

Cultural Timbre as Unvoiced Knowing: An Audiovisual Spectral Synthesizer That Transforms Chinese Musical Instruments

Yong Zhao*

12533227@mail.sustech.edu.cn

Southern University of Science and Technology
Shenzhen, China

Marcel Zaes Sagesser*†

marcel.sagesser@ustp.at

University of Applied Sciences St. Pölten
Sankt Pölten, Austria

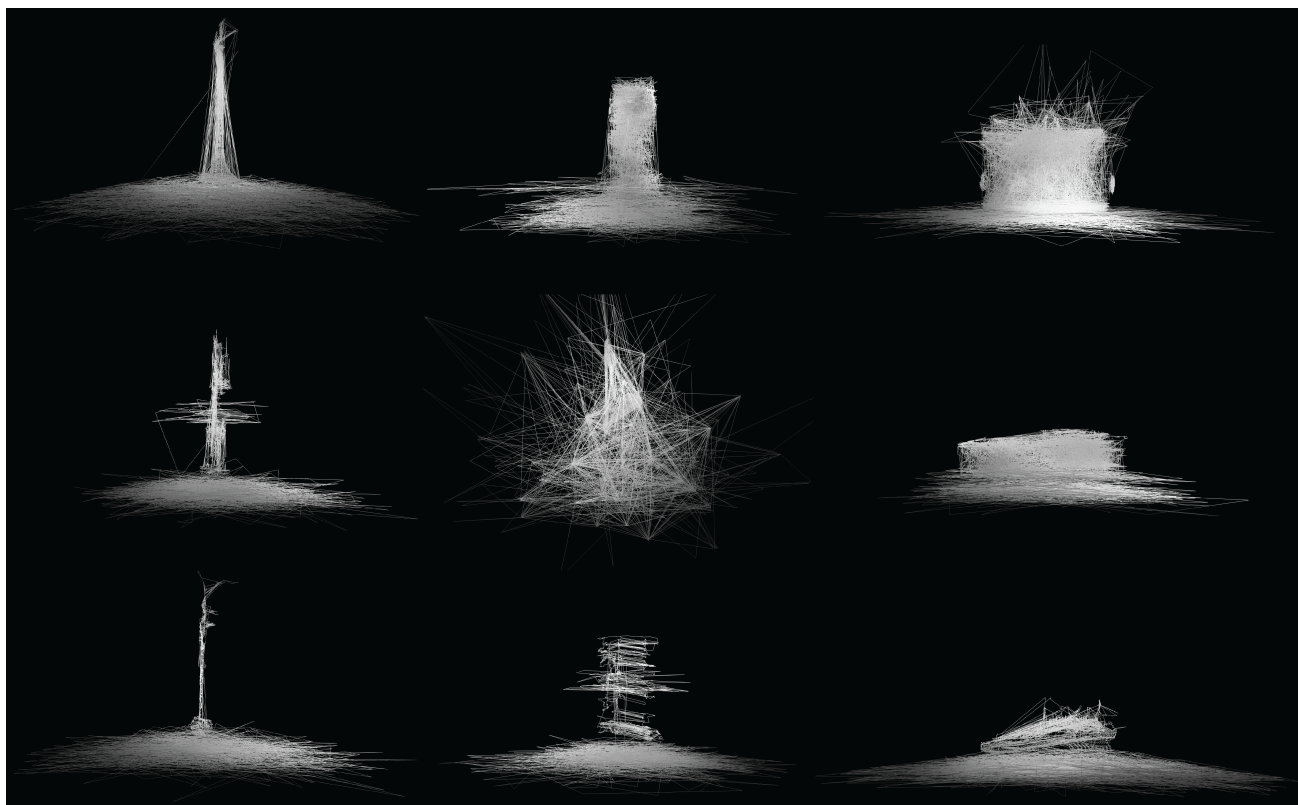


Figure 1: Nine screenshots from the installation, *SonoChrom*, showing interactive point cloud animations that visualize transitions from one Chinese instrument to another with spectral sound synthesis.

Abstract

SonoChrom is an installation based on an audiovisual spectral timbre synthesizer made out of four traditional Chinese music instruments. The installation demonstrates how sonic timbre is connected to a specific culture, activates unvoiced knowing, and bears potential for expressivity similar to that of a human voice. Timbre – sometimes under-used given a widespread dominance of pitch, rhythm, or musical tone – might matter in everyday (technological) culture as a signifier of, and activator for, unvoiced cultural knowledge. Encounters with particular timbres can activate embodied memories and personal associations.

*Both authors contributed equally to this research.

†corresponding author



This work is licensed under a Creative Commons Attribution 4.0 International License.

NIME '26, June 23–26, 2026, London, UK

© 2026 Copyright held by the owner/author(s).

Inspiring a shift in thinking towards a timbre-first, culturally situated, and multi-sensory interaction model, the authors present a digital installation, in which audiences play with remodeled sounds of Chinese instruments. In ancient Chinese aesthetic discourses, timbre is framed as an expressive device similar to the human voice, as a parameter through which thought and emotion can be conveyed with immediacy and depth. Through recording sessions, a corpus of samples was created and spectrally decomposed into partials – which are then recombined in hyper-realistic ways in the installation. A screen-based, aestheticized visualization based on the shapes of the four instruments renders timbral changes as evolving motion textures.

This paper documents the installation and advances a timbre-first approach, arguing that technologically shaped timbre, when deployed as an interactional device, offers significant potential for contemporary digital applications. As a multi-sensory system, *SonoChrom* offers experiential engagement with Chinese sound culture through timbre and visualizations.

