# Screenless Optical Theremin with Tremolo (ScOTT)

MICHAEL GANCZ, Yale University, USA
JUSTIN BERRY, Yale University, USA
SHU WEI, Yale University, USA
JAKE SHAKER, Yale University, USA
KIMBERLY HIEFTJE, Yale University, USA
ASHER MARKS, Yale University, USA

Additional Key Words and Phrases: gesture, MIDI, extended reality, XR, head-mounted display, collaborative music-making, human-computer interaction, musical mapping

## 1 Program Notes

"Programmability is a curse." -Perry Cook, 2001 [2]

Music and gesture are *entangled*. We describe music in terms of gesture; likewise, we must engage with gesture in order to produce musical utterance. Music draws gesture from within us, often without our conscious intervention. Likewise, gestural decisions affect the character of musical performance.

When we map gesture to music, we fragment a space of infinite variety into operative metaphors: semantic, embodied, spatial, tactile, synesthetic. We classify the space of gestural possibility in terms of dialectical or combinatorial oppositions: ergonomic or effortful, strong or light, pressing or slashing [1, 3]. Regardless, gesture remains ambiguous, endlessly programmable, despite the language that we construct around it.

The present performance does not attempt to challenge existing schemes for understanding music and gesture. Instead, it explores the intersection between these two infinite spaces. A screenless XR headset that tracks movement, but does not render it visually, allows our performance to operate both in the virtual world—where gesture manifests as vectors and variables—and the physical world, where gesture manifests as a conversation with the saxophone, the guitar, and the audience.

## 2 Project Description

### 2.1 Abstract

Head-mounted extended-reality (XR) interfaces provide a flexible platform for immersive and embodied musical instrument design. By combining spatial audio, ergonomic first-person gestural control, and networked interactivity, these interfaces can facilitate expressive, interesting, and emotionally resonant performances. Unlike data gloves and hyperinstruments, XR headsets can be calibrated to the physical properties of their individual users. However, these headsets tend to obscure the eyes and other parts of the face, limiting the user's capacity to establish eye contact and transmit facial expressions. These subtle communicative elements play a crucial role in real-world collaborative musical settings where performers utilize facial cues to negotiate surface parameters such as timing, dynamics, and breath, as well as more complex qualities like atmosphere and interpretive mimesis.

In this remote performance, we present the **Sc**reenless **O**ptical **T**heremin with **T**remolo (ScOTT), a novel gestural MIDI controller powered by a modified screenless XR headset. ScOTT focuses on the hands and arms, mapping broad gestures to coarse-grained musical parameters (e.g. pitch and velocity) and small movements to more complex musical ornaments (e.g. tremolo width and frequency). In a structured improvisation that highlights the role of ornament in musical texture, we explore the ScOTT's capacity to balance social presence, embodied interaction, and expressivity.

Authors' Contact Information: Michael Gancz, Yale University, New Haven, CT, USA; Justin Berry, Yale University, New Haven, CT, USA; Shu Wei, Yale University, New Haven, CT, USA; Jake Shaker, Yale University, New Haven, CT, USA; Kimberly Hieftje, Yale University, New Haven, CT, USA; Asher Marks, Yale University, New Haven, CT, USA.


*NIME '25, June 24–27, 2025, Canberra, Australia*

Fig. 1. A collage of hand gestures used to control the ScOTT

### 2.2 Hardware and software specifications

We run ScOTT on Unity 2022.3.23f1 using a Windows 11 Home (23H2) computer (Intel i7-13700KF, NVIDIA GeForce RTX 4080, 32 GB RAM), with a Quest Link wire connection to a modified screen-free Meta Quest 3. We employ RtMidi for Unity[1] and loopMIDI[2] to route ScOTT's MIDI output into Reason 10[3] or Ableton Live 12[4].

### 2.3 Gestural mapping scheme

ScOTT tracks the following relationships between parts of the performer's body:

- distance between the chest and left wrist (left protrusion; $pl$)
- distance between the chest and right wrist (right protrusion; $pr$)
- distance between the left middle fingertip and wrist (left closure; $cl$)
- distance between the right middle fingertip and wrist (right closure; $cr$)
- speed of the center point between the right middle fingertip and wrist (right speed; $sr$)

In either the Unity inspector or the graphical user interface, performers may select pitch-classes in standard tuning (from 0 to 11) and minimum and maximum MIDI pitch values (from 0 to 127) in order to define an ordered list $P$ of possible pitches that ScOTT can trigger. They may also select minimum and maximum velocity, tremolo frequency, and tremolo width values, as well as map parameters $a_p$ and $a_v$, velocity modifier $m$, and closure threshold $t$. The magnitude of map parameters $a_p$ and $a_v$ represents how 'exponentially' the available pitch and velocity values, respectively, are distributed with respect to their relevant tracked distance. For instance, a very high $a_p$ would cluster many low notes near the performer's chest, leaving high notes sparsely distributed as they approach arm's reach. A low $a_p$ would distribute pitches relatively evenly in space.[5] These live-tracked and preselected values trigger MIDI notes in accordance with the following transformations:

$$\mathrm{pl}_{norm} = \mathrm{clamp}(\mathrm{norm}(\mathrm{pl}, \mathrm{pl}_{min}, \mathrm{pl}_{max}), 0, 1) \quad \mathrm{pr}_{norm} = \mathrm{clamp}(\mathrm{norm}(\mathrm{pr}, \mathrm{pr}_{min}, \mathrm{pr}_{max}), 0, 1) \tag{1}$$

[1] https://github.com/keijiro/jp.keijiro.rtmidi
[2] https://www.tobias-erichsen.de/software/loopmidi.html
[3] https://www.reasonstudios.com/reason
[4] https://www.ableton.com/en/live/
[5] $a_p$ and $a_v$ are meaningful in the exclusive range $(1, \infty)$

Fig. 2. The modified screen-free Quest 3 headset used to play ScOTT.

$$\mathrm{cl}_{norm} = \mathrm{clamp}(\mathrm{norm}(\mathrm{cl}, \mathrm{cl}_{min}, \mathrm{cl}_{max}), 0, 1) \quad \mathrm{cr}_{norm} = \mathrm{clamp}(\mathrm{norm}(\mathrm{cr}, \mathrm{cr}_{min}, \mathrm{cr}_{max}), 0, 1) \tag{2}$$

$$\mathrm{pitch}_{norm} = \mathrm{logmap}_{a_p}(\mathrm{pr}_{norm}) \quad \mathrm{pitch}_{idx} = \lfloor \mathrm{scale}(\mathrm{pitch}_{norm}, 0, \|P\|) \rfloor \quad \mathrm{pitch} = P|_{\mathrm{pitch}_{idx}} \tag{3}$$

$$\mathrm{vel}_{norm} = \mathrm{logmap}_{a_v}(\mathrm{clamp}(\mathrm{pl}_{norm} + m\mathrm{sr}_{norm}, 0, 1)) \quad \mathrm{vel} = \mathrm{scale}(\mathrm{vel}_{norm}, \mathrm{vel}_{min}, \mathrm{vel}_{max}) \tag{4}$$

$$\mathrm{tremfreq}_{norm} = \begin{cases} 0 & \mathrm{cr}_{norm} \leq \mathrm{t} \\ \mathrm{cr}_{norm} & \mathrm{cr}_{norm} > \mathrm{t} \end{cases} \quad \mathrm{tremfreq} = \mathrm{scale}(\mathrm{tremfreq}_{norm}, \mathrm{tremfreq}_{min}, \mathrm{tremfreq}_{max}) \tag{5}$$

$$\mathrm{tremwidth} = \lfloor \mathrm{scale}(\mathrm{cl}_{norm}, \mathrm{tremwidth}_{min}, \mathrm{tremwidth}_{max}) \rfloor \tag{6}$$

Here *vel* indicates velocity, *tremfreq* and *tremwidth* indicate tremolo frequency in Hz and width in semitones, respectively, and the *logmap*, *scale*, *norm*, and *clamp* functions are defined as follows:

$$\mathrm{logmap}_a(x) = 1 + \log_a\left((1 - \frac{1}{a})x + \frac{1}{a}\right) \quad \mathrm{scale}(x, a, b) = (a - b)x + a \tag{7}$$

$$\mathrm{norm}(x, a, b) = \frac{x - a}{b - a} \quad \mathrm{clamp}(x, a, b) = \max(a, \min(x, b)) \tag{8}$$

ScOTT also has an "invert pitch map" button, which rewrites the *logmap* function as follows:

$$\mathrm{invlogmap}_a(x) = -\log_a\left((1 - \frac{1}{a})x + \frac{1}{a}\right) \tag{9}$$

## 3 Technical Notes

We will present a prerecorded 10-minute composition for ScOTT, saxophone, and guitar, alongside a score-follower animation. If the conference schedule contains a dedicated streaming block for remote performances, we prefer to be included in this block. Otherwise, we request our performance to be made available either on a dedicated screen or via a link embedded into conference materials.

## 4 Media Links

In our supplementary materials we include a demo video showcasing our initial workshop for ScOTT, clarinet, and guitar, as well as the graphical score composed by Michael Gancz for the official performance. Our performance video will be presented as part of conference proceedings, and we will make a stable link to this performance available post-fact.

## 5 Ethical Standards

This project is partially funded by Reality Labs Research, and the headset we use is a Quest 3 that was modified and donated by Reality Labs Research.

## Acknowledgments

The authors would like to thank Abigail Crocker for her clarinet performance in our demo recording, and Wes Lewis for his saxophone performance in our performance recording.
This work was supported by Reality Labs Research and the Yale Center for Immersive Technology in Pediatrics.

## References

[1] Peter Bennett, Nicholas Ward, Sile O'Modhrain, and Pedro Rebelo. 2007. DAMPER: a platform for effortful interface development. In *Proceedings of the 7th international conference on New interfaces for musical expression*. 273–276.
[2] Perry Cook. 2017. 2001: Principles for designing computer music controllers. *A NIME Reader: Fifteen years of new interfaces for musical expression* (2017), 1–13.
[3] Nicholas Ward, Kedzie Penfield, Sile O'Modhrain, and R Benjamin Knapp. 2008. A Study of Two Thereminists: Towards Movement Informed Instrument Design.. In *NIME*. 117–121.