Qianji: A Resilient Framework for Orchestrating "A Thousand Machines" in Distributed Performance

Ruilei Duan
Zhejiang Conservatory of Music
Department of Music Engneering
Hangzhou, China
drl@zjcm.edu.cn

Zhengyang Kenny Ma*
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong SAR
zmaaf@connect.ust.hk

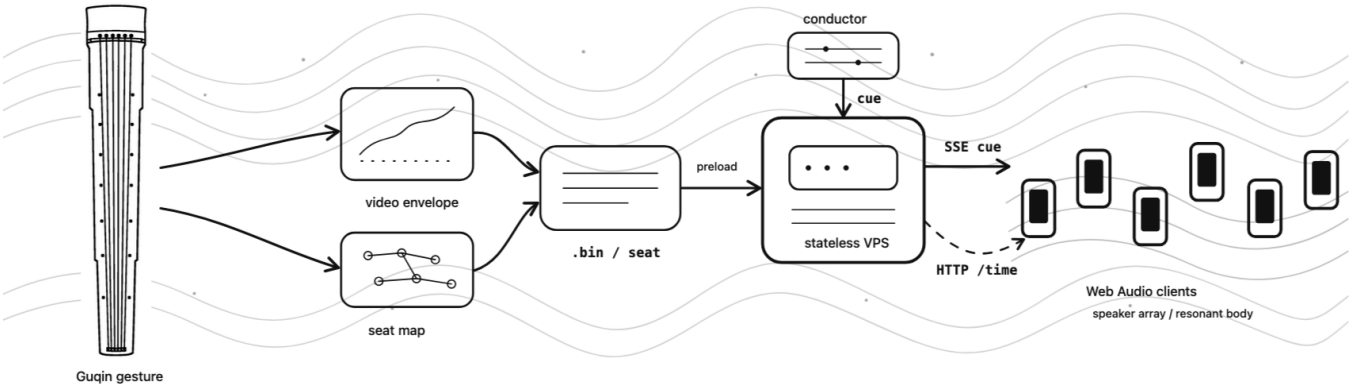

Figure 1: Qianji's resilience-first architecture for transforming Guqin gestures into per-seat scores and broadcasting them through a stateless server to audience smartphones.

Abstract

This paper introduces **Qianji** (literally "A Thousand Machines"), a web-based framework designed for massive distributed audio performance over audience members' own cellular networks. When hundreds of smartphones must function as a synchronized speaker array in a single venue, the congestion and jitter inherent to high-density 4G/5G environments demand an architecture that prioritizes connection survival over bidirectional interactivity. Qianji addresses this with a "resilience-first" unidirectional design utilizing Server-Sent Events (SSE) and stateless HTTP clock synchronization, ensuring textural coherence even under severe network degradation.

We validate this framework through two public performances of "The Discourse of an Instrument, and a Thousand Machines," deploying up to 421 audience smartphones as a distributed resonance chamber for a traditional *Guqin*, with server-side stress tests confirming scalability to 2000+ concurrent connections. To manage the complexity of such an array, we present the "Video-to-Volume" workflow, which allows composers to treat the audience as pixels in a low-resolution display, mapping spatial gestures directly to sound without the overhead of real-time synthesis control. Grounded in the philosophy of *cosmotechnics*, this case study demonstrates how Qianji enables a new scale of acoustic intervention, reconciling the characteristic of the zither with the macro-granular digital presence of the crowd.

CCS Concepts

• **Applied computing** → **Sound and music computing**; • **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods.*

Keywords

Web Audio API, Distributed Performance, Resilience, Server-Sent Events, Massive Interaction, Cosmotechnics, Guqin

ACM Reference Format:
Ruilei Duan and Zhengyang Kenny Ma. 2026. Qianji: A Resilient Framework for Orchestrating "A Thousand Machines" in Distributed Performance. In *Proceedings of New Interfaces for Musical Expression (NIME '26).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The vision of the "audience as a speaker array" has been inspiring computer music researchers for decades. From the transistor radio experiments of the 1970s [15] to the smartphone symphonies of the Mobile Phone Orchestra (MoPhO) [20], the goal has been consistent: to transform a passive crowd into a massive, distributed

*Corresponding author.

NIME '26, London

ACM ISBN 978-1-4503-XXXX-X/26/06
https://doi.org/XXXXXXX.XXXXXXX

sound system. With the ubiquity of modern smartphones and the Web Audio API, this vision is theoretically more accessible than ever. However, a significant gap remains between theory and practice: **reliability at scale**.

Orchestrating several hundred to over a thousand devices in a single physical space introduces chaotic network conditions that remain largely unexplored in existing web audio frameworks. Most current systems rely on stateful, bidirectional connections (e.g., WebSockets), which are designed for interactive ensembles but have not been tested under the heavy packet loss and jitter typical of congested cellular networks (4G/5G) at this scale.

To address this, we introduce **Qianji**, a resilient web-based framework designed specifically for massive distributed audio over audience cellular networks. Qianji adopts a "resilience-first" philosophy: we prioritize **textural coherence** over phase accuracy. By utilizing a unidirectional broadcast architecture (Server-Sent Events) and stateless clock synchronization, the system ensures that the collective instrument remains functional even when individual connections are unstable.

We validate this framework through "The Discourse of an Instrument, and a Thousand Machines," a performance piece featuring a traditional *Guqin*. The choice of the *Guqin*—an instrument of quiet, singular intimacy—serves as a rigorous stress test for the system's ability to maintain a coherent, subtle acoustic environment amidst the digital noise of hundreds of smartphones. Grounded in the notion of *cosmotechnics* [8], this work demonstrates a viable pathway for large-scale acoustic interventions where the audience functions not as a collection of users, but as a unified, resilient body of sound.

The contributions of this paper are as follows:

- **Qianji: A resilient framework for massive distributed performance.** We present a system architecture that decouples command broadcasting (SSE) from clock synchronization (stateless HTTP), enabling reliable orchestration of several hundred concurrent live clients, with server-side capacity validated to over 2000 concurrent connections.
- **A resilience-first synchronization strategy.** We introduce a method for maintaining acoustic coherence in high-jitter environments. By utilizing a multi-stage filtering pipeline—outlier rejection, RTT-based sample selection, and Exponential Moving Average (EMA) smoothing—Qianji prioritizes the stability of musical textures over phase accuracy, ensuring continuous playback even under severe packet delay variation.
- **A validated case study in large-scale distributed performance.** Through two pilot tests and two public concerts deploying up to 421 audience smartphones alongside a live *Guqin*, we demonstrate the practical viability of the framework and contribute an artistic model for treating the audience as a macro-granular resonant body.

2 Related Work

This research intersects with two primary domains of NIME scholarship: the orchestration of distributed audience interaction and the engineering challenges of large-scale web audio deployment.

2.1 The Audience as Speaker Array

The concept of utilizing the audience's personal devices as a distributed loudspeaker array has evolved alongside mobile technology. Early precedents, such as Levin's *Dialtones* [1], relied on the inherent latency and low fidelity of cellular telephony to create coarse-grained sonic textures. As smartphones gained programmable synthesis capabilities, projects like the Mobile Phone Orchestra (MoPhO) [13, 20] demonstrated that these devices could function as expressive instruments for trained ensembles.

In the context of mass participation, frameworks often prioritize *input* aggregation (e.g., voting, collaborative control) over *output* synchronization. Systems like *massMobile* [21] and *Open Symphony* [24] facilitate collective decision-making, treating the phone primarily as a controller. While *echobo* [10] explores the audience as a sound source, it operates within a harmonic framework that tolerates significant latency. In contrast, creating a coherent, continuous acoustic texture across thousands of heterogeneous devices on congested cellular networks requires a synchronization strictness closer to phased speaker arrays than to voting systems, a challenge that remains under-explored.

The challenge of performing music over networks has also received growing attention. Bukvic's *L2Ork Tweeter* [2] addresses latency, synchronization, and bandwidth through a control-data-driven protocol that anticipates future cues, enabling tightly-timed crowdsourced musicking. Xambó and Goudarzi [22] explore the mobile audience as a collective "digital musical persona," using web audio and sensor APIs to aggregate distributed participants into a single sonic identity. Liloia and Dannenberg [11] reframe network latency not as a defect but as a compositional resource for percussive improvisation. Our work shares this pragmatic stance toward network imperfection, but shifts the focus from performer-to-performer collaboration to a one-to-many broadcast scenario where hundreds of co-located audience devices must maintain textural coherence over congested cellular networks.

2.2 Web Audio Network Performance & Scalability

The democratization of the Web Audio API has led to robust frameworks for networked performance. *Soundworks* [12] represents the current state-of-the-art, offering a modular architecture for synchronization and distributed state management via WebSockets. Critically, Soundworks maintains a synchronized copy of each client's state on the server, enabling real-time remote monitoring and control of individual devices—a powerful capability for interactive compositions where a performer dynamically adjusts parameters per-client. However, this design requires two persistent WebSocket connections per client and $O(N)$ server-side state—an architecture well-suited to ensembles of tens to hundreds of devices where bidirectional interactivity is essential. In our scenario—a strictly conductor-driven broadcast to several hundred or more devices where individual client state is irrelevant—this bidirectional overhead becomes unnecessary. Qianji therefore complements frameworks like Soundworks by trading per-client interactivity for a stateless, unidirectional fan-out optimized for the specific constraints of massive one-to-many broadcast over unreliable cellular networks.