Keywords

Timbre, Chinese musical culture, Spectral synthesis, Audiovisual installation, Sound visualization, Timbre-first thinking

1 Introduction

The study of sound has long been of interest across disciplines. One of its core characteristics, *timbre*—often described as the “color of sound” [17]—fundamentally shapes auditory experience and the perceived identity of instruments, human voices, and sonic environments. Timbre also functions as a cultural signifier [1, 20]. As such, it can activate culturally learned listening, long-term memory, and personal associations—even in infants [18]. It arouses emotional resonance [9] and unvoiced knowing [15].

Despite its salience and omnipresence, timbre remains difficult to define [2], and therefore, to access for untrained listeners, and to communicate outside of specialized fields. Compared to the relatively standardized treatment of pitch and rhythm, timbre is often underutilized [13] because of its perceptual and technical complexity. Due to the relatively difficult access to timbre as a perceptual-technical category, in the realm of cultural analysis, whenever sound is at the fore, timbre is more often than not neglected and remains unaddressed. McDermott and Oxenham, for example, regard music perception as culturally specific and auditory system-related, while arguing that timbre is “less well linked to basic perceptual mechanisms” [14].

But how do culture and timbre actually relate? In this paper, we explore this rather complex connection through a sound-installation-driven case study that foregrounds the timbral qualities of four traditional Chinese musical instruments and connects these with contemporary technology-driven sonic culture. We present *SonoChrom*, an audiovisual spectral timbre synthesizer grounded in the sonic worlds of the Chinese instruments *guzheng*, *suona*, *erhu*, and *tanggu*. The system is built from a curated, self-recorded corpus of sound samples that includes a number of performance techniques. This sample library is then decomposed into partials and recombined within a manipulable synthesis space that forms the installation. The audience explores overtone structures and hybrid crossovers which produce timbral “in-between” identities that wouldn’t be possible on actual instruments. The audience’s ECG is captured as a simple bio-signal input and mapped to modulate timbral transitions in real time, while a screen-based visual layer—derived from instrument-shaped visual motifs—renders timbral changes as evolving motion textures, supporting multi-sensory attention to timbre, with the intention that a playful engagement with timbre as a form of cultural knowledge can activate unvoiced knowing for the audience.

SonoChrom offers a structured yet culturally sensitive approach to timbre exploration with further potential for a timbre-first thinking in contemporary digital sound applications. It operates as a device for constructing culturally situated, hyperrealistic timbral identities, enabling affective audience engagement through tacit cultural recognition. It uses existing and (for many in China) well known sounds, and presents those in surprising ways. In so doing, this paper contributes, through documenting the installation, a timbre-centered workflow, which probes how an audience responds to the hyper-realistic in-between timbres that this technological system outputs. At this early stage of the research, informal, non-generalizable exhibition observations suggest strong audience engagement and a clear interest in sharing anecdotal reflections on Chinese musical culture.

2 Theoretical Background

2.1 Timbre as a Technical Parameter

Technically, any sound’s timbre is constituted of the unique configuration of its overtone series—its spectrum. How the amplitude of each partial in the overtone series is distributed over the sound’s spectrum decides on its timbre. A pitched sound can be described as a superposition of an integer number of sinusoidal components, constituted of a fundamental frequency and additional frequencies occurring at approximately integer multiples of the fundamental (e.g., f , $2f$, $3f$, ...), forming a harmonic series [3]. Instrumental sounds are recognizable not only because of the way their partials are composed, but also because of how each partial’s amplitude fades or rises over time. Timbre thus helps humans to distinguish, recognize, and memorize sounds.

Timbre has been described as one of four fundamental sonic attributes—alongside pitch, intensity, and duration—constituting musical performance and aesthetics [12]. In practical terms, timbre enables listeners to distinguish the sounds of a flute from an oboe. Even when a clear pitch is absent, for example in the sound of a gong, timbre still allows us to recognize the unpitched sound. However, due to its elusive and ambiguous essence [19], others describe it as a “wastebasket” parameter, filled with everything beyond pitch and dynamic level [4].

2.2 Timbre as Culture

In European Music, historically, timbre was long treated as subordinate to harmonic and melodic developments [4]. The Romantic period began to explore timbre as an essential element of musical storytelling. Notably, within this increasing attention to timbre, instruments capable of imitating the voice were esteemed.

This tradition partly resonates with the long-term musical philosophy in ancient Chinese instrumental discourses, where timbre is frequently framed through the metaphor of the human voice [12]. Within four traditional classifications—percussion, plucked string, bowed string, and wind—many instruments are appreciated for their ability to vividly mimic the human voice through their multifaceted performing techniques. The pursuit of the human voice, in this context, claims that it is the most expressive form, aligning with the idea that the content to be expressed is the “human” in itself. Timbre is closely tied to performative techniques that extend to the imitation of human and non-human vocalities both in the past and in modern (Chinese) electronic music, emphasizing the culturally situated approach to expression [11, 12]. Across both traditions, timbre appears not merely as a technical parameter, but as a culturally situated device that reflects deeper values.

In this context, timbre contributes to sound as a cultural signifier, to its expressive quality. As such, it contributes to meaning-making, depending not only on its production, but also on who listens: Feld argues for sound as a social dimension (because of its dynamism and its pure physical energy) [7]. It is clear that timbral parameters contribute to this social and also bodily dimension of sound. We believe that the meaning partially carried by timbre links material sound making to culturally legible (and situated) listening, exemplified in *SonoChrom* in the sense of surprising renderings of well-known sounds that let audiences draw connections with their respective memories of those instruments—thereby culturally situating the experience of timbre.

