iXeRemin: Designing a Polyphonic MR Instrument for Artistic Performance and Interaction

Gwangyu Lee
gwangyulee@dongguk.edu
MARTE Lab. Dept. of Multimedia, Dongguk University
Seoul, South Korea

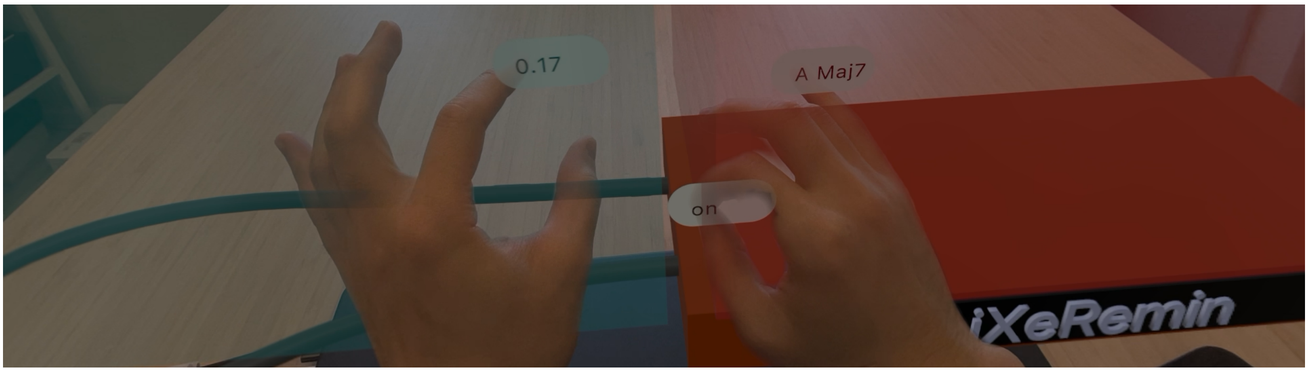


Figure 1: Performing iXeRemin in mixed reality: playing an A major 7 chord through spatial hand gestures.

Abstract

This paper introduces iXeRemin, a theremin-inspired mixed reality (MR) instrument for Apple Vision Pro that combines hand tracking, frequency modulation (FM) synthesis, chordal gesture mapping, fingertip labels, and Open Sound Control (OSC) output. The right hand controls pitch, note triggering, and vibrato, while the left hand controls amplitude, FM index, and harmonic mode. Chords are selected through left index-thumb distance, with seventh chords activated by left pinky-thumb distance. Real-time 3D labels display amplitude, note state, and pitch or chord information near the performer's fingertips. Rather than claiming a direct continuation of traditional theremin technique, iXeRemin uses the theremin as a conceptual reference for touchless musical control and adapts it to a spatial computing context. The contribution of this short paper is the design and public release of a polyphonic MR instrument that connects gestural sound control with visual feedback and OSC-based interoperability.

Keywords

Mixed reality musical instruments, Polyphonic theremin, Hand gesture control, FM synthesis, Spatial computing

1 Overview

iXeRemin is a mixed reality (MR) musical instrument developed for Apple Vision Pro. It draws on the theremin as a reference point for touchless musical interaction, but its mappings are designed for MR hand tracking rather than as a direct reproduction of traditional theremin technique. The system combines FM synthesis, fingertip-based visual feedback, chordal hand gestures, and OSC output for use with external audiovisual environments.

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

In relation to earlier extended-reality theremin prototypes such as VRMin [2] and MR:emin [1], iXeRemin emphasizes polyphonic control, visible performance parameters, and public availability. The system was publicly released on the Apple App Store on June 10, 2025 [3], and is available for download. A demonstration project file for receiving and utilizing OSC data through Max/MSP or TouchDesigner is also available on GitHub [4]. These resources support reuse, inspection, and further exploration in artistic and technical contexts.

1.1 Background and Related Work

The theremin provides an important historical reference for touchless musical control, while contemporary extended reality (XR)-based systems have explored how spatial visualization can reshape musical interaction. Research on mixed-reality musical interfaces demonstrates how visual augmentation can support spatial awareness and performer orientation [1, 2]. Santini's work on augmented piano interfaces highlights the importance of accurate spatial mapping and real-time feedback for intuitive performance [6]. Furthermore, Li and Kristensson [5] describe *sensorimotor regularities* as relationships between bodily action and sensory response in MR environments. In this paper, that work is used as a design inspiration for making relationships between hand motion, sound, and visual feedback legible, rather than as a formal evaluation framework.

2 Design of iXeRemin

The design of iXeRemin adapts touchless musical control to an MR setting through three linked concerns: continuous sound control, visible spatial feedback, and chordal hand interaction. Implemented for Apple Vision Pro, the system uses real-time hand tracking to map the right hand to pitch, note triggering, and vibrato, while the left hand controls amplitude, FM index, and chord selection. This two-handed design is intended to support melodic, timbral, and harmonic control within the same visible performance space.

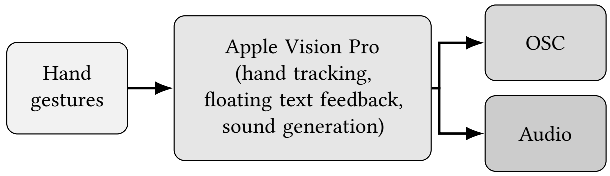


Figure 2: System overview of iXeRemin showing data flow from hand gestures to OSC and audio output.

2.1 Overall Structure

The iXeRemin system is organized into three interconnected modules operating in real time: (1) a hand-gesture interface that translates spatial movement into musical control of pitch, amplitude, vibrato, and harmonic mode; (2) an FM synthesis engine implemented in Swift and integrated with AVAudioEngine, generating continuously modulated sound shaped by the performer's gestures; and (3) an OSC networking layer that enables iXeRemin to function as an MR-based OSC controller, transmitting performance data-such as hand position, fingertip distances, and synthesis parameters-to external platforms including Max/MSP and TouchDesigner.

Together, these modules connect physical motion, auditory feedback, and networked communication within a unified MR environment. The overall system architecture is shown in Figure 2, illustrating the data flow from Apple Vision Pro hand tracking to the FM synthesis engine and OSC output modules.

2.2 UX/UI Design

The UX/UI design of iXeRemin focuses on making touchless performance space visible, interactive, and configurable within mixed reality. It integrates spatial control, visual feedback, and parameter customization into a coherent environment for gesture performance. The system consists of four key components: (1) a boundary box that defines and visualizes the active performance region, (2) fingertip labels that provide real-time visual feedback, (3) gesture mappings informed by action-response relationships, and (4) a settings window for detailed configuration of synthesis and communication parameters. Together, these elements create a spatially aware interface for MR performance.

2.2.1 Boundary Box and Spatial Interaction. A semi-transparent boundary box defines the active performance area. Sound is generated only when the relevant fingers (both left and right) are fully within the boundary; gestures outside the box do not trigger notes. This spatial constraint ensures precise gesture control and prevents unintended activation, following the "accurate space sampling" design strategy proposed by Santini in augmented reality instrument design [6]. Additionally, performers can treat the boundary itself as a performative element-moving their hands into and out of the box region to intentionally trigger or shape sound events. Figure 3 shows an overview of the boundary box during performance, while Figure 4 illustrates the four possible gesture states (box-in pinch, box-in release, box-out pinch, and box-out release) that determine whether sound is triggered or suppressed.