Regarding synchronization, Lambert et al. [9] demonstrated that HTML5-based systems can achieve 1–10ms precision using linear regression. However, their evaluation focused on controlled network environments. In the unmanaged, heterogeneous landscape of public cellular networks (4G/5G), packet jitter often exceeds the operational thresholds of traditional synchronization algorithms. Our work addresses this gap with a "resilience-first" architecture: rather than pursuing phase accuracy through persistent connections, we utilize stateless HTTP polling and adaptive filtering to maintain textural coherence under conditions where bidirectional protocols struggle to sustain stable connections.

3 System Design & Interfaces

The Qianji framework addresses the challenge of distributed performance through a "thick-client" philosophy [16]: the server acts as a minimalist signaling broker, while the complexity of synthesis and scheduling is offloaded to the edge devices. This design manifests in three distinct interfaces, each reflecting a different temporal relationship to the performance: the *Compositional Interface* translates spatial gestures into pre-computed per-device scores offline; the *Conductor Console* issues timestamped cues at runtime without requiring continuous feedback; and the *Audience Client*, once initialized, operates as a passive resonator—autonomously receiving and sounding the performance without further user interaction. Figure 2 illustrates the overall system architecture.

3.1 Composition: The Video-to-Volume Workflow

To manage the spatial complexity of hundreds of discrete sound sources, we developed a custom "Stage Editor" application. Distributing real-time spatialization commands to such a large array would require continuous per-client control messages, saturating both the server and the cellular network. Instead, the system shifts this complexity offline through a **Video-to-Volume** pipeline, trading runtime bandwidth for pre-computation.

(1) **Mapping:** The venue seating chart is mapped to a normalized 2D grid (x, y).
(2) **Sampling:** Spatial gestures (e.g., a wave sweeping left-to-right) are rendered as standard grayscale video files. The pipeline samples the pixel luminosity $L(x, y, t)$ for each seat coordinate.
(3) **Encoding:** These values are quantized into 8-bit amplitude envelopes and stored in compact binary files ('.bin').

This workflow effectively treats the audience as a low-resolution "sonic screen," decoupling the complexity of the spatial gesture from the runtime bandwidth limits. The Stage Editor accepts arbitrary (x, y) coordinates rather than enforcing a rectangular grid, so non-standard layouts—curved auditorium rows, theater-in-the-round, free-standing audiences, or multi-tier balconies—are handled by directly editing seat coordinates in the venue map. When no meaningful spatial layout is available (e.g., outdoor mobile crowds), the pipeline degrades gracefully to a purely temporal envelope: every device receives the same luminosity track sampled from a single video pixel, preserving textural choreography without spatialization. Figure 3 illustrates this pipeline.

3.2 Conductor Interface: Asynchronous Control

The Conductor Console is designed for reliability over interactivity. It utilizes a "fire-and-forget" mechanism [17], broadcasting timestamped cues ($T_{exec} = T_{now} + \delta$) with a significant safety buffer ($\delta \approx 2000ms$). This interface allows the performer to trigger complex, pre-loaded spatial textures without maintaining continuous control streams, ensuring that the conductor's intent is executed regardless of momentary network latency.

3.3 Audience Interface: The Resonator

The design of this interface draws inspiration from the acoustic tradition of the Chinese *Guqin*, where the instrument is placed upon a resonating surface—a wooden table (*qinzhuo*), or a natural medium such as stone—whose material properties passively shape and amplify the zither's tone [18, 19, 23]. We adopt this principle as a design metaphor: each audience device functions as a discrete passive resonator, receiving and sounding the performance throughout the Qianji System.

The client-side application runs within any standard mobile WebView, accessed via a QR code scan from the audience's native camera or messaging app. Upon entry, the audience member self-reports their seat through a structured three-level selector (section, row, seat number), which maps the device to its pre-computed (x, y) coordinate and corresponding `.bin` spatial score. To realize this passive resonator model, the interface adopts a **Zero-Interaction** design. Once the "Join" button is pressed, the screen displays a dynamic luminosity field that mirrors the device's current amplitude envelope—brightening and dimming in sync with the audio gain. The client autonomously downloads its specific `.bin` file and schedules audio playback locally upon receiving SSE triggers, ensuring acoustic continuity even if the connection drops mid-phrase.

4 Engineering Resilience

Deploying synchronized audio on hundreds of audience devices requires navigating a hostile environment of congested cellular networks and aggressive mobile OS resource throttling. Qianji achieves stability through a multi-layered resilience strategy covering network transport, temporal synchronization, and device state management.

4.1 Layer 1: Network Resilience (SSE)

A key architectural decision in Qianji is the choice of Server-Sent Events (SSE) over WebSockets for the command broadcast channel. While both protocols maintain persistent connections, they differ in ways that matter at scale [4]. WebSocket establishes a full-duplex channel via an HTTP Upgrade handshake, requiring the server to maintain per-connection frame buffers and bidirectional state; each connection also demands application-level heartbeat (ping/pong) logic to detect failures. SSE, by contrast, operates as a standard HTTP response stream: the server simply writes `data:` lines to an open connection, with no framing overhead, no client-to-server masking, and no upgrade negotiation.

For our scenario—a strictly one-to-many broadcast where clients never send data upstream—this difference is decisive. SSE

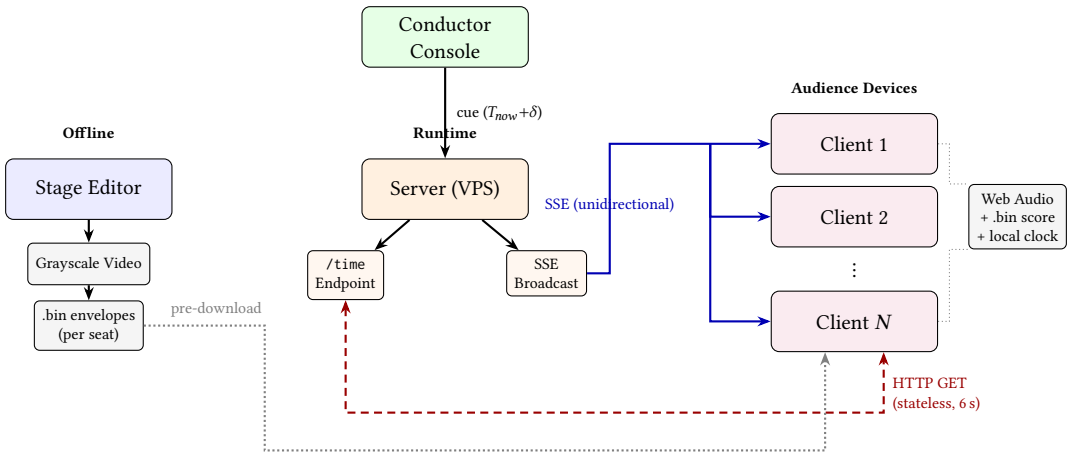

Figure 2: Qianji system architecture. Command broadcasting (SSE, blue) and clock synchronization (HTTP, red dashed) operate on separate channels. The server maintains no per-client state. Clients autonomously download their spatial score (.bin) and schedule playback locally.

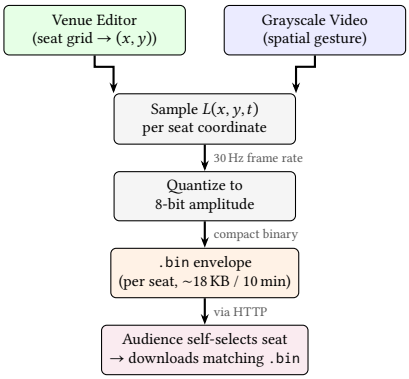

Figure 3: The Video-to-Volume pipeline. The Venue Editor defines seat coordinates; grayscale video encodes spatial gestures. Both are merged offline to produce per-seat binary envelopes. At runtime, each audience member self-selects their seat and downloads the corresponding `.bin` file.

reduces the server to a stateless fan-out operation, and benchmarks at 10,000 concurrent connections show approximately 60% lower memory consumption compared to WebSocket due to the absence of bidirectional frame buffers. Most critically, the browser's native `EventSource` API provides automatic reconnection with configurable backoff, handled entirely by the browser engine. When a device on a congested 4G/5G network loses its connection during a cell tower handoff, the `EventSource` transparently re-establishes the stream without any application-level intervention—a behavior that WebSocket requires developers to implement manually.

4.2 Layer 2: Temporal Resilience (Adaptive Filtering)

Clock synchronization operates on a channel entirely separate from SSE: clients issue independent, stateless HTTP `GET` requests to a `/time` endpoint every 6 seconds, exchanging four Network Time Protocol (NTP)-style timestamps to compute clock offsets. This decoupled design ensures that synchronization traffic does not interfere with the command broadcast channel. However, raw offset values in public cellular networks are highly noisy. We employ a multi-stage filtering pipeline to estimate the true clock offset from noisy measurements. First, samples whose Round-Trip Time (RTT) deviates beyond 2σ from the cycle mean are rejected. The remaining samples are ranked by RTT, and only the best 50% are retained. Finally, the per-cycle mean offset is smoothed via an **Exponential Moving Average** (EMA, $\alpha = 0.3$):

$$\hat{x}_k = \alpha \cdot z_k + (1 - \alpha) \cdot \hat{x}_{k-1} \tag{1}$$

This progressive filtering—from 15 raw NTP samples down to a single stable estimate—ensures that transient network spikes do not propagate into the synchronization clock. The goal is not sub-millisecond precision but *stability*: maintaining a usable estimate even when network conditions degrade severely, so that the textural coherence of the collective instrument is preserved. Figure 4 illustrates this multi-stage noise reduction under two representative jitter conditions, simulated using the deployed parameters.

4.3 Layer 3: Device Resilience (Mobile Contexts)

Mobile operating systems aggressively suspend background audio threads to save battery. To prevent the "instrument" from falling asleep, we implement two specific countermeasures:

- **Wake Locks & Silent Video:** We utilize the Screen Wake Lock API combined with a looped, muted 1×1 pixel video element. This forces the OS to classify the browser as an active media application, preventing background suspension.
- **Visibility Recovery:** If a user switches apps, the audio clock may drift. We monitor the `visibilitychange` event; upon return, the system calculates the drift Δ_t. If $\Delta_t > 500ms$, the playhead hard-seeks to the server time; if $\Delta_t <$

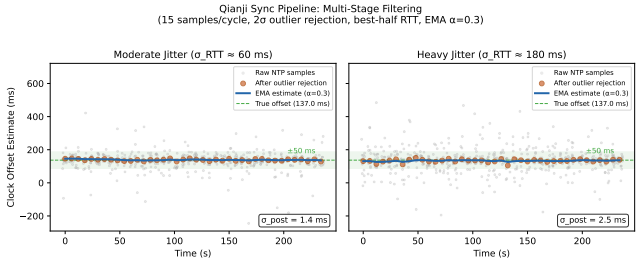

Figure 4: Multi-stage sync pipeline under moderate ($\sigma_{RTT} \approx 60$ ms) and heavy ($\sigma_{RTT} \approx 180$ ms) jitter. Gray: raw NTP samples (15 per cycle); orange: per-cycle mean after 2σ outlier rejection and best-half RTT selection; blue: EMA-smoothed estimate (α=0.3). Simulated using deployed parameters. Post-convergence $\sigma < 3$ ms in both scenarios.

500*ms*, a variable playback rate is applied to smoothly realign the phase without audible pitch artifacts.

Additionally, we implemented platform-specific adaptations to account for differences in how iOS and Android handle the Web Audio API—including audio context suspension policies, volume control interfaces, and output buffer sizes.

5 Case Study: The Discourse of an Instrument, and a Thousand Machines

To validate the resilience of the Qianji framework, we developed "The Discourse of an Instrument, and a Thousand Machines," a live piece accompanying a *Guqin* performance with hundreds of audience smartphones. This piece serves not only as a technical stress test but as an artistic exploration of the tension between the singular acoustic authority of the zither and the distributed digital presence of the crowd.

Research into the *Guqin* within NIME has largely focused on gesture acquisition [5] or augmented reality pedagogy [25]. These projects typically enhance the instrument's input capabilities or visual feedback. Our work, however, focuses on the instrument's relationship to its environment. In this work, we do not aim to augment the instrument's playability, but rather to recontextualize its cultural function within a networked acoustic ecosystem.

5.1 Artistic Intent: Resonance over Melody

The *Guqin* is historically an instrument of quiet intimacy, played not only in the scholar's studio but in mountain retreats and secluded gardens, where the surrounding environment—stone, wood, water—served as a resonating extension of its delicate voice.[1] Our artistic intent was to reconstruct this relationship at the scale of the venue: rather than treating the smartphones as independent instruments that "play along" with the zither, we designed the audience array to function as the *Guqin*'s **Resonant Body**—a distributed *qinzhuo* (琴桌) that receives and extends the instrument's resonance across the entire hall.

[1]See James Watt, "The *Qin* and the Chinese Literati," *Orientations Magazine*, Nov. 1981, pp. 38–49. Available at https://www.silkqin.com/10ideo/wattart.htm.

The phones thus form not a counterpoint but a reactive environment: a digital *shanshui* (山水, "mountains and water"), the Chinese literati ideal of landscape as a site of spiritual communion between self and nature [6], here reconstructed as a networked acoustic ecology that responds to the zither's gestures. This demanded a sound design strategy that could absorb temporal jitter (±50*ms*) without collapsing into incoherence, leading us to adopt what we term a *macro-granular* approach: while each device plays a continuous audio sample rather than performing granular synthesis internally, the audience as a whole functions as a macro-scale granular instrument, with each phone acting as an independently enveloped grain source [14].

5.2 Realization: Macro-Granular Texture

Leveraging Qianji's offline "Video-to-Volume" workflow, we implemented this macro-granular texture across the audience array.

- **The Audience as Grains:** Each smartphone acts as a single grain generator. The source material consists of recordings of the *Guqin* itself, processed through GRM Tools [7]—time-stretched, reversed, and spectrally shaped to produce dense, evolving timbres that retain the spectral character of the original instrument while becoming unrecognizable as discrete notes. The resulting textures are concentrated in the mid-to-high frequency range where human hearing is most sensitive, ensuring perceptual clarity even at low playback volumes from smartphone speakers.
- **The Video as Envelope:** We composed the spatial movement using grayscale motion graphics. A "cloud" moving across the video frame translates to a wave of amplitude envelopes opening and closing across the audience. The screen of each device mirrors this envelope as a dynamic luminosity field, creating a synchronized visual counterpart to the sound.

Because the system allows for pre-calculated spatial data, we could orchestrate complex, organic textures—such as a "rain" of sound sweeping from the back of the hall to the stage—that would be impossible to calculate in real-time. The inherent network jitter, rather than being a defect, served to "blur" the edges of these grains, softening the digital array into a cohesive, organic texture that supported rather than overpowered the acoustic instrument.

5.3 Agency and Cosmotechnics

This interaction model establishes a feedback loop reminiscent of Freeman's *Glimmer* [3], where audience behavior and orchestral performance mutually influence one another. However, where *Glimmer* used light sticks to aggregate audience input into orchestral cues, our system inverts the flow: the singular authority of the *Guqin* is diffracted through the digital lens of the crowd.

Drawing on Kuzmin et al.'s framework of *musical cosmotechnics* [8], this design recontextualizes the traditional zither. The *Guqin*, historically an instrument of solitary cultivation, here becomes a mechanism for communal resonance. The "Video-to-Volume" pipeline serves as the technical mediator of this cosmology, translating the gestural intent of the performer into a thousand discrete points of light and sound, blurring the boundary between the instrument and the audience-as-environment.

5.4 Deployment & Technical Evaluation

The Qianji server was deployed on a 2-core, 4GB Virtual Private Server (VPS) with 200Mbps peak bandwidth. Audience devices connected via their own cellular networks (4G/5G), accessing the client application through a QR code scanned with their phone's native camera or messaging app. To preserve the zither's acoustic presence against the distributed array, the *Guqin* was amplified using two boundary microphones and one small-diaphragm condenser, routed to a left-center-right speaker array at the front of the audience. A Sennheiser AMBEO VR Mic was placed at the center of the venue to capture the spatial audio field of the distributed array.

We validated the system through two pilot tests. The first ($N = 100$) was conducted in a university lecture hall with the students to verify end-to-end integration across a diverse range of device models and OS versions. The second pilot ($N = 50$) served as both a perceptual evaluation of synchronization quality and a server-side stress test: alongside the 50 physical devices, we ran concurrent simulated connections from five geographically distributed machines under varying network conditions, confirming stable SSE operation at 2000+ simultaneous clients.

Following these pilots, the system was deployed in two public concert performances of "The Discourse of an Instrument, and a Thousand Machines" in a 500-seat venue. Server logs recorded approximately 84% of the seated audience as active connections ($N = 421$ and $N \approx 320$ respectively). Both performances ran without critical failure, with the SSE reconnection mechanism transparently recovering devices that experienced momentary cellular dropouts. From the audience perspective, the spatial gestures—sweeping waves and granular clouds moving across the hall—were perceptibly coherent, with no audible tearing or temporal discontinuity between neighboring devices.[2]

6 Discussion

6.1 The Audience as Resonant Body: An Organological Inquiry

In the history of the "audience as speaker array" [15], agency is typically granted to the audience-member as a performer. Our approach with **Qianji** diverges by deliberately stripping the audience of performative agency. We frame this not as a loss of control, but as an act of "borrowing" the ubiquity of personal devices to form a unified acoustic ecosystem.

The choice of the unidirectional SSE protocol is central to this inquiry. It does not merely solve a bandwidth problem; it acts as a form of technological *Aufhebung* (sublation). By utilizing the medium's capacity for massive, one-way broadcasting, we preserve the smartphone's physical capacity for sound generation while canceling its inherent tendency towards bidirectional distraction. The device is effectively "hollowed out" of its social media functions, transforming from a portal of communication into a pure vessel of resonance.

[2]A spatial audio recording of the live performance, project documentation, and source code are available at: https://zmk5566.github.io/qianji/.

Crucially, this technical intervention reconfigures the audience's subjective experience. Although the individual momentarily "loses" their device—surrendering it to the collective network—this dispossession is not a silence, but an opening. It is precisely this suspension of digital control that marks the inception of listening. Enmeshed within the resonant web, the audience member is liberated from the impulse to broadcast, allowing them to inhabit the network not as a user, but as a listener.

6.2 Toward a Musical Cosmotechnics

This transformation is not generic; it is tuned to the specific locality of the performance. Drawing on Kuzmin et al.'s framework of *musical cosmotechnics* [8]—developed through their bamboo-based DMI *zhu nao*, which embeds environmental data into instrument design—we position our system as a parallel exploration: where *zhu nao* **internalizes** locality within a single instrument, our system **externalizes** it across a distributed crowd.

The *Guqin* is historically inseparable from the Chinese literati tradition of *shanshui* (山水)—the contemplation of landscape as a site of spiritual cultivation. The scholar plays not to an audience but to mountains and rivers; the instrument's quiet voice is calibrated to intimate space, relying on the environment to carry the tone. Our system reinterprets this cosmology in the digital age. The hundreds of smartphones, having been sublated into a Resonant Body, become a **Digital *Shanshui***—a synthetic landscape that the *Guqin* addresses. The unidirectional SSE architecture mirrors this relationship: the instrument speaks, the landscape resonates, but does not talk back. In this silence of the interface, the technology mediates a return to the acoustic ecology of the scholar's studio, reconciling the granular ubiquity of the machine with the singular authority of the tradition.

6.3 Limitations

While our stress tests confirm server-side scalability to 2000+ connections and our largest live deployment reached 421 devices, the emergent acoustic properties of significantly larger arrays—aggregate sound pressure, spatial masking, and inter-device phase interactions—remain unexplored. Scaling from hundreds to thousands of simultaneously sounding devices in a single venue presents challenges that cannot be fully predicted from smaller deployments.

Additionally, we acknowledge the absence of quantitative synchronization data from the live performances. As discussed in our ethical considerations, we prioritized lowering the barrier to audience participation over data collection; instrumenting client devices with logging would have introduced additional consent requirements and technical overhead that risked reducing participation rates. The simulated convergence analysis (Figure 4) uses parameters extracted directly from the deployed codebase, but real-world cellular conditions may differ. Future deployments will explore opt-in, anonymized telemetry to bridge this gap.

We also note two comparative evaluations that this paper does not provide. First, we have not run a controlled head-to-head benchmark of SSE against WebSockets under matched high-jitter conditions; our protocol choice rests on architectural reasoning (stateless fan-out, native browser reconnection) and published

memory benchmarks rather than an in-house bake-off. Second, we have not directly compared our multi-stage filtering pipeline against established synchronization schemes such as those used in Soundworks or the linear-regression approach of Lambert et al. on identical cellular traces. Both comparisons would strengthen the empirical case for the resilience-first design and are planned for future work, ideally on instrumented opt-in deployments where ground-truth offsets can be recovered without compromising audience privacy.

6.4 Ethical Considerations

All participants in both pilot tests and public performances were informed in advance that they were taking part in a research study involving their personal devices. Participation was voluntary; audience members could decline by simply not scanning the QR code.

By design, the Qianji system does not collect any personally identifiable information from audience devices—no IP addresses, device identifiers, carrier information, or usage analytics are recorded. This privacy-by-design principle is also the reason no client-side synchronization logs were retained from the live performances: we chose not to instrument the audience's devices beyond what was strictly necessary for playback. The source code of the framework, together with project documentation and a spatial audio recording of the performance, is publicly available at the project page.[3]

6.5 Future Work

While the current deployment uses a strictly offline Video-to-Volume workflow for stability, the underlying SSE architecture is capable of real-time parameter control—the server can broadcast updated amplitude envelopes or spatial mappings on the fly. Future iterations will explore this online mode, enabling live compositional gestures that respond to the performer's actions in real time. We also plan to implement scheduled command sequences, functioning like a distributed "loopbox," where audience feedback loops can be pre-scheduled to avoid network congestion. Additionally, we aim to introduce mechanisms for audience choice, allowing individuals to vote on or adjust global performance parameters, thereby reintroducing a layer of agency into the resonant ecosystem without compromising the stability of the broadcast clock.

7 Conclusion

This work demonstrates that the web platform, when architected with resilient unidirectional data flows and adaptive client-side filtering, is capable of sustaining massive-scale acoustic interventions where traditional WebSocket-based approaches falter. Through two pilot tests and two public performances of "The Discourse of an Instrument, and a Thousand Machines," we validate Qianji not merely as a theoretical architecture, but as a production-ready framework for distributed performance. This technical mediation embodies a form of musical cosmotechnics [8], reconciling the singular, intimate authority of the *Guqin* with the granular ubiquity of the digital crowd. Ultimately, by transforming the audience from passive spectators into the synchronized resonant body of the performance, we suggest that the future of networked music lies not in complex hardware, but in the precise, communal alignment of the technology already in our hands.

[3] https://zmk5566.github.io/qianji/

References

[1] [n. d.]. Golan Levin : Dialtones (A Telesymphony). https://www.fondation-langlois.org/html/e/page.php?NumPage=229

[2] Ivica Ico Bukvic. 2022. Latency-, Sync-, and Bandwidth-Agnostic Tightly-Timed Telematic and Crowdsourced Musicking Made Possible Using L2Ork Tweeter. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. doi:10.21428/92fbeb44.a0a8d914

[3] Jason Freeman. 2005. Large audience participation, technology, and orchestral performance. In *ICMC*.

[4] Ilya Grigorik. 2013. *High Performance Browser Networking*. O'Reilly Media.

[5] Jingyin He, Ajay Kapur, and Dale A Carnegie. 2015. Developing A Physical Gesture Acquisition System for Guqin Performance. (2015).

[6] Yuk Hui. 2021. *Art and Cosmotechnics*. University of Minnesota Press.

[7] INA-GRM. [n. d.]. GRM Tools. https://inagrm.com/en/store Audio effects suite developed by Groupe de Recherches Musicales.

[8] Iurii Kuzmin, Zhengyang Ma, and Raul Masu. 2024. Toward Musical Cosmotechnics: the case of zhu nao 竹脑—a bamboo-based instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 582–590.

[9] Jean-Philippe Lambert, Sébastien Robaszkiewicz, and Norbert Schnell. 2016. Synchronisation for distributed audio rendering over heterogeneous devices, in html5. In *2nd Web Audio Conference*.

[10] Sang Won Lee and Jason Freeman. 2013. echobo: A mobile music instrument designed for audience to play. *Ann Arbor* 1001, 48109-2121 (2013), 10–44.

[11] Ari Liloia and Roger Dannenberg. 2024. Exploiting Latency In The Design Of A Networked Music Performance System For Percussive Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

[12] Benjamin Matuszewski. 2020. A Web-Based Framework for Distributed Music System Research and Creation. *Journal of the Audio Engineering Society* 68, 10 (Dec. 2020), 717–726. doi:10.17743/jaes.2020.0015

[13] Jieun Oh, Jorge Herrera, Nicholas J Bryan, Luke Dahl, and Ge Wang. 2010. Evolving The Mobile Phone Orchestra. (2010).

[14] Curtis Roads. 2001. *Microsound*. MIT Press.

[15] Benjamin Taylor. 2017. A history of the audience as a speaker array. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 481–486.

[16] Maarten Van Steen and Andrew S. Tanenbaum. 2023. *Distributed Systems* (4th ed.). distributed-systems.net.

[17] Markus Völter, Michael Kircher, and Uwe Zdun. 2004. *Remoting Patterns: Foundations of Enterprise, Internet and Realtime Distributed Object Middleware*. John Wiley & Sons. doi:10.1002/0471726877

[18] Chris Waltham, Kimi Coaldrake, Evert Koster, and Yang Lan. 2016. Acoustics of the Qin. In *Studies in Musical Acoustics and Psychoacoustics*. Springer, 49–74.

[19] Chris Waltham, Yang Lan, and Evert Koster. 2016. An acoustical study of the qin. *The Journal of the Acoustical Society of America* 139, 4 (2016), 1592–1600.

[20] Ge Wang, Georg Essl, and Henri Penttinen. 2008. Do mobile phones dream of electric orchestras?. In *ICMC*.

[21] Nathan Weitzner, Jason Freeman, Stephen Garrett, and Yan-Ling Chen. 2012. massMobile-an Audience Participation Framework.. In *NIME*, Vol. 12. 21–23.

[22] Anna Xambó and Visda Goudarzi. 2022. The Mobile Audience as a Digital Musical Persona in Telematic Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. doi:10.21428/92fbeb44.706b549e

[23] Ouyang Xiao. 2017. Van Gulik's The Lore of the Chinese Lute Revisited. *Monumenta Serica* 65, 1 (2017), 147–174.

[24] Leshao Zhang, Yongmeng Wu, Mathieu Barthet, et al. 2016. A web application for audience participation in live music performance: The open symphony use case. (2016).

[25] Yingxue Zhang, Siqi Liu, Lu Tao, Chun Yu, Yuanchun Shi, and Yingqing Xu. 2015. ChinAR: Facilitating Chinese Guqin Learning through Interactive Projected Augmentation. In *Proceedings of the Third International Symposium of Chinese CHI*. ACM, Seoul Republic of Korea, 23–31. doi:10.1145/2739999.2740003