3 Related Works

In the scope of this paper, we focus on art installations that playfully render us aware of timbre, thereby showing where our work *SonoChrom* draws inspiration and thematic connections.

Fritz Winckel's *3D spectrogram* is often cited as an early "sound sculpture" that materializes spectral analysis in a physical form [5]. Published as early as 1960, it renders an eight-second excerpt of Beethoven's Eighth Symphony as a 3D spectrogram. By spatializing the sample's spectral structure, the artwork invites audiences to perceive sonic attributes through another sense, inspiring *SonoChrom* to extend the engagement with timbre by enabling multi-sensory engagement.

Bryan Yueshen Wu and Ke Peng's *Evolution* integrates luminous powders with sound in a dynamic audiovisual system [21]. The work foregrounds a material-sonic coupling in which vibrations and frequencies function as active agents that trigger transformations in inorganic materials. It offers a strong precedent for treating sonic attributes as drivers of perceptual and material change, supporting *SonoChrom*'s emphasis on timbre as an active design parameter.

Fei Jun's *Theatre of Emotions* employs affective computing to deliver a personalized audiovisual journey structured around emotional recognition [6]. The system captures ECG-derived measures to infer emotional states across categorical labels, which then generates music and visuals aimed at a therapeutic, immersive experience. This work is directly informative for *SonoChrom*'s use of ECG as an affect-related input in multi-sensory experience design.

All works discussed treat sound (or timbre) as a multi-modal entity that is translated across media and speaks to several senses. All of them present timbre of partially known sounds in a form that is unusual and surprising. This is where *SonoChrom* draws inspiration from.

4 Methodology

This research explores how timbre can activate culturally situated memories ("unvoiced knowing") and how it holds potential for expression akin to those of the human voice, and how, by recombining existing timbres into new sounds, a timbre-first thinking can be applied in technology-driven sound design scenarios.

Methodologically, this goal is met by building a theoretical framework through background research, by prototyping an installation with a spectral timbre synthesizer (based on a self-made sample library of four traditional instruments recorded by ourselves with volunteering amateur musicians) and a visualization interface, by exhibiting this installation in the End of Year Show 2025 at the School of Design, Southern University of Science and Technology, Shenzhen, and by discussing the informal observational results gained in the exhibition.

The technical pipeline entails that we dissected the recorded samples into their individual partials, and programmed a real-time system in Max/MSP, where, triggered by user input from real-time ECG data, existing instrumental samples are slowly interpolated into otherworldly hyper instruments made of recombining partials from several unrelated sound samples in our library. In TouchDesigner, we created a visualization of timbre that engages the audience in real time.

Additional outputs of this research include a poster presentation about the initial prototype and preliminary research results at the *Timbre and Orchestration in Popular Songs (TOPS)* conference 2025, McGill University, Canada.

5 System Design

5.1 Building a Sample Library



Figure 2: Recording session with amateur musicians playing *guzheng*, *erhu*, *suona*, and *tanggu*.

Traditional Chinese instruments are commonly grouped into four families: plucked strings, bowed strings, winds, and percussion. To establish a focused yet representative timbral corpus, we selected one widely recognized instrument from each family: *guzheng*, *erhu*, *suona*, and *tanggu*. We conducted a recording session with four amateur musicians (students from Southern University of Science and Technology who volunteered and consented to their recordings being used in this research), capturing a range of performance techniques for each instrument (e.g., single tones, vibrato, tremolo, etc.).

The selected playing techniques (excluding *tanggu*, which is primarily regarded as a rhythmic instrument) reflect the core of Chinese instrumental practice, aiming to mimic the human voice [12]. *Erhu* closely resembles the human voice due to its continuous sound and dynamic control. *Suona* shares vocal qualities due to its air column resonance. *Guzheng*, although not directly imitating the human voice, uses techniques like "press-and-release" and "rolling finger" to transform discrete notes into more continuous expressions, approximating speech intonation.

The resulting corpus comprises 113 audio files of 8 seconds duration each: 24 files from *guzheng*, 31 files from *erhu*, 30 files from *suona*, and 28 files from *tanggu*. All recordings were subsequently denoised and brought to the same level to produce clean source materials suitable for spectral recomposing.

5.2 Spectral Re-synthesis

The recorded samples have different timbres, as visualized here with spectrograms in Max/MSP shown in figure 3. This figure illustrates one recorded single long note played by four different instruments. To render our sound library interactable, we manually dissected each recording sample into its first eight partials and stored these as separate audio files. This step was done manually in Adobe Audition, by ensuring that each stem contains perceptually meaningful (yet unusual sounding) components of the original timbre. All versions were exported time-aligned and equal in length, to ensure further looping and blending among instruments. The process yielded in 904 audio files.

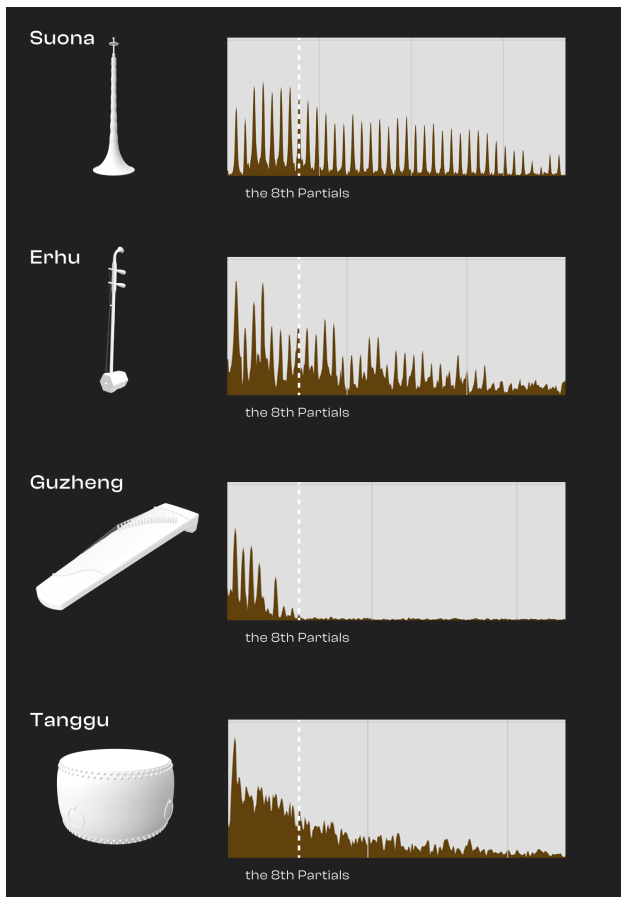


Figure 3: Spectral composition of different recorded instrument samples visualized in Max/MSP using spectrograms. In *SonoChrom*, we used only the first 8 partials for spectral re-synthesis. The white dotted lines indicate where the 8th partial of each instrument is located in their spectrogram.

As shown in figure 4, the first eight partials of each instrument and playing technique were then integrated into a Max/MSP patch for spectral re-synthesis. The system will then loop among pre-set combinations of partials of a selected given instrument, so that the audience "dives" into the inner life of a sample via different partial combinations. The trend in adding/removing partials is alternatively upward and downward among instruments, ensuring that the sum of partials played is always eight when mixing instruments. Once the transition is complete (at step 9), more intricate combinations of even/odd partials follow (see figure 5).

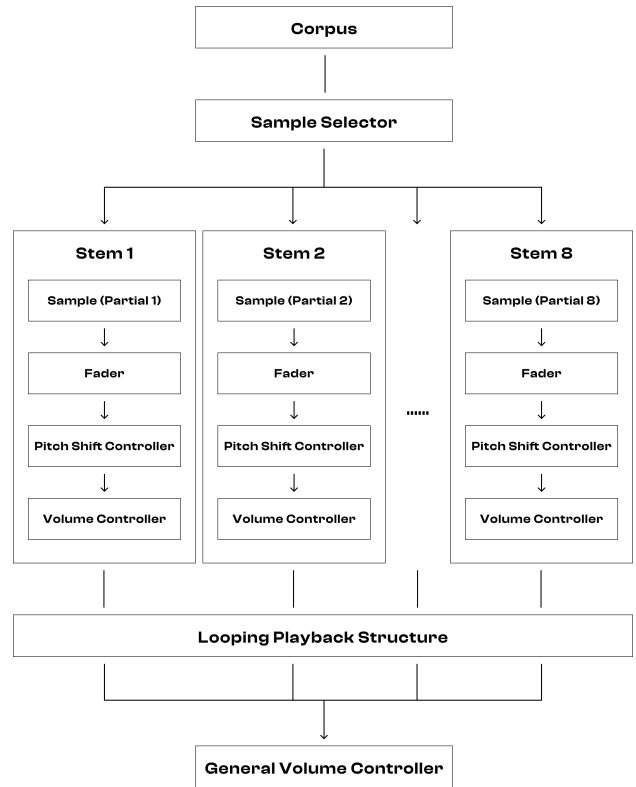


Figure 4: Diagram of the audio processing system for one instrument.

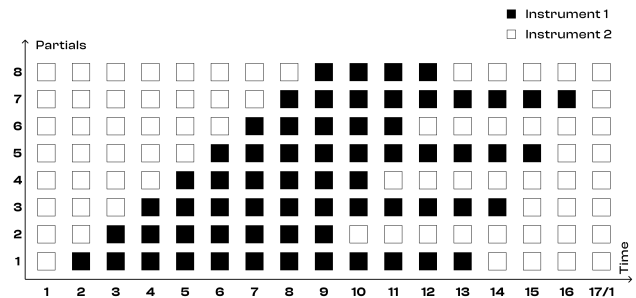


Figure 5: Schematic of the playback structure of the installation, which cycles through predefined states. It illustrates how the playback system is designed to play partials over time. There are 16 steps in total. Each step will be played twice, before the player steps to the next sound. In the idle mode, if no "In-between" timbres are played, the system will leave one instrument muted while the other playing.

5.3 Parameter Mapping

As to render the instrumental sound corpus interactable, all samples/stems were then positioned on a valence/arousal map, positioning the sounds of the four instruments at the four corners of the map, dividing up playing techniques, faster/slower looping cycles, and purposely introduced artificial detuning factors on the different areas of this map, which later shall be controlled through simple bio-data by the user (see next section). The map is shown in figure 8. Translating the produced sound corpus into a fluid sonic space not only makes it ultimately interactable in the installation, but it also allows for surprising

hyper sounds—mixtures of instruments, of playing techniques, and of partial combinations—which would not be playable on actual instruments. "In-between" timbres emerge through cross-instrument blending, where selected stems from one instrument are combined with those of another, as described afore. The resulting sounds from this fluid mapping, while faintly reminiscent of actual instruments, have a distinct characteristics: they have a certain "artificial-ness" to them. The idea was to surprise listeners with novel versions of culturally familiar sounds.

5.4 Timbre Visualizations

The visual layer in TouchDesigner (TD) externalizes timbral changes as a texture that is continuously in motion. Four 3D instrument models were imported into TD and converted into point-coordinate datasets to enable intricate parametric control; then trajectory-tracing and motion-accumulation effects over these point sets were generated; additional texture operator nodes (e.g., Noise, Threshold) were added to provide modulation in flow, density, and persistence; a linear material was applied across all models to maintain stylistic visual coherence and to emphasize line-based forms as the primary visual language.

The choice to build the visualization from instrument-shaped models was made specifically for timbre to be rendered perceptible across instrumental blends. By grounding the visual layers in a recognizable form of a source instruments, the system provides audiences an immediate visual anchor for understanding which instrumental timbre is currently dominant in the sound mixture. As sound mixtures are dynamically blending, the visual layer progressively dissolves/reassembles the initial/target instrumental shape, ultimately supporting the audience in understanding the perceptual sonic transitions.

The visuals correspond to the sonic output of the installation and produce dynamic variations—for instance, more tremulous and unstable visual motion when the sound shifts from a single sustained note toward a tremolo. Thereby, they reinforce perceptual coupling between timbral and visual motion. When hybrid "in-between" timbres emerge through cross-instrument blending, the visual layer likewise fuses trajectories from multiple models, producing a transient "in-between" point cloud shape. In this sense, the visualization makes timbral states legible, supports the audiences' orientation within the system, and renders the emergence of a hyper-instrument salient.

The installation has two modes, the idle mode (no audience inputting bio-signals), in which case both audio and visuals cycle through partial variations of one single instrument (the one selected last by the last user), ensuring that the installation never remains silent. The other mode is the bio-signal input mode (explained in the next section), where the user inputs simple bio-data to both trigger and control the next target position in the map (a specific instrument, playing technique, looping tempo, detuning factor), in which case the visualization also switches to showing the instrumental transition, as indicated in figure 7.

5.5 Audience Engagement: Bio-signal Input

As a way of producing variation in the installation with simple audience input, we decided to use audiences' physiological bio-signals, with which they might unconsciously generate connections with the installation rather than following the idea of conscious interaction as would be the case with playing a musical instrument. We chose ECG as bio-signal input and applied an ECG capturing sensor, since it is simple, reliable, fast, and

exhibition-proof. The key point here is that the ECG provides a neutral, universally accessible physiological signal. We acknowledge how limited ECG bio-signals are, and thus use them as simple interaction parameter rather than as a complex way of understanding the audience's emotional state. When different audiences engage with *SonoChrom*, their ECG is measured, which may or may not be directly influenced by what they hear from the installation; yet in any case it will change the status of the installation, triggering a next target sound and visual specifically for them.

The ECG sensor has been integrated into the installation setup, as shown in figure 6. *SonoChrom* uses ECG as a physiological input that triggers timbral motion. We treated ECG-derived features as continuous indicators of physiological arousal and valence [16], and reflections of autonomic variation, acknowledging the limited scope and interpretability of ECG data, and used them to modulate how timbral states evolve over time. Figure 6 shows an audience member interacting with *SonoChrom*.

From the ECG stream, the system derives two-dimensional control signals: heart rate (HR) and heart rate variability (HRV). These features exemplify moment-to-moment variations in cardiovascular activity, which are widely utilized as references of physiological activation and regulatory dynamics [10, 16]. In *SonoChrom*, HR primarily reflects the fluctuation in arousal, controlling parameters such as tempo, volume, playing techniques, and articulation of looped samples, while HRV serves as an indicator of valence, influencing the pitch variation range of each stem [8], as shown in figure 8. HR and HRV together define the next target state of the installation. For instance, when an audience member's arousal increases while their valence remains low, *SonoChrom* assigns *suona* as the playing instrument, and loops tremolo samples with relatively fast speed at a high volume. The duration and cycling speed through the 16 steps (see figure 5) are both shortened, rendering the output faster and more hectic. The detuning value, which randomly detunes the partials within a stem slightly to produce auditory instability, increases, resulting in a slightly more dissonant spectral mixture. In this way, all interactable parameters are accessible and controllable by the audience through their bio-signal input, always generating fluid new audiovisual states for each audience member.



Figure 6: An audience member is interacting with *SonoChrom* by touching the ECG sensor, watching visualizations on a projected screen, and listening to changing timbres on loudspeakers.

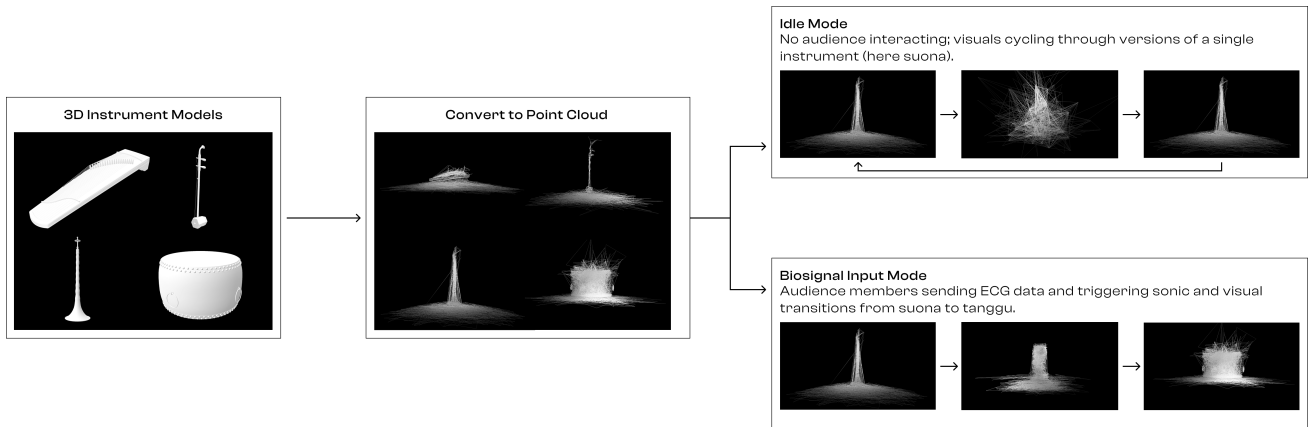


Figure 7: Conceptual framework of the visualization system, realized in TouchDesigner.

6 Exhibition Observation Data

The installation *SonoChrom* was exhibited in the "End of Year Show" at the School of Design, Southern University of Science and Technology, Shenzhen.

Estimatingly, far over 100 people have experienced the installation in the exhibition. We observed that many followed the invitation—they put on headphones, touched the ECG sensor, and engaged with *SonoChrom*. The visual layer appeared to make it easier for many to play with the installation. Several audience members also stepped up to us and commented, or started a conversation with us, on what they experienced. One audience member expressed a particular curiosity regarding the meaning behind the visual representations, especially when the visuals transitioned between different instruments and showed abstract shapes. They said they were unable to interpret the white lines, which prompted them to focus more intently on deciphering their significance. Others recalled their memories of childhood experiences during the Spring Festival. One specific audience member found the auditory experience differing from what they had expected after having read about the installation. They commented on a perceived lack of access points to these Chinese instruments, and hence found it difficult to engage with *SonoChrom* (they did not know the instruments before). One audience member stated that they were picturing a vague impression of an ancient performer, dressed in traditional Chinese attire, floating and playing suona. Other audience members perceived the sound of mixed guzheng and tanggu as resembling Western electronic music, associating it with a musical festival atmosphere. Some felt that the project conveyed a strong sense of Chinese culture. Many audience members have talked to us during the exhibition and asked about the specific backgrounds, histories, and motivation of the installation and about the choice of these specific four Chinese instruments.

7 Discussion

The discussion is based on informal observations in the exhibitions of the pilot installation *SonoChrom*. The goal of the installation was the rendering experiential of the connection between culture and timbre, and particularly, how a timbre-first thinking can raise awareness around, and activate, unvoiced knowing of culturally situated sounds, such as those of traditional Chinese instruments. Through building a bio-signal-driven audiovisual

spectral synthesizer, novel sounds were produced and made accessible for a broad, untrained audience with (but not exclusively) a Chinese cultural background.

In *SonoChrom*, timbre becomes an object for play. Some audience members associated specific *suona* sounds with grand events in their own lives, such as funerals and weddings, where *suona* is traditionally often used in the Chinese context. This observation demonstrates that some drew connections between technologically produced timbre and associations they brought into the exhibition—in this sense, unvoiced knowing might have become articulated through their play with the installation. *SonoChrom* enables hybrid "in-between" timbres across the four instrumental identities, realizing sounds grounded in culturally recognizable sources yet extending beyond what a single actual-world instrument could physically produce. Several visitors have mentioned that they struggled to identify the instruments but still found some familiarity in the unrecognizable mixtures, saying that it reminded them of some instruments. Despite the small number and the informal character of the observations, it becomes evident that different visitors produced different interpretations, and many shared a sense of something unknown yet familiar in the timbre. The variations in the audience's cultural backgrounds might have amplified the diverse interpretations of the audiovisual timbral states presented in *SonoChrom*. Above all, many pointed out that the installation conveyed a sense of unknown yet familiar. In line with the intention of defamiliarizing existing instruments and producing culturally situated technological timbral variations, the results show that several were surprised by what they heard. One non-Chinese audience member noted that the perceived timbral qualities differed from their expectations based on the research description, and others were reminded of (Western) electronic music.

We believe that the results show a certain expressive potential—not unlike that of the human voice in Chinese musical traditions—which could be applied in contemporary sound products. *SonoChrom* shows sound as a flexible material rather than a fixed state, thereby fostering a shift in thinking towards timbre-first, culturally situated, and multi-sensory sonic interaction, with potential for application in contemporary technological products where we believe a timbre-first thinking could add additional expressivity. Bio-signals, while limited, helped the users to individually engage with the installation and connect the sounds heard with their own body and their unvoiced knowing.

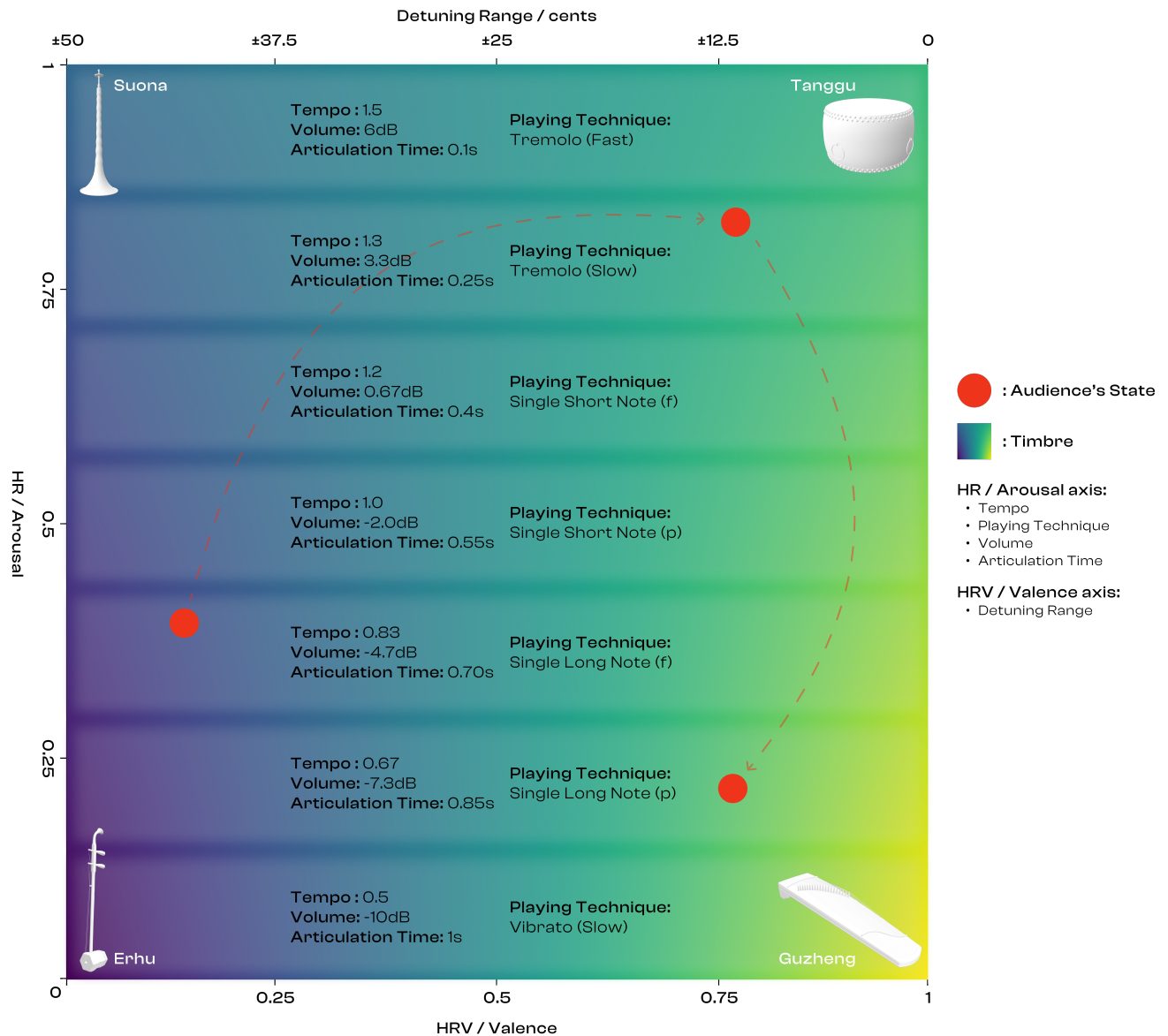


Figure 8: Mapping space of ECG signal to sounds. Except for playing techniques, all parameters are changed linearly based on real time ECG bio-signal input from audience members, ultimately producing a fluid space of sonic crossovers.

7.1 Limitations and Future Work

At this stage, ECG bio-signals serve as a single, simple physiological input to the installation. While enabling continuous embodied interaction, it provides only a limited insight into the audience's emotional state. Future work will integrate additional bio-signals, such as respiration or EEG, supported by a fuller data pipeline, to examine complementary contributions of multiple physiological channels and enable richer timbre-centered interaction. *SonoChrom* also remains as a small-scope prototype. Future work will conduct structured user evaluations.

8 Conclusion

In this paper, we argue for technologically and purposefully-shaped timbre as a culturally specific device in contemporary digital tools, and propose unvoiced knowing as a design anchor for timbre-centered interaction. We present *SonoChrom* as a case

study that links (i) a partial-oriented timbral system that supports controllable navigation and hybridization, (ii) a multi-sensory strategy that externalizes timbral motion through real-time visual representation, and (iii) a bio-coupled interaction approach that uses ECG as a continuous modulation signal. We instantiate these contributions through an audiovisual spectral timbre synthesizer grounded in four Chinese instruments (*guzheng*, *erhu*, *suona*, and *tanggu*). *SonoChrom* demonstrates how culturally situated timbral resources can be remodeled into a contemporary synthesized space where audiences explore spectrally re-synthesized timbres and attend to unvoiced knowing as it unfolds through a timbre-first, multi-sensory, culturally situated, and expressivity-focusing engagement that is inspired by the human voice as a "role model" for expressivity in Chinese culture. More broadly, we hope this practice-oriented system of interfacing with culturally specific timbre-first thinking will sharpen awareness of cultural timbre for other researchers and practitioners, and encourage further

work that examines the cultural and experiential dimensions of a sonic attribute as mathematically structured and perceptually complex as timbre.

9 Ethical Standards

The research leading to, and studying, the installation *SonoChrom* and the accompanying present paper strictly followed the institutional and national guidelines for human research ethics, and was funded primarily by the Southern University of Science and Technology, Shenzhen, China. The research did not include any formal study with recruited participants. The exhibition took place in a university-hosted exhibition space, open for free to general visitors. In this semi-public setting, all interactions with audience members were voluntary, transient, spontaneous interactions initiated by the visitors. No data of these interactions was recorded, no names or identifiable elements were collected or stored, therefore, the audience remains anonymous. Under the institutional and national guidelines, for such a kind of observation, no formal ethics approval was required, as it was considered low-risk, as the installation posed minimal to no risk for harm to the visitors.

Acknowledgments

We extend big thanks to the four amateur student musicians producing the sound corpus for *SonoChrom*; the School of Design at Southern University of Science and Technology, Shenzhen, for their continuous support; colleagues, faculty and the administrative team of the School of Design for their open-minded approach to research and for their continuous feedback; and all exhibition visitors for their interest in *SonoChrom*. A special thanks goes to Mirna Zordan at the School of Design for inspiring and advising the interactive visualizations.

References

- [1] Laura-Lee Balkwill and William Forde Thompson. 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music perception* 17, 1 (1999), 43–64.
- [2] Albert S Bregman. 1994. *Auditory scene analysis: The perceptual organization of sound*. MIT press.
- [3] Marcelo Caetano, Charalampos Saitis, and Kai Siedenbueg. 2019. *Audio Content Descriptors of Timbre*. Springer International Publishing, Cham, 297–333. https://doi.org/10.1007/978-3-030-14832-4_11
- [4] Emily I. Dolan. 2013. *The Orchestral Revolution: Haydn and the Technologies of Timbre* (1 ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139235976>
- [5] Pierre Dragicevic. [n. d.]. 3D Spectrogram. Data Physicalization (dataphys.org) entry. <https://dataphys.org/list/3d-spectrogram/> Accessed: 2026-02-05.
- [6] Jun Fei. 2023. Theatre of Emotions. ZCOOL project page. <https://www.zcool.com.cn/work/ZNjg3Mzg5ODQ%3D.html> Accessed: 2026-02-05.
- [7] Steven Feld and Keith H Basso. 1996. Senses of place. (1996).
- [8] Alf Gabrielsson and Patrik N Juslin. 2002. Emotional Expression in Music. In *Handbook of Affective Sciences*. Oxford University Press. <https://doi.org/10.1093/oso/9780195126013.003.0026>
- [9] Julia C Hailstone, Rohani Omar, Susie MD Henley, Chris Frost, Michael G Kenward, and Jason D Warren. 2009. It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly journal of experimental psychology* 62, 11 (2009), 2141–2155.
- [10] Sylvia Kreibitz. 2010. Autonomic Nervous System Activity in Emotion: A Review. *Biological psychology* 84 (04 2010), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- [11] Censong Leng. 2007. Cong yinse hecheng de fangshi kan dianzi yinyue yu Zhongguo yinyue chuantong de lianxi [Connections between electronic music and Chinese musical tradition seen through methods of timbre synthesis]. *Huangzhong (Wuhan Yinyue Xueyuan Xuebao)* 4 (Jan. 2007), 42–47. Publisher (pinyin): Wuhan Yinyue Xueyuan.
- [12] Chenghua Liu. 2023. Renshenghua de yinxiang quwei yu meigan xiaoying — Zhongguo chuantong yinyue yinse de shenmei quxiang [Vocalized sonic interest and aesthetic effects: The aesthetic orientation of timbre in Chinese traditional music]. *Minzu yishu yanjiu* 36, 1 (Feb. 2023), 100–109. <https://doi.org/10.14003/j.cnki.mzysyj.2023.01.11> Publisher (pinyin): Zhoukou Shifan Xueyuan Yinyue Wudao Xueyuan (Jiangsu, Nanjing 210013).
- [13] Stephen McAdams and Bruno L. Giordano. 2016. The Perception of Musical Timbre. In *The Oxford Handbook of Music Psychology*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198722946.013.12>
- [14] Josh H McDermott and Andrew J Oxenham. 2008. Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology* 18, 4 (2008), 452–463. <https://doi.org/10.1016/j.conb.2008.09.005> Sensory systems.
- [15] Michael Polanyi. 2009. The tacit dimension. In *Knowledge in organisations*. Routledge, 135–146.
- [16] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (Dec. 1980), 1161–1178. <https://doi.org/10.1037/h0077714>
- [17] Wayne Slawson. 1981. The Color of Sound: A Theoretical Study in Musical Timbre. *Music Theory Spectrum* 3, 1 (03 1981), 132–141. <https://doi.org/10.2307/746139> arXiv:<https://academic.oup.com/mts/article-pdf/3/1/132/3946931/3-1-132.pdf>
- [18] Laurel J Trainor, Luann Wu, and Christine D Tsang. 2004. Long-term memory for music: Infants remember tempo and timbre. *Developmental science* 7, 3 (2004), 289–296.
- [19] Isabella Van Elferen. 2018. Timbrality: The Vibrant Aesthetics of Tone Color. In *The Oxford Handbook of Timbre* (1 ed.), Emily I. Dolan and Alexander Rehding (Eds.). Oxford University Press, 68–91. <https://doi.org/10.1093/oxfordhb/9780190637224.013.28>
- [20] Xin Wang, Yujia Wei, Lena Heng, and Stephen McAdams. 2021. A cross-cultural analysis of the influence of timbre on affect perception in western classical music and chinese music traditions. *Frontiers in psychology* 12 (2021), 732865.
- [21] Bryan Yueshen Wu and Ke Peng. 2022. Evolution. Personal website project page. <https://bbryan.com/Evolution> Accessed: 2026-02-05.