Drifting in Currents of Currents: A Hydro-Acoustic Interface for Gesture-Driven Neural Sound Synthesis and Entropic Memory

Tak-Cheung Hui
tachui@hkmu.edu.hk
Department of Creative Arts, Hong Kong Metropolitan University
Ho Man Tin, Hong Kong

Xiaoqiao Li
xili@hkmu.edu.hk
Department of Creative Arts, Hong Kong Metropolitan University
Ho Man Tin, Hong Kong

Chun-ting Chan
vctchan@hkmu.edu.hk
Department of Creative Arts, Hong Kong Metropolitan University
Ho Man Tin, Hong Kong

Cheuk-kit Chung
chkchung@hkmu.edu.hk
Department of Creative Arts, Hong Kong Metropolitan University
Ho Man Tin, Hong Kong

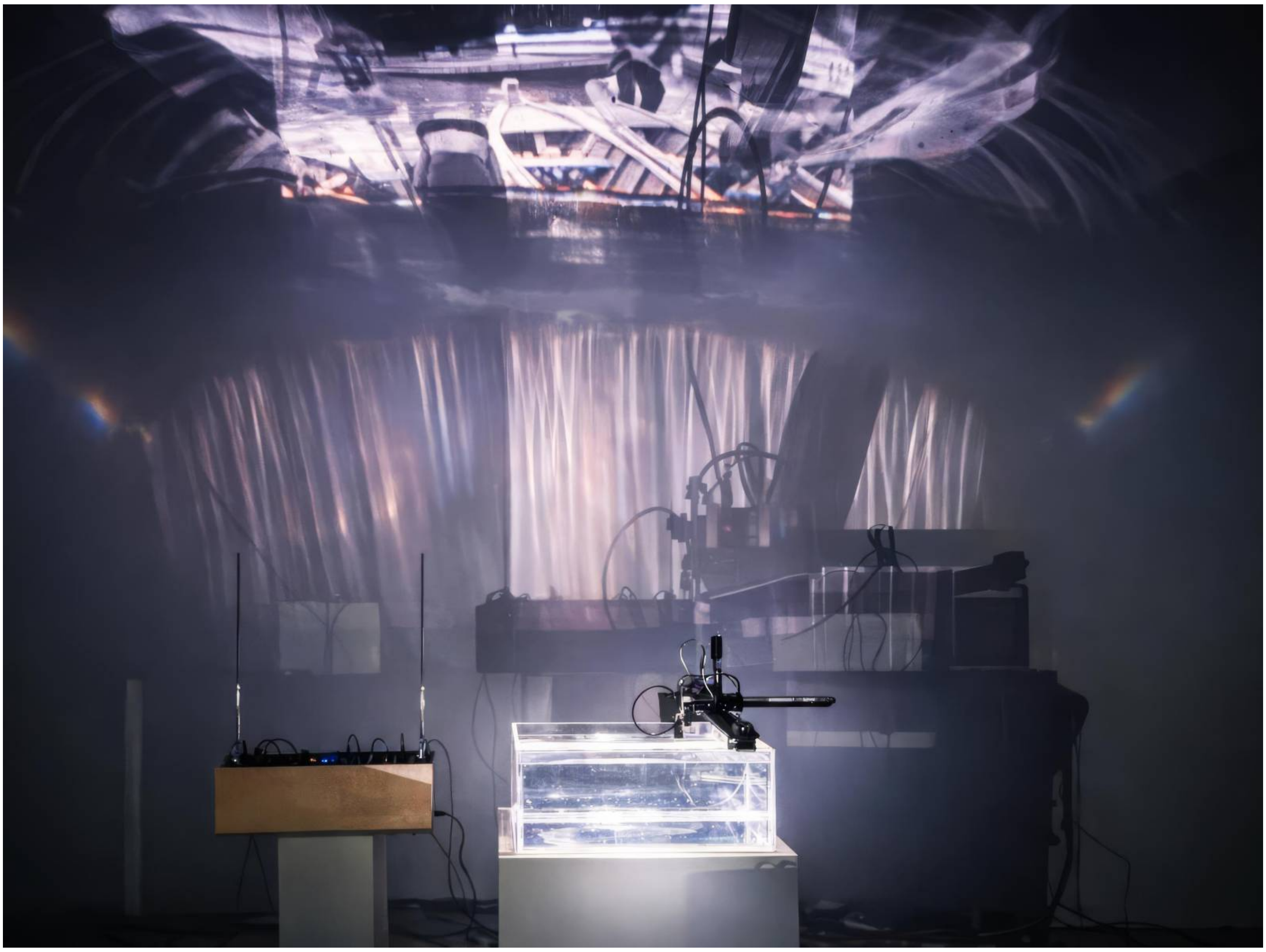

Figure 1: Installation overview showing the water tank, overhead XY plotter, theremin interfaces, and caustic projection.

Abstract

Drifting in Currents of Currents is an interactive hydro-acoustic system that explores the displacement of Hong Kong's Tanka (boat people) through gesture, fluid dynamics, and neural sound synthesis. A solo interactor controls two theremin interfaces to drive a motorized XY plotter that drags a hydrophone through a water tank. The resulting water disturbance functions simultaneously as sound material and excitation signal for a neural audio

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

model (RAVE) trained on coastal soundwalks recorded in Tai O and Aberdeen. Rather than reproducing historical vocal materials, textual fragments from documented Saltwater Songs operate as symbolic gating mechanisms within the synthesis pipeline: each fragment triggers a time-limited amplitude envelope that conditions the audibility of neural output without producing vocal sound. Audio is spatialized through multiple transducers attached directly to the tank and pedestal, establishing a physical hydro-acoustic feedback loop. Grounded in theories of cosmotechnics and wet ontology, and informed by perspectives on machine learning as a creative musical tool, the system frames memory as an entropic, materially mediated process—one that resists retrieval and remains perceptible only through instability, latency, and drift.

Keywords

Sound Art, Kinetic Installation, Neural Audio Synthesis, Cultural Heritage, Hydro-informatics, Tangible Interface, Fluid Dynamics

1 Introduction: Drift, Displacement, and the Problem of Retrieval

The Tanka people constitute a historically amphibious culture within the Pearl River Delta whose livelihood, social organization, and cosmology were inseparable from water. Living on boats along the Hong Kong coastline, Tanka communities developed maritime practices and vocal traditions, commonly known as Saltwater Songs (Ham Shui Go), that emerged from labor, navigation, and tidal rhythms, grounded in the rhythms of daily maritime life and communal experience.

During the twentieth century, rapid urbanization and land reclamation displaced Tanka communities, dissolving the environmental conditions that sustained their way of life. Although Saltwater Songs have been documented through transcription and archival initiatives [3], such efforts inevitably transform fluid, situated practice into static cultural artifacts. Contemporary approaches to cultural preservation that rely on artificial intelligence risk further abstraction by converting historically contingent sound practices into datasets optimized for reproduction, retrieval, and stylistic generalization. This system avoids this risk by training the neural model exclusively on environmental field recordings (Section 3.3).

Prior work on sensor-based and tangible musical interfaces [9, 13] has explored hydro-acoustic instruments emphasizing surface interaction or tactile feedback [7]. More recently, RAVE-based instruments have explored tangible latent space navigation through magnetic interfaces [10] and low-latency architectural redesign for instrumental interaction [2]. In contrast, the present work integrates fluid dynamics directly into the excitation layer of neural audio synthesis, using water turbulence as the primary input signal to the RAVE model. Instead of granting access to historical sound, *Drifting in Currents of Currents* deliberately withholds it. By embedding water, gesture, and machine learning within a single feedback system, the installation stages memory as unstable, delayed, and materially resistant. An overview of the physical installation is shown in Figure 1. Audience gestures do not activate historical recordings; they disturb a hydro-acoustic environment whose responses are only indirectly conditioned by historical traces.

The contribution of this work lies in three interconnected aspects. It introduces a hydro-acoustic interaction model in which water operates simultaneously as excitation source, resonator, and feedback medium for neural synthesis. It proposes an ethically constrained approach to historical material that relies on symbolic gating without sonic reproduction. Finally, it demonstrates a gesture-mapping strategy that enforces restraint as an interactional value, reframing masking, latency, and irreversibility as expressive resources.

2 Theoretical Framework: Cosmotechnics and Wet Ontology

This work is informed by Yuk Hui's concept of *cosmotechnics*, which challenges the assumption that technology operates as a culturally neutral and universally applicable apparatus. Hui argues that technical systems are always embedded within specific cosmologies—configurations of humans, environments, and knowledge—and that modern technological universalism often erases local technical traditions [5]. Tanka maritime practices constituted a water-based cosmotechnics whose logic was fundamentally incompatible with land-oriented urban modernity.

Drifting in Currents of Currents foregrounds the absence of this cosmotechnics, deliberately refusing to reconstruct it through representation or simulation. The system juxtaposes mechanical precision and neural synthesis with the instability of water, reflecting the displacement of a water-based cosmology by computational regimes that privilege stability and abstraction.

The project also draws on Astrida Neimanis's theory of *wet ontology*, which emphasizes water as a medium that resists fixity, transparency, and linear temporality [8]. Water retains traces only through continuous transformation and cannot be stably archived. By inserting water directly into the signal chain as a sound source, optical refractor, and vibrational body, the system resists the abstraction characteristic of digital media. Taken together, cosmotechnics and wet ontology frame the installation as a site where technical mediation and material instability co-produce memory through ongoing transformation.

3 System Overview: Hydro-Acoustic Organology as Feedback

The system functions as a hybrid organology [12], linking mechanical actuation, fluid dynamics, and neural synthesis into a recursive structure (Figure 2). Drawing on distributed systems frameworks for digital musical instruments [6], the installation integrates sensing, actuation, and synthesis across physical and computational layers. In this system, hydro-acoustic turbulence operates as a dynamic modulator of neural latent space, linking nonlinear material behavior to generative structure.

3.1 Physical Configuration

At the center of the system is a transparent acrylic tank filled with saline water, functioning simultaneously as an acoustic medium, a vibrational body, and an optical refractor. Its transparency allows projected imagery to pass through the water and undergo continuous distortion. Mounted above the tank, a motorized XY plotter using a CoreXY architecture controls the horizontal movement of a submerged hydrophone (Figure 3). As the plotter drags the hydrophone through water, it generates turbulence, cavitation, and resonant noise that constitute the system's primary sonic material.

Interaction is mediated through two theremin interfaces (Figure 4). The left theremin controls the X-axis, corresponding to historical distance, while the right theremin controls the Y-axis, representing environmental turbulence. Interaction is designed

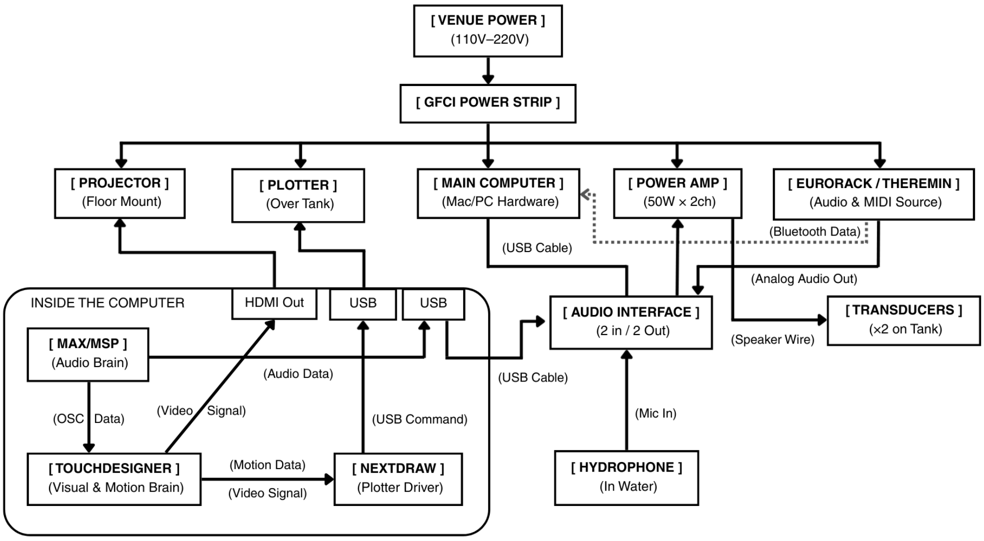

Figure 2: System signal flow diagram. A detailed architecture showing the integration of hardware (sensors, actuators, transducers) and software (Max/MSP, TouchDesigner, RAVE). The diagram illustrates the closed-loop topology where hydro-acoustic input excites the neural synthesis engine, which feeds back into the water tank via transducers.

for a single active interactor, with others observing and listening. The theremin was selected for its non-contact sensing, which enforces a physical gap between interactor and instrument. Unlike wearable or touch-based interfaces, influence remains indirect and without tactile confirmation. This characteristic reinforces the idea that historical memory cannot be grasped directly but only influenced at a distance.

Figure 3: Hydro-acoustic actuation. A motorized XY plotter drags a hydrophone through the water tank, while transducers mounted on the tank walls establish a physical feedback loop.

3.2 Audio Architecture, Feedback, and Spatial Listening

Sound in the system emerges from three concurrent sources. A direct hydrophone signal captures real-time water noise generated by mechanical motion. A modular synthesizer provides low-latency gestural feedback. In parallel, a neural synthesis layer generates audio using a RAVE model trained on approximately thirty hours of coastal soundwalk recordings from Tai O and Aberdeen [1].

The modular synthesizer responds immediately to changes in theremin capacitance, whereas the neural model introduces a perceptible inference latency of approximately 90-250 milliseconds. In practice, interactors perceive these as two temporally distinct layers. This latency is preserved as a compositional element, producing a layered texture in which gesture and its neural response coexist without synchronizing.

Sound is spatialized through multiple transducers attached directly to the tank walls and pedestal. Since vibrational energy propagates through the acrylic structure and water before radiating into the air, listeners perceive the tank itself as the acoustic source. The distinction between sound source and resonator collapses, encouraging listening oriented toward material presence over spatial localization.

3.3 Neural Synthesis and Ethical Abstraction

Neural synthesis operates as a machine listening layer whose behavior emerges from unstable excitation. Unlike conventional digital signal processing techniques that preserve deterministic input-output relationships, the RAVE model introduces a latent representation trained on environmental sonic textures associated with Tanka displacement. When excited by hydro-acoustic turbulence, the model produces spectrally drifting and temporally smeared responses that are non-repeatable and delayed. These properties are necessary, ensuring that sonic output remains contingent and irreducible to gestural intent.

The neural model is deliberately not trained on vocal recordings in order to avoid reproducing or simulating Tanka voices. The dataset captures the broader sonic ecology of displacement, including diesel engines, water striking stilts in Tai O, and distant construction drones from nearby reclamation sites. When excited by the hydrophone, the model resynthesizes these textures, aligning with perspectives that frame machine learning as a creative musical tool [4].

Saltwater Songs enter the system only as symbolic gating structures. Textual fragments from documented lyrics trigger amplitude envelopes that momentarily open the neural synthesis signal path, regulating temporal audibility without influencing spectral features or model parameters. At moments of heightened turbulence, the neural synthesis signal remains active even as it is perceptually overwhelmed by hydro-acoustic noise.

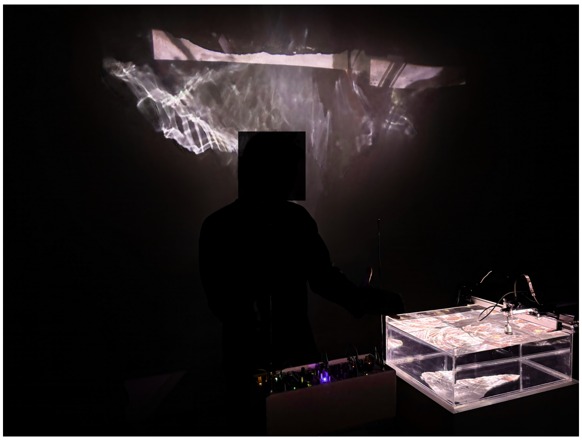

Figure 4: Gesture control interface showing theremin-based non-contact gestures translated into plotter motion, neural-synthesis audibility, and modular-synthesis parameters.

4 Gesture Mapping: Time, Turbulence, and an Ethic of Restraint

Gesture data from the theremin interfaces is mapped onto a two-dimensional control space defined by historical distance (X-axis) and environmental turbulence (Y-axis). The X-axis governs the balance between raw hydro-acoustic sound and neural synthesis, while the Y-axis controls plotter speed and agitation intensity.

This mapping enforces an ethic of restraint. Neural memory becomes audible only when gestures slow and turbulence subsides. Excessive force destabilizes the system and obscures the signal being sought. Because the hydro-acoustic feedback loop continuously alters the tank's resonant state, identical gestures do not yield identical outcomes. The interface, therefore, resists optimization and rewards exploratory listening over instrumental control.

5 Interaction Experience, Settling, and Managed Drift

The installation is designed for exhibition contexts, where one person operates the theremin interfaces; the artistic purpose is to offer an embodied encounter with displacement.

Audience members typically begin with exaggerated gestures that produce dense hydro-acoustic noise. Early interaction often takes the form of searching, characterized by rapid sweeps and large movements that generate broadband cavitation and effectively obscure the neural signal. As interactors pause or reduce motion, the water settles, and the delayed neural response emerges, producing a call-and-response dynamic in which silence becomes a prerequisite for audibility.

In overextended interaction, resonant frequencies reinforce one another through feedback. An automated mechanism attenuates low frequencies when amplitude exceeds a threshold, while continuous modulation displaces emerging resonances. Through this process, sonic history accumulates temporarily before being redistributed, ensuring the system remains in a state of managed drift without resolution.

6 Visual System and Temporal Lag

Textual fragments are also transmitted to a generative visual pipeline implemented using a latent-diffusion-based model [11], integrated into a real-time audiovisual environment via TouchDesigner. Generated imagery is projected downward into the tank, where surface agitation refracts light into unstable caustics.

While the audio system resists retrieval, the visual system attempts, and necessarily fails, to illustrate it. Caustic distortion prevents visual material from stabilizing. Neural inference latency is left uncompensated, aligning visual behavior with the system's broader treatment of memory.

7 Discussion and Conclusion: Material Mediation, Memory, and Entropic Process

Drifting in Currents of Currents resists optimization, transparency, and reproducibility—values often prioritized in interface design. Water introduces nonlinearity, turbulence, and irreversibility that cannot be fully modeled or predicted, positioning the system against control-oriented paradigms and emphasizing situated interaction.

During exhibitions, interactors initially attempted rapid, control-oriented gestures, but aggressive movement masked the neural output. Only when interactors slowed and the water settled, did the RAVE signal emerge, staging a contrast between what Hui terms technological universalism and the materially contingent engagement that the system demands.

Within this framework, machine learning operates strictly as a contingent listener. By exciting the neural model exclusively with hydro-acoustic input, the neural output remains grounded in material conditions. Symbolic gating further reinforces this stance by allowing historical material to condition temporal structure without becoming audible content. In this sense, composition shifts from producing fixed sonic form to designing conditions under which form continually emerges and dissolves.

This work reframes neural sound synthesis as an interface of resistance. By embedding water, feedback, masking, and latency into the system's core logic, the work proposes an alternative model for AI-mediated musical interfaces—one that accepts loss, instability, and delay as constitutive conditions for engaging with displaced cultural memory.

Ethical Standards

This project did not involve human participants or personal data collection. Audience interactions occurred in public performance settings without documentation or evaluation. The work adheres to the NIME Principles and Code of Practice by fostering inclusive, accessible, and non-invasive engagement with cultural heritage.

Acknowledgments

To Robert, for the bagels and explaining CMYK and color spaces.

References

[1] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011* (2021).

[2] Franco Caspe, Andrew McPherson, and Mark Sandler. 2025. Waveform Autoencoding at the Edge of Perceivable Latency. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Canberra, Australia. https://doi.org/10.5281/zenodo.15699550

[3] EcoArtAsia. 2017. *Ballad On The Shore*. Retrieved February 9, 2026 from https://ecoartasia.net/BOS/BOS_chi.html A documentary and archival project on Hong Kong Tanka fishermen songs..

[4] Rebecca Fiebrink and Baptiste Caramiaux. 2016. The machine learning algorithm as creative musical tool. *arXiv preprint arXiv:1611.00379* (2016).

[5] Yuk Hui. 2019. *The question concerning technology in China: An essay in cosmotechnics*. Vol. 3. MIT Press.

[6] Robert Jack, Jacob Harrison, and Andrew McPherson. 2020. Digital musical instruments as research products. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 446–451.

[7] Steve Mann, Ryan Janzen, and Mark Post. 2006. Hydraulophone design considerations: Absement, displacement, and velocity-sensitive music keyboard in which each key is a water jet. In *Proceedings of the 14th ACM international conference on Multimedia*. 519–528.

[8] Astrida Neimanis. 2017. *Bodies of water: Posthuman feminist phenomenology*. Bloomsbury Academic.

[9] Dan Overholt. