The Moving Drone: Negotiating Agency Between the Voice and the Virtual

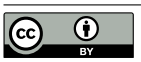
NITHYA SHIKARPUR, Massachusettes Institute of Technology, USA
VICTOR ARUL, Harvard University, USA
ANNA HUANG, Massachusettes Institute of Technology, USA

Additional Key Words and Phrases: Practice-based research; Hindustani music; Musical agency; New musical instruments; Human-AI interaction

Fig. 1. A still from the premiere of 'The Moving Drone' at Harvard University, featured in the Hydra concert series.[1]

[1]The speaker set up in this figure is different, please refer figure 3 for the NIME performance proposed set up.

Authors' Contact Information: Nithya Shikarpur, Massachusettes Institute of Technology, Cambridge, USA; Victor Arul, Harvard University, Cambridge, USA; Anna Huang, Massachusettes Institute of Technology, Cambridge, USA.



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

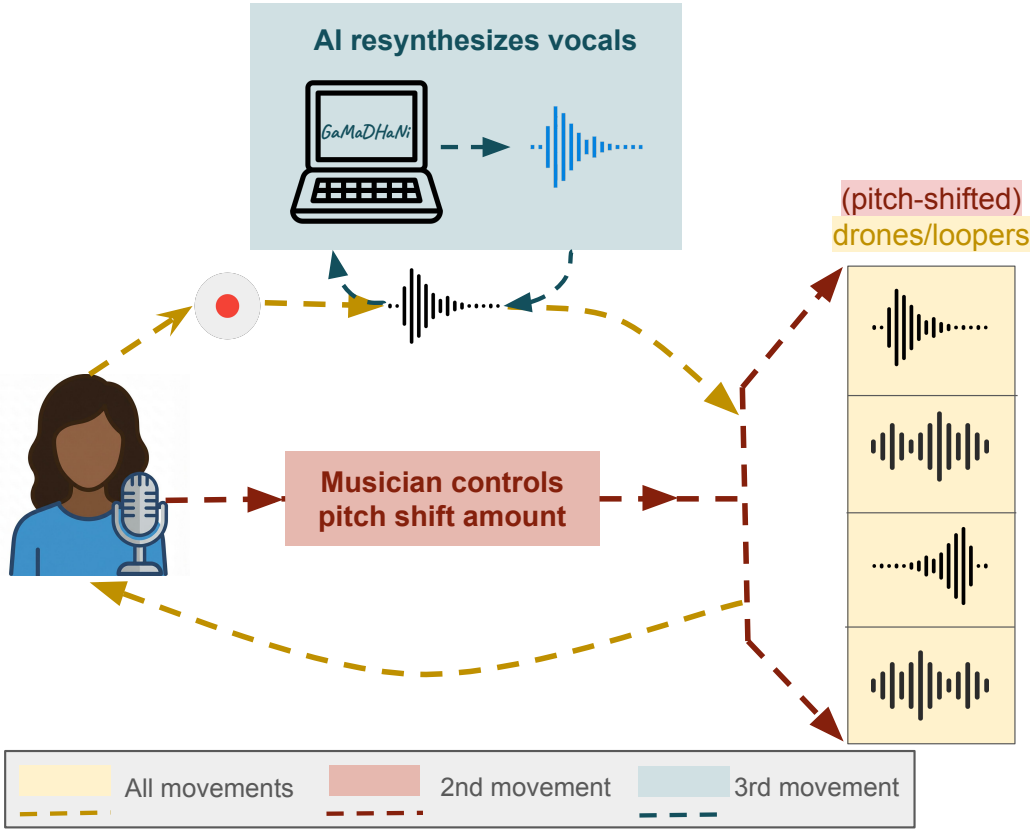


Fig. 2. High-level system design for the performance across three movements.

1 Program Notes

Melodic material in Hindustani music is presented in relation to a tonic, usually sustained by the *tanpura*, a four-stringed drone instrument. Rooted in Hindustani music, 'The Moving Drone' sets the traditionally static drone into motion that, throughout the performance, gains increasing agency transitioning from reactive to more proactive roles. The work employs four independent loopers in Max/MSP to function as 'virtual' drones. They are populated cyclically in real-time as the vocalist improvises, creating an organic and evolving feedback loop between the voice and the virtual drone. This relationship further evolves melodically by pitch shifting the loops, which introduces a dimension of sudden, explicit movement. Then it changes timbrally, via the integration of GaMaDHaNi, a singer conditioned pitch-to-voice generative AI model to resynthesize looped audio. While current music AI approaches prioritize high-fidelity and realism of generated content which has sparked anxiety over job replacement for the music community, this work intentionally utilizes low-fidelity generative outputs, further necessitating human interpretation and situational context in order to be complete. 'The Moving Drone' positions technology and generative AI within established socio-cultural musical practices, proposing a virtual drone as an active, responsive, and co-creative musical agent.

2 Project Description

Hindustani music, i.e. North Indian Classical music, is a modal tradition where melodic meaning of each note is derived from it's relationship with a fixed tonic [2]. This relationship is externalized by the *tanpura*, a four-stringed drone that plays the tonic in addition to other fixed notes to lay the steady melodic foundation upon which the performer improvises. Through this performance, we explore giving increased agency to this conventionally passive drone and probe the negotiation between vocal improvisation and drone.

In this work, the drone is reimagined as a virtual entity realized through four independent loopers that are cyclically populated by the performer's live improvisations, creating a real-time feedback loop. We conceptualize the drone's agency along two axes, pitch and timbre, each ranging from reactive (low agency) to proactive (high agency) [21] [2].

[2]This is a work in progress, and we acknowledge that the full scale of the theoretical framework has not been utilized yet, but we hope to use this design space as a guide for future work.

Raga	Pitch shift amount	A#	B	C	C#	D	D#	E	F	F#	G	G#	A
Bihag	0	**S**		R		**G**	**m**	M	**P**		D		**N**
Bhairavi	-1	r		**g**		m		P	**d**		n		**S**
Basant	0	**S**	r			G	m	**M**	P	d			N
Kafi	+2	d		**n**		**S**		R	**g**		m		**P**

Table 1. Table showing the different ragas used during Movement 2. Notes are depicted using solfege notation in [8] with A# as the tonic. Notes highlighted in red are used as 'portals' to jump into and out of each raga as the pitch shifting of loopers occurs.

Each looper, implemented in Max/MSP [1], has a fixed buffer length of three seconds, and individual controls on the amount of crossfade to prevent clicks as the loop repeats. The performer controls what is recorded into loops by means of a manually controlled record button which cyclically populates the loops. The high-level system layout is depicted in figure 2. The performance consists of three movements: (1) organically evolving loops, (2) jumping loops, and (3) AI-resynthesized loops, each exploring different aspects of melody and timbre elaborated below.

2.1 Movement 1: Organically Evolving Loops

The performance starts with the loops playing the tonic (*S*) and the dominant (*P*), similar to a traditional *tanpura*. The vocalist improvises in raga *Bihag*, a raga replete with feelings of love and happiness [3, 6]. Structurally, the emphasis on the notes *S, G, P, N* (analogous to a Major 7th chord) provides a stable and fairly consonant harmonic landscape for the drone. As the vocalist sings, specific tones are recorded into the loops to create a symbiotic, ever-evolving virtual drone. In this movement, the drone's agency remains limited, functioning reactively across both melodic and timbral dimensions.

2.2 Movement 2: Jumping Loops

This movement introduces a more explicit, 'forced' movement along the melodic axis, transitioning the drone into a relatively more proactive role. While the movement begins with the standard tonic (*S*) and fifth (*P*), the system no longer simply records and loops the vocalist. Instead, movement is shown by pitch-shifting loops across four pre-determined presets, each corresponding to a distinct raga: *Bihag*, *Bhairavi*, *Basant*, and *Kafi*. These presets are designed so that most absolute pitches remain consistent across the four ragas, similar to ideas in melodic mode switching or *grihabedham* [2] as seen in table 1. Specific notes function as 'portals' or pivots that allow the performer to transition between these varied tonal landscapes and the raga characteristic phrases pose as anchors to help anchor oneself in the constantly changing raga space. In this phase, the system actively shapes the vocal improvisation rather than merely reflecting it.

2.3 Movement 3: AI resynthesized loops

In the final movement, the drone achieves timbral agency, mediated by GaMaDHaNi [17], a hierarchical generative model trained on Hindustani vocal music. The model employs two stages: a pitch generator that defines the 'melodic idea' and a spectrogram generator conditioned on a singer ID that realizes the melodic idea into sound. This is implemented by routing audio from Max/MSP to a dedicated GPU-powered laptop for real-time synthesis. Using GaMaDHaNi's spectrogram generator, the input is transformed into another timbre before being fed into the loopers, with each looper assigned a pre-determined singer ID.

Artistically, we choose to create a sense of chaos and tension in this section by leaning into the noisy and distorted quality of generated samples from the model, in part due to the 16 kHz sample rate and also the Griffin-Lim algorithm [9] used to estimate phase. Additionally, raga *Shree* chosen for improvisation, lends chromatic clusters that add to the harmonic dissonance in the drone.

3 Performer's Reflection: The Process of Negotiation

The aesthetic impetus for this work was the transformation of the traditionally static drone into a dynamic, moving entity. Negotiating agency with this virtual drone necessitated a process of cognitive and technical adaptation, which I have distilled here to contextualize the design choices detailed in section 2.

First, the lack of a constant tonic from the tanpura, an element I had taken for granted, initially resulted in significant musical disorientation. Through practice, I developed an improvisation style that periodically revisited the tonic, ensuring that it was present in at least one of the four loops at all times. I also found myself relying on raga as a cognitive framework to find my footing in this new, changing sonic environment. Raga-characteristic phrases played the role of perceptual

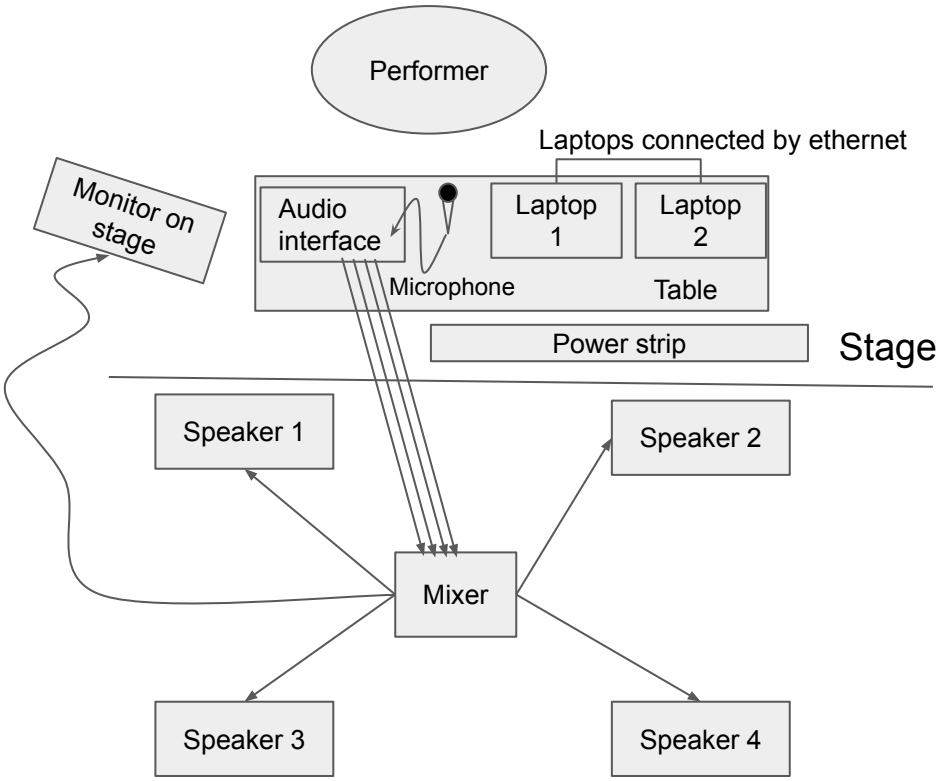


Fig. 3. Technical Floor Plan

anchors for me, while the consonance or dissonance of notes in the raga guided the emotional intention for each movement. Furthermore, I recognized that populating the drone exclusively with my live vocal input constrained the system's spectral width to the physical limits of my own vocal range. As a result, I adopted layers of octave-shifted loops to make the sound more full. The artificial voice felt uncanny marked by a quality that was both not-me and distinctly non-human. This tense and chaotic aesthetic was further amplified by the chromatic note clusters and austere nature of raga *Shree*.

Exploring this interaction with the drone took me through a process of learning, understanding and unlearning that I hope to keep exploring further.

4 Reflections: AI and Music

The rapid advancement of text-to-music generative models has sparked strong backlash from music communities [12, 15], exposing divergent motivation across, artistic, academic, and commercial entities [7, 14]. Previous work [4] highlights the importance of situating AI application for music within cultural and artistic practices. However, the prevalence of standardized benchmarks [5, 13] often narrows the scope of evaluation, thus marginalizing the experimental, niche and culturally-informed approaches to music making and generation. Furthermore, Eurocentric bias-spanning datasets, model architectures, and task definitions continue to be a point of contention for artists from diverse musical backgrounds [16–19].

This work adopts an intentional stance of grounding AI within the first author's practice of Hindustani music. By challenging the metrics prioritized by current evaluation benchmarks, such as high-fidelity and realism, we advocate for a more critical adoption of AI in music. We posit that music does not exist in a vacuum; it is a practice forged through centuries of culture and community. Acknowledging this lineage is essential when developing new musical technologies.

5 Technical Notes

This performance is a solo improvised set of around 13 minutes duration. Two laptops will be used, one to run Max/MSP and another laptop to run the model. The piece has a 4-channel output that can be reduced to stereo if required. See figure 3 for a technical floor plan. We list the equipment required below:

- Equipment provided by artists:
  - Audio interface

  - 2 Laptops
  - Ethernet Cable
- Equipment requested from organizers:
  - Microphone along with a stand
  - XLR for microphone
  - 4 speakers with cables, each connected to 4 channel output from laptop (provided through Dante Virtual Soundcard)
  - 1 Monitor on stage for the performer
  - Table to place laptops during performance
  - Power strip close to the table to charge laptops during performance

6 Media Links

Media link: https://youtu.be/3dJOzoxGx_c

7 Ethical Standards

This is a solo performance and thus was not subject to the institute's ethical board review. This performance involves generative models trained on open source datasets [10, 11, 20] comprising of around 120 hours of data. Although appropriate permissions were obtained from stakeholders, artists and labels, at the time of data collection, we acknowledge that the scope of generative modeling possibly fell outside of the original intent of researchers and artists. This discrepancy highlights the ongoing need for renewed ethical and licensing standards within the context of AI-driven musical co-creation. Additionally, we acknowledge the costly environmental impact of AI models and as a result strive to maintain small models that support both fast inference and also consume less power.

Acknowledgments

Nithya would like to thank Hans Tutschku for his invaluable guidance and feedback throughout the course of developing this performance. She also expresses deep gratitude to Weilu Ge and Manaswi Mishra who, through conversations, helped fuel and spark ideas for the performance. Most importantly she is grateful for all her music teachers who gave her the vocabulary and insight to develop these ideas.

References

[1] Cycling '74. 1997. Max. https://cycling74.com/products/max. [Accessed 12-02-2026].
[2] Sandeep Bagchee. 1998. *Nad.* BPI Publishing.
[3] Sabyasachi (Rahul) Bhattacharya. 2014. Raga Bihag: An uplifting melody. https://sarod.com.au/raga-bihag-an-uplifting-melody/. [Accessed 12-02-2026].
[4] Giovanni Bindi, Nils Demerlé, Rodrigo Diaz, David Genova, Aliénor Golvet, Ben Hayes, Jiawen Huang, Lele Liu, Vincent Martos, Sarah Nabi, et al. 2023. AI (r) evolution–where are we heading? Thoughts about the future of music and sound technologies in the era of deep learning. *arXiv preprint arXiv:2310.18320* (2023).
[5] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2023. Simple and Controllable Music Generation. In *Thirty-seventh Conference on Neural Information Processing Systems*.
[6] Ashwini Bhide Deshpande. 2023. Raag Ki Tasveer - Bihag Family | Batiyan Daurawat | Dr. Ashwini Bhide Deshpande. https://www.youtube.com/watch?v=BptPdYSxRGs&list=PLEfmeUsduoEadg1CroTkD1pghArNZ7HkT&index=2. Accessed 04-30-2026.
[7] Eamonn Forde. 2025. Musicians are deeply concerned about AI. So why are the major labels embracing it? https://www.theguardian.com/music/2025/dec/16/musicians-are-deeply-concerned-about-ai-so-why-are-the-major-labels-embracing-it. [Accessed 12-02-2026].
[8] Kaustuv Kanti Ganguli and Preeti Rao. 2018. On the Distributional Representation of Ragas: Experiments with Allied Raga Pairs. *Trans. Int. Soc. Music. Inf. Retr.* 1, 1 (2018), 79–95.
[9] Daniel Griffin and Jae Lim. 1984. Signal estimation from modified short-time Fourier transform. *IEEE Transactions on acoustics, speech, and signal processing* 32, 2 (1984), 236–243.
[10] Sankalp Gulati, Joan Serra, Vignesh Ishwar, Sertan Sentürk, and Xavier Serra. 2016. Phrase-based rāga recognition using vector space modeling. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 66–70.
[11] Sankalp Gulati, Joan Serrà Julià, Kaustuv Kanti Ganguli, Sertan Sentürk, and Xavier Serra. 2016. Time-delayed melody surfaces for rāga recognition. In *ISMIR*.
[12] Jonathan Herington, Raffaella Borasi, Benjamin J Guerrero, David E Miller, Blaire Koerner, Yu Jung Han, Zenon Borys, and Rachel Roberts. 2025. Musicians' ethical concerns about AI: an interview study. *AI & SOCIETY* (2025), 1–14.
[13] Kevin Kilgour, Mauricio Zuluaga, Dominik Roblek, and Matthew Sharifi. 2018. Fr\'echet audio distance: A metric for evaluating music enhancement algorithms. *arXiv preprint arXiv:1812.08466* (2018).
[14] Andrew McPherson, Fabio Morreale, and Jacob Harrison. 2019. Musical instruments for novices: Comparing NIME, HCI and crowdfunding approaches. In *New directions in music and human-computer interaction*. Springer, 179–212.
[15] Adam Neely. 2026. Suno, AI Music, and the Bad Future. https://www.youtube.com/watch?v=U8dcFhF0Dlk. [Accessed 12-02-2026].
[16] Michele Newman, Lidia Morris, and Jin Ha Lee. 2023. Human-AI Music Creation: Understanding the Perceptions and Experiences of Music Creators for Ethical and Productive Collaboration.. In *ISMIR*. 80–88.

[17] Nithya Shikarpur, Krishna Maneesha Dendukuri, Yusong Wu, Antoine Caillon, and Cheng-Zhi Anna Huang. 2024. Hierarchical generative modeling of melodic vocal contours in hindustani classical music. In *ISMIR*.
[18] Nithya Shikarpur and Cheng-Zhi Anna Huang. 2024. Exploratory Study Of Human-AI Interaction For Hindustani Music. *NeurIPS Creative AI Track* (2024).
[19] Ahmet Solak, Florian Grötschla, Luca A Lanzendörfer, and Roger Wattenhofer. 2025. Bias beyond Borders: Global Inequalities in AI-Generated Music. *arXiv preprint arXiv:2510.01963* (2025).
[20] Ajay Srinivasamurthy, Sankalp Gulati, Rafael Caro Repetto, and Xavier Serra. 2021. Saraga: Open datasets for research on indian art music. *Empirical Musicology Review 2021; 16 (1): 85-98.* (2021).
[21] Notto JW Thelle and Philippe Pasquier. 2021. Spire muse: A virtual musical partner for creative brainstorming. In *NIME 2021*. PubPub.