2.2.2 Visual Feedback through Fingertip Labels. Real-time 3D text labels appear near the performer's fingertips to display performance data: the left index fingertip shows the current amplitude (amp), the right index fingertip indicates the note on/off state, and the right pinky fingertip displays the current pitch or chord name (e.g., C, C major, C minor). These annotations float naturally with finger movement, providing immediate visual feedback without obstructing gesture recognition. To clearly indicate whether the hand is inside or outside the boundary box, the labels dynamically change their content based on spatial context: when the fingers are inside the box, the labels show numerical amplitude values and explicit note-on or note-off states, whereas outside the box they revert to generic placeholders such as "amp" and "note". Similarly, the fingertip labels for harmony control continuously reflect the current pitch or chord name, ensuring that performers can always verify which chord type is active during performance. This inside-outside distinction is illustrated in the gesture-state examples shown in Figure 4.

Figure 3: Screenshot of the iXeRemin interface showing the semi-transparent boundary box and real-time fingertip labels for amplitude, note-on/off state, pitch or chord name.

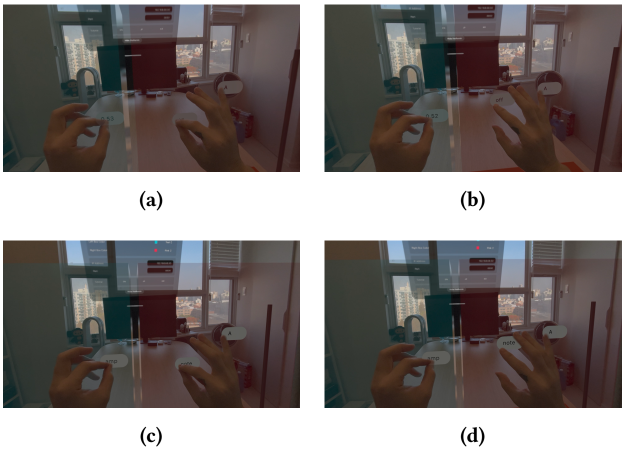


Figure 4: Examples of gesture states relative to the boundary box: (a) box-in pinch, (b) box-in release, (c) box-out pinch, and (d) box-out release. Sound is produced only in case (a) when a pinch gesture is performed inside the boundary box.

2.2.3 Gesture Mapping. The gesture-to-sound mapping is informed by the broader idea that action and response should be legible to the performer [5]. For instance, closing the right-hand pinch gesture triggers a note-on event, while opening the hand stops the sound. A detailed description of the complete gesture-to-sound mapping strategy is provided in Sound Design.

Figure 5: The iXeRemin settings window.

2.2.4 Settings Window for Parameter Control. In addition to the gesture-based interface, iXeRemin includes a dedicated settings window that allows performers to configure instrument parameters in real time. The settings window provides several real-time configuration options: (1) sliders for adjusting the spatial position of the iXeRemin system, (2) waveform selection and a toggle for FM synthesis, (3) ADSR envelope controls, (4) customizable boundary box colors, and (5) input fields for OSC IP address and port number. This configuration panel bridges physical and digital performance layers, offering both expressive immediacy through gesture and refined control through parameter adjustment, as illustrated in Figure 5.

2.3 Sound Design

The sound generation of iXeRemin is based on an FM synthesis model that allows continuous parameter variation. The current prototype uses a fixed modulator frequency of 220 Hz as a stable reference across synthesis modes. This value is a design choice for the present implementation rather than an optimized synthesis parameter; future versions will expose modulator frequency or ratio settings for performer control.

2.3.1 Pitch and Vibrato Control. The right pinky fingertip's horizontal (x-axis) position controls the carrier frequency, ranging from 220 Hz (A3) to 1760 Hz (A6), thus covering approximately three octaves. The right pinky fingertip's vertical (y-axis) movement simultaneously modulates both the vibrato rate and vibrato depth in real time. The vibrato rate is mapped from 0 Hz to 10 Hz, while the vibrato depth varies proportionally between 0% and 20%, allowing both speed and intensity of modulation to respond continuously to hand motion. This relative, percentage-based mapping lets performers shape vibrato through a single gesture rather than adjusting multiple independent parameters. Representative examples of these pitch and vibrato mappings are shown in Figure 6.

2.3.2 FM Index and Amplitude Control. The left index fingertip's horizontal (x-axis) position controls the FM index within a range of 0 to 25, shaping harmonic complexity and spectral brightness. Meanwhile, its vertical (y-axis) position determines amplitude, enabling continuous loudness control through hand movement. Together, these spatial mappings connect timbral and dynamic modulation within the same left-hand gesture. Representative examples of these amplitude and FM index mappings are shown in Figure 7.

2.3.3 Waveform Selection and Timbre Configuration. Within the settings window, performers can select the waveform type-sine, sawtooth, triangle, or rectangle-to define the instrument's tonal character. FM synthesis can also be enabled or disabled through the settings window; when enabled, it uses the fixed modulator frequency described above. Synthesis parameters are smoothed through envelope interpolation to reduce audible discontinuities between dynamic and harmonic states.

2.4 Chord and Polyphonic Control System

A central innovation in iXeRemin is the gesture-driven chord and polyphonic control system. The performer's left-hand index-thumb distance determines whether a single tone or a chord is played, with the distance range (1-18 cm) divided into six discrete harmonic states: single tone, major, minor, sus4, diminished, and augmented. Each chord is generated through multiple FM tone layers tuned according to the selected chord's interval structure-for example, major chords use [0, +4, +7] semitone offsets, while minor chords use [0, +3, +7].

Additionally, the system extends this mapping by incorporating the distance between the left pinky and thumb to control seventh chords. When the performer is playing a single tone and the pinky-thumb distance exceeds 5 cm, a dominant seventh chord is automatically triggered. Similarly, when performing a major chord, exceeding this distance activates a major seventh chord. The same logic applies across other chord modes, allowing performers to access extended harmonies through changes in left-hand posture. This enables smooth harmonic expansion within the same spatial gesture framework, providing expressive continuity between triads and seventh chords. Illustrative examples of these distance-based chord transitions are shown in Figures 8 and 9.

3 OSC Output and External Use

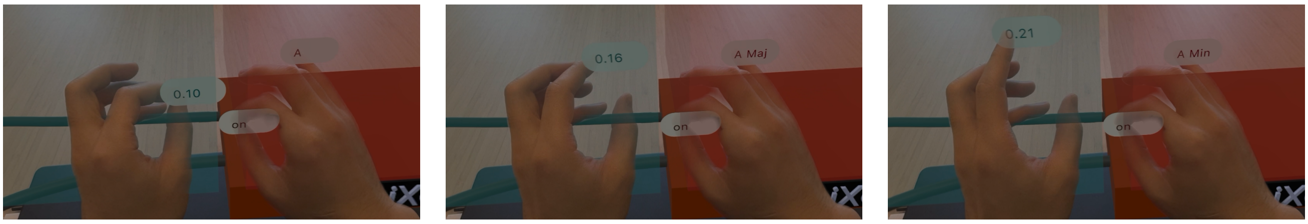
iXeRemin transmits key performance parameters via OSC for external communication. These parameters include normalized fingertip positions-specifically, the left index fingertip's (x, y) values controlling FM index and amplitude, and the right pinky fingertip's (x, y) values controlling pitch and vibrato. Additional

(a) (b) (c)

Figure 6: Right pinky fingertip control: (a) E4, low vibrato; (b) G6, low vibrato; (c) G6, high vibrato.

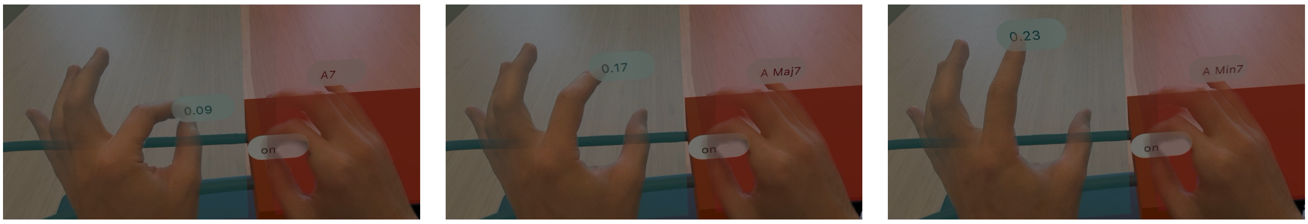
(a) (b) (c)

Figure 7: Left index fingertip control: (a) amp=0.87, low FM; (b) amp=0.17, low FM; (c) amp=0.18, high FM.



(a) (b) (c)

Figure 8: Triad selection via left index-thumb distance: (a) A (1-5 cm), (b) A major (5-9 cm), (c) A minor (9-13 cm).



(a) (b) (c)

Figure 9: Seventh-chord extension via left pinky-thumb distance: (a) A7, (b) Amaj7, (c) Amin7 (activated when pinky-thumb distance exceeds 5 cm).

OSC messages transmit pitch (frequency), note-on/off state, vibrato rate, vibrato depth, amplitude, FM index, ADSR envelope values, waveform selection, and chord mode. This allows iXeRemin to operate as a standalone MR instrument or as an OSC-based controller for external synthesis and audiovisual systems, e.g., Max/MSP, TouchDesigner, or custom software. Multiple iXeRemin devices can also send OSC data to a single host computer, enabling ensemble configurations in which streams from different performers are combined or reinterpreted.

4 Conclusion

iXeRemin presents a theremin-inspired MR instrument for Apple Vision Pro that combines continuous hand tracking, FM synthesis, chordal gesture mapping, fingertip labels, and OSC output. Its main contribution is a publicly available design that makes pitch, amplitude, timbre, and harmonic state visible and controllable around the performer's hands.

The current work should be understood as a design-oriented short paper rather than a formal evaluation study. Because no performer study is reported here, claims about learnability, gestural precision, and long-term performance practice remain provisional. The current implementation also depends on Apple Vision Pro, whose cost, availability, weight, and field-of-view constraints may limit adoption. Future work will examine the usability of the chord-distance mappings through performer workshops and will explore additional timbral controls, including hand rotation, curvature, motion trajectory, and performer-adjustable FM parameters.

5 Ethical Standards

This research was self-funded with no external funding sources. The iXeRemin application is publicly available on the Apple App Store for research and artistic use. No conflicts of interest exist. No human subjects or animals were involved in this research.

References

[1] David Johnson, Daniela Damian, and George Tzanetakis. 2020. Evaluating the effectiveness of mixed reality music instrument learning with the theremin. *Virtual Reality* 24 (06 2020). https://doi.org/10.1007/s10055-019-00388-8

[2] David Johnson and George Tzanetakis. 2017. VRMin: Using Mixed Reality to Augment the Theremin for Musical Tutoring. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Aalborg University Copenhagen, Copenhagen, Denmark, 151–156. https://doi.org/10.5281/zenodo.1176205

[3] Gwangyu Lee. 2026. *iXeRemin: Play FM Theremin in XR Space*. Retrieved February 4, 2026 from https://apps.apple.com/app/ixeremin/id6746877322

[4] Gwangyu Lee. 2026. *iXeRemin: Source Code Repository*. Retrieved February 4, 2026 from https://github.com/gwangyu-lee/iXeRemin

[5] Jingyi Li and Per Ola Kristensson. 2025. On the Benefits of Sensorimotor Regularities as Design Constraints for Superpower Interactions in Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 2568–2578. https://doi.org/10.1109/TVCG.2025.3549876

[6] Giovanni Santini. 2020. Augmented Piano in Augmented Reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 411–415. https://doi.org/10.5281/zenodo.4813449