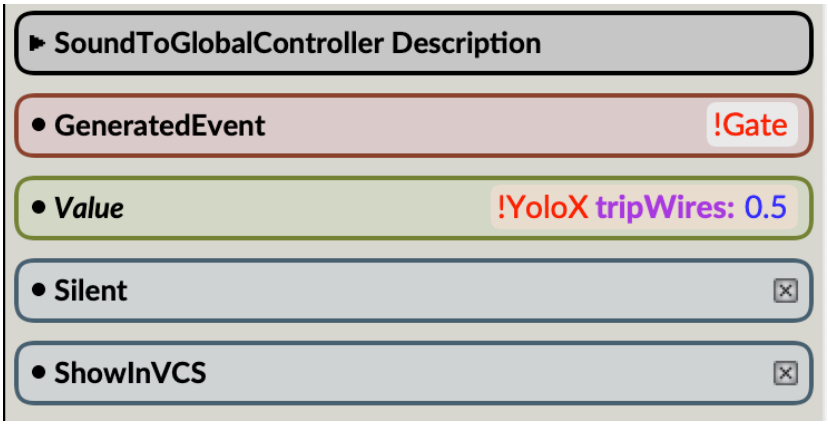


Figure 7. The tripWires Triggering Mechanism

A third method is used on a number of occasions where data arising from the brush's position in the two-dimensional space provides ongoing control of a sonic parameter. For example, near the end of the piece, the Y-axis is mapped to the frequency parameter of a particular voice in a choral texture modeled on the timbral characteristics of Tibetan throat singing. A fixed base frequency is offset by the incoming Y-axis value, whose range of values 0 to 1 is mapped to the range of -1 to 1, allowing upward and downward physical movement of the brush to produce proportional rises and falls in pitch. Thus, the movement of the brush and the movement of the sound in three dimensional space become one connected gesture. This mapping carries specific ethnomusicological resonance. In Tibetan throat singing, a practitioner sustains a foundational drone while selectively amplifying upper partials through precise nasal and pharyngeal modulation, producing the perceptual effect of simultaneous multi-voiced singing. By mapping the Y-axis data to frequency in this sonic context, the performer's

body assumes an analogous role—shaping the harmonic contour of the sound through the continuous, intentional arc of physical movement of the brush.

The three data mapping mechanisms outlined above are not mutually exclusive; their combination enables a significantly expanded range of compositional and performative possibilities. This combinatorial logic resonates with a principle deeply embedded in the Chinese written language: a character stands complete in itself, carrying its own meaning, its own history—yet when placed alongside another, something new emerges. The data mapping strategies employed in this work operate along a similar logic: each mechanism is self-sufficient in isolation, yet their combination produces compositional possibilities that no single strategy could generate alone.

The first combination allows multiple trigger types to coexist within a single, unified performance space. Returning to the concluding strokes of the character 具, the section integrates both continuous data mapping and threshold-breaching triggering within the same gestural field. As the brush traces a figure-eight trajectory from left to right, continuous Y-axis data modulates the frequency parameter of a folk flute sound, producing a wide vibrato effect. Simultaneously, when the brush reaches the lower-left or lower-right corners of the imagined canvas, threshold conditions are met, triggering two kinds of drum sounds. The sonic complexity of this passage arises not from any single mechanism but from the superimposition of threshold-breach and continuous control within one shared spatial domain.

The second combination operates through spatial separation rather than superimposition. In the fast section (time 3:34-7:00), a drum combination is triggered by

sweeping the brush across a tripWires positioned at the midpoint of the X-axis. Concurrently, a rhythmic bell sounds in counterpoint to the drum combo, its timbral character subject to continuous modulation via vertical brush movement. Crucially, as long as the brush remains on one side of the wire—within the open territory of either half of the imaginary canvas—the drum trigger is held inactive, granting the freedom to shape the bell sound through uninterrupted gestural motion of the brush. The single central wire thus performs a double function: it defines the condition for percussive triggering while simultaneously demarcating a zone of expressive freedom. This spatial logic directly recalls the two-fold practice square of calligraphy, in which the central division does not constrain but rather clarifies—organizing the space so that each half may be inhabited with intention and precision.

CHAPTER V

Performance Techniques

There is something fundamentally unusual about performing in a space that cannot be seen. Conventional interfaces offer the audience a point of orientation: interfaces such as the Wacom tablet or the iPad establish a direct, visible correspondence between the performer's hand and a physical surface. In each case, the surface mediates and makes legible the relationship between body and sound. In this work, no such surface exists. The performance space is the imaginary canvas—an intangible square suspended between the performer's body and the camera positioned at a distance. From the audience's perspective, the performer writes upon nothing, and yet sound emerges.

The absence of a visible interface does not diminish the precision of the mapping; rather, it transfers the legibility of the performance entirely onto the relationship between the physical positioning of the brush and the sonic response, asking the audience to perceive the imaginary canvas through the evidence of the sound itself. The unseen surface, too, echoes the principle of negative space—its absence does not leave the audience without orientation; it invites them inward, toward an image that each mind must complete for itself. In withholding the visible, the performance opens something richer: the imaginary canvas becomes most vivid precisely because it cannot be seen.

I developed two performance techniques in response to the specific demands of the invisible interface, each addressing the challenge of making the brush's position-sound relationship perceptible to the audience. The first concerns the mechanics of threshold-based triggering. Because the threshold value is embedded within the invisible canvas rather than marked on a physical surface, the trigger point cannot be felt, only crossed. To ensure reliable triggering, a degree of physical overcompensation is necessary. When a threshold is set at !YoloX is greater than 0.5, the motion begins at the far left and continues long past the moment the sound is heard, the body's momentum carrying the brush onward to value 0.7, 0.8, or sometimes further. This overextension is not an imprecision but a structural condition of performing within an intangible space.

The second technique is perhaps the most counterintuitive: the expression of staying still. In passages where continuous data is mapped to one or several parameters, movement of the brush transforms sound—but only an audience that first hears the

sound unchanged can fully perceive the nature of that transformation. The performer therefore pauses after such sonic introduction, holding the body (and the brush) still long enough so that the unaltered sound can be heard and comprehended. Then, with the first movement, everything shifts—and because stillness preceded it, the change arrives with absolute clarity. The body's rest is not absence, but preparation; the stillness speaks before the brush movement does.

Collectively, the two performance techniques serve a single overarching function: to make the relationship between gestural cause (brush movement) and sonic effect legible to the audience, bridging the perceptual gap introduced by the invisibility of the interface.

CHAPTER VI

Intentionality of Spatialization Design

Space, in this work, carries multiple dimensions of meaning. It is the acoustic field that surrounds the audience; it is the invisible canvas upon which the performer writes; it is the philosophical dimension that the Chinese aesthetic tradition has long understood as generative rather than empty. These three conceptions of space do not operate in isolation but fold into one another, each informing how sound is placed, directed, and withheld. From this convergence, four panning strategies emerge that together shape the spatial architecture of the composition.

First Panning Strategy

The first panning design is derived directly from the sonic and spatial characteristics of its source materials. The bell sounds employed throughout the piece

carry specific cultural and architectural significance: they represent the 铎 (duo), a type of Dharma instrument (法器) traditionally suspended at the four corners of Chinese Buddhist temple structures (See Figure 8). These bells hung there for two main reasons. Practically, the bell ringing startles birds away from the eaves, protecting the wooden structure from the slow damage of nesting. Religiously, they are there as a sonic offering—a continuous, wind-borne chant, a form of worship that requires no human hand. When the wind moves through the temple grounds, it does not ring the bells in sequence or symmetry; it rings them stochastically, and the sound arrives from wherever it arrives, scattered and unplanned across the four corners of the building. It is this quality—the spatial unpredictability of bells moved by wind rather than intention—that the spatialization of this first strategy seeks to inhabit.



Figure 8. Bells that Hang on the Guanyin Pavilion

At the opening of the piece, bell sounds are introduced that symbolize the religious reference to the temple. The first sound arrives as the brush sweeps from left to right—a gesture of wind—and the bell sounds from a single, fixed location, unpanned, as a corner bell rings from where it hangs. The second follows the reverse arc, right to left: this time, the sound begins to stir, moving halfway through a full circular sweep before settling. With the third bell, the scale of the wind changes altogether. A continuous, expansive gesture summons something larger, the panning is as though all four corner bells have been set swinging together. In this configuration, one bell sound is anchored at the center, while the left and right bell sounds circulate in a surround pattern. The movement of both the left and right bell sounds are designed to sit still in their original position for a second then start moving (See Figure 9). The result is a spatial scattering of overlapping bell voices across the listening space—an architectural echo of the temple's four corners rendered in sound.

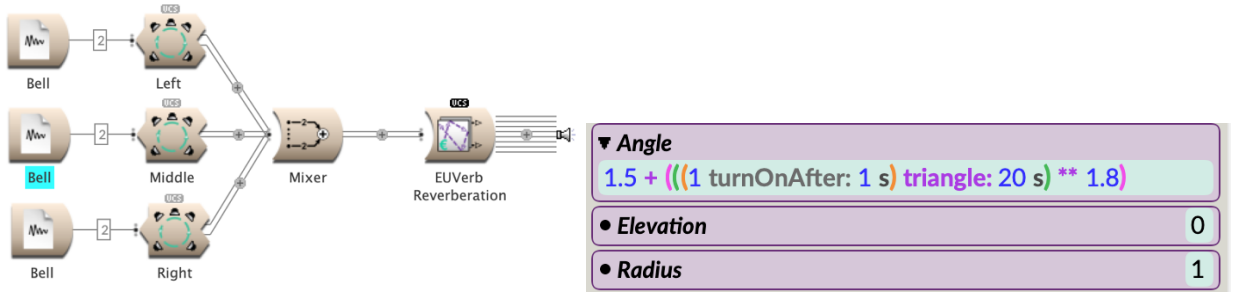


Figure 9. Third Bell Sound Spatial Design with Example of Left Panning Algorithm

Second Panning Strategy

The surround panning design of this work is grounded in the conceptual framework established by the piece's title: *one* carries within it the meaning of unity. Sonic elements are first dispersed across the listening space before gradually drawing

together into a unified whole. Following the spatial logic of the two-fold practice square, the listening space is divided into two symmetrical halves, establishing a framework from which the movement toward unity departs (See Figure. 10).

The first spatial division is horizontal two-fold: the listening space is split into front and back halves with the idea that the listener sits in the middle of the listening space. In the closing section, a sustained legato sound signals the ending of the piece, and the panning is designed to envelop the audience entirely. The sound is distributed across two panning objects moving in mirrored opposition—one sweeping right to left to right through the front half, the other moving left to right to left through the back half. The two voices travel together, yet remain within their own spatial territories, encircling the audience in a symmetrical, converging embrace.

The second division is vertical two-fold: the listening space is split into left and right. As shown in Figure 10, two panning objects depart from the same center point (value: 0.5), moving in the same direction but with a slight temporal offset between them. The sound itself is granulated—and through this combination, the audience perceives something in motion: the sound seems to shift between left and right while simultaneously drifting backward, from the front of the space toward the rear (See Figure 11a, b).

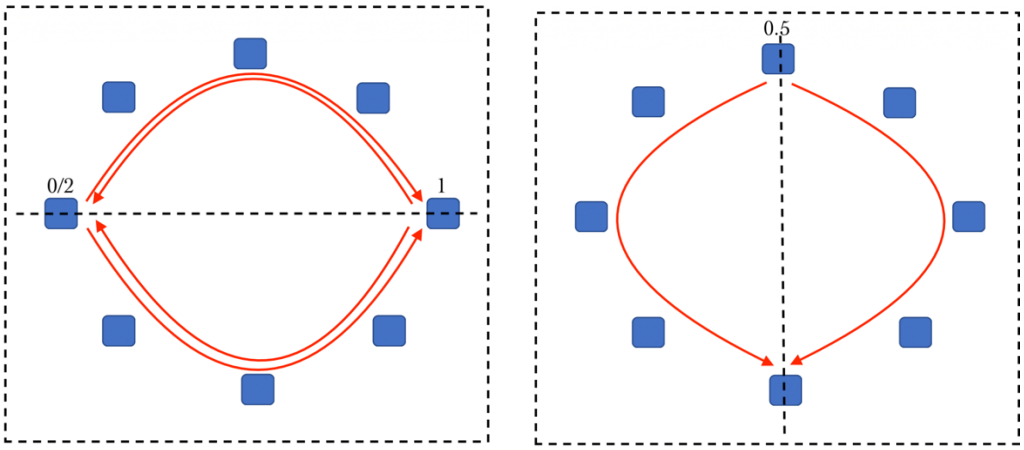


Figure 10. Two Kinds of Two-fold Panning Space

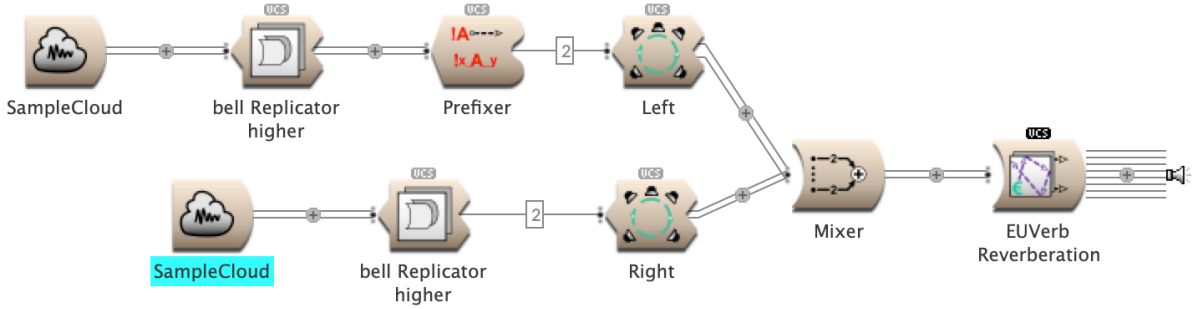


Figure 11a. Delay Example Structure

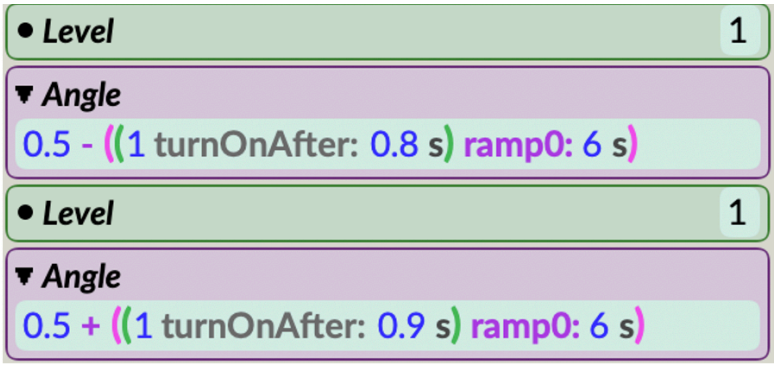


Figure 11b. Delay Example Algorithm

Third Panning Strategy

The third panning strategy is circular, drawn from the character 圆 in the title (which means circle, completeness). Circular panning may be a common technique, but

its use here is constrained in order to achieve a specific meaning: because the full circle physically embodies the form of 圆, it cannot be used indiscriminately without diminishing that correspondence. In the final section, a throat singing sound traces the full circle—but departs from convention in one respect. In the final section, the throat singing sound moves through a complete circular pan—but with a subtle inversion: rather than looping continuously in one direction, the sound completes a full circle forward, then retraces it in reverse (See Figure 12).

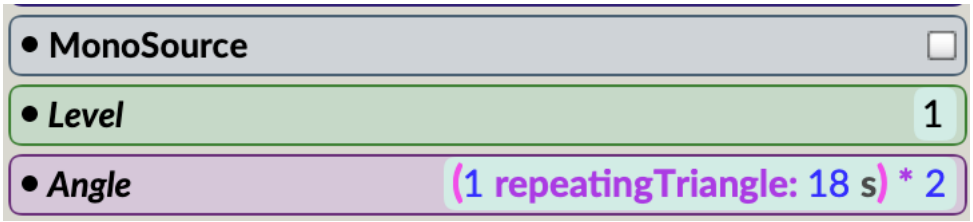


Figure 12. Throat Singing Panning Algorithm

Fourth Panning Strategy

The fourth spatial consideration is radius—not where a sound is located, but how far away it feels (See Figure 13). Spatialization in this work extends beyond the circular plane to encompass depth: the perceived distance between sound and listener becomes its own expressive dimension, shaped with the same intentionality as panning and direction.

In the second section, a choral sound undergoes a gradual accelerando, creating a mounting sense of urgency that propels the music forward. The radius parameter is synchronized to the rhythmic pulse of this choir: rather than changing linearly, the value changes by 0.1 with each beat, moving from 2 toward 0—from distance toward proximity. Every 20 beats, the radius resets to 2 and the cycle begins again. The result is

an illusion of a chanting procession advancing toward the audience in waves rather than in a single unbroken approach. This rhythmic modulation of distance also serves a perceptual function: by periodically resetting the spatial trajectory, it alleviates listener fatigue and sustains the ear's engagement across the duration of the section.

The radius modulation described above is automated through Capytalk expressions. A second approach, however, places direct control of the radius parameter in the hands of the performer via continuous data streams. Immediately following the active middle section, when everything quiet down again, a bell sound is heard again. In this sound, the X-axis is mapped to the radius parameter and multiplied by two, giving the full sweep of the brush its full range of spatial depth. Moving toward the right, I place the bell far away; drifting leftward, I draw it closer. The change in distance is not programmed but performed—each shift in the brush's position in space carrying the sound nearer or further. The spatial trajectory of the bell is thus inseparable from the physical displacement of the brush.

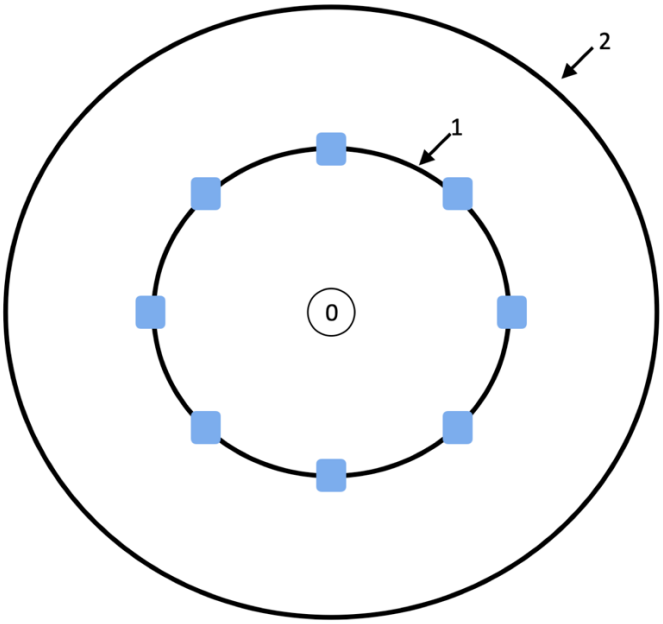


Figure 13. Value in Terms of the Distance to Audience.

CHAPTER VII

Compositional Structure

The structure of this composition reflects the architecture of the Guanyin Pavilion. In his architectural survey, Sicheng Liang described the building as follows: “it rises to three stories in height, yet its exterior presents the appearance of only two—between the upper and lower levels, a concealed intermediate floor is inserted, functioning as a mezzanine that remains invisible from the outside.”[11] The piece is structured by a similar logic of surface and depth. To the first-time listener, three sections emerge: an exposition (time 0:00-3: 34), an active middle passage (time 3:34-7:00), and a closing legato finale (time 7:00-11:18). Beneath this surface, however, the composition is organized into four

[11] The original text is: “阁高三层，而外观则似二层；其上下二主要曾之间，夹以暗层，如西式 Mezzanine 者，自外部观之不见. English text translated by Tao Li.

sections, each corresponding to one of the four characters of the Chinese title, in sequence. Each character lends its individual meaning and characteristic to its respective section; and as the four characters together articulate a progressive structural arc, so too does each section unfold in a manner that reflects the particular semantic weight of the corresponding character.

The first section is governed by the character 具 (ju). Within the four-character phrase, it carries the meaning of preparation. As a standalone character, its semantic range extends to include detail, completeness, and the notion of being equipped or furnished with the necessary means. This constellation of meanings aligns naturally with the function of an exposition: a section that introduces the materials, establishes the sonic vocabulary, and prepares the ground for what follows.

The second section is governed by the character 足 (zu). Within the four-character phrase it carries the meaning of sufficiency or fulfillment. As a standalone character, it also denotes the foot and functions as a radical in numerous verbs connoting movement and forward motion. The section it inhabits is the most physically alive in the piece, characterized by dense brush movement and a pronounced rhythmic energy that provides sharp contrast to the quiet exposition preceding it. Variations of drum combinations constitute the primary sonic material, layered above continuous rhythmic figures to produce a polyrhythmic texture. The precise timing required to trigger the drum combinations is a deliberate compositional choice—one that pays homage to my background as a percussionist.

Drumming executed with rhythmic exactitude appears as a recurring trait across my output. A comparable though more virtuosic instance can be found in the opening of my piece *纸上“弹”兵 Fantasia on Battle Tactics* for Wacom tablet and Kyma where the relationship to rhythm is somewhat different. The running figure in the background of this piece is not merely an accompaniment to be played over, but a partner to be listened to, its pulse setting the conditions within which each drum trigger must land. There is a particular pleasure in this arrangement—the experience of ensemble collaboration, even when the ensemble is entirely oneself.

The third section is governed by the character 圆 (yuan). Within the four-character phrase it carries the meaning of readiness or fulfillment. As a standalone character, it denotes the circle—a form that resonates deeply with the Buddhist concept of 轮回 (lunhui), the cycle of reincarnation. The circle, in this cosmological framework, signifies not termination but return: the end of one cycle is simultaneously the beginning of the next. This meaning is reflected both in the bodily movement of the performer, which is organized around circular gestures that enact the character's form, and in the spatialization design of the section, which shifts from the two-sided panning logic of the preceding sections toward full circular panning across the surround field (0-2), tracing the complete circumference of the listening space.

The written form of 圆 is itself structurally significant: the character encloses an interior component within a square outer frame, symbolizing a circle that contains everything within it. This morphology informs the performative actions of the section.

Before tracing the circle itself, the performer first explores the outer boundaries of the imaginary canvas—beginning at the uppermost edge, moving laterally, then descending—following both the directional principles of Chinese character writing and the act of locating the invisible square before inscribing within it. Notably, 圆 is the only character among the four that is not written out directly in the performance. It is instead rendered metaphorically through circular motion—a decision that prioritizes gestural clarity, offering the audience the most immediate and legible embodiment of the character's meaning.

The final section is governed by 成 (cheng)—and the character arrives with weight. It is the last of the four, the closing word of the phrase, and its meaning does not lighten that position: within the compound, it speaks of enlightenment earned through practice; standing alone, it carries the fullness of completion, of success, of something genuinely accomplished. The Chinese expression 成了—"it is done"—says everything in two syllables. Done is not a simple word. It holds a journey inside it, often a difficult one. The writing of the character forms a melodic line layered above a repeated folk flute melody, making this the most lyrical moment in the entire piece. The choice to close with melodic, pitched material is a conscious act of return. In the composer's recent works for data-driven instruments, the focus moved away from equal temperament and toward sound as texture and timbre-focused approach. This dissertation piece, conceived as a bridge between past and future, calls back to the composer's acoustic background—and that return is concentrated in the final musical sentence. Inspired by an ancient

architectural and artistic tradition, a looking back feels not only fitting but honest and grounded: it reflects my core belief that music, like architecture and other art forms, possesses the capacity to transcend the boundaries of time and space.

The character 成 is written out in full as the closing gesture of the piece—an act that simultaneously marks the finale and echoes the calligraphic inspiration that underlies the work as a whole. Notably, although Chinese characters are conventionally written from top to bottom, 成 concludes with a stroke that returns to the top (See Figure 14). In this single detail, the accumulated metaphors of the piece converge: the full circle, the becoming of one, the end that is also a beginning.



Figure 14. Stroke Orders of Character 成

Postlude

Writing the character 成 is not without hesitation. It marks the ending of this piece, but it also marks the close of an eleven-year chapter of study and pursuit in electronic music—and I was not sure that a single character, however weighted with

meaning, could hold all of that. Perhaps it cannot. Perhaps what this work represents is not a conclusion but an intermediate node: a place where one thread of the journey becomes visible before continuing forward. 成了—the piece is done. The brush lifts from the canvas, the sound fades, and yet the character ends on a stroke that rises. The brush returns to the top and the next line, whenever it comes, will begin from there.

BIBLIOGRAPHY

Chiang, Yee. *Chinese Calligraphy: An Introduction to Its Aesthetic and Technique*. Harvard University Press, 1974.

Google for Developers. "Overfitting," n.d. https://developers.google.com/machine-learning/crash-course/overfitting/overfitting

He, Wu, et al. *Mitigating Hallucinations in YOLO-based Object Detection Models: A Revisit to Out-of-Distribution Detection*. Arxiv.org, 2025, https://arxiv.org/html/2503.07330v2. Accessed 6 Oct. 2025.

Jocher, Glenn., Qiu, Jing. *Ultralytics YOLO*, Version 26, Ultralytics, 2026.

Liang, Sicheng. *Liang Sicheng Quan Ji*. Vol. 1. Zhongguo Jianzhu Gongye Chubanshe, 2001.

Scaletti, Carla. *Kyma X Revealed*. Champaign, Symbolic Sound Corporation, 2004.

_____. *The Kyma Language for Sound Design*, Version 7.45f0, Symbolic Sound Corporation, 2025.

Stolet, Jeffrey. *Do: Notes about Action in the Creation of Musical Performance with Data-driven Instruments*. Lulu. com, 2021.

Wikipedia Contributors. "*Dule Temple.*" Wikipedia, Wikimedia Foundation, 24 May 2025. https://en.wikipedia.org/wiki/Dule_Temple. Accessed Oct. 2, 2025.

Zhang, Aston, et al. *Dive into Deep Learning*. Cambridge University Press, 2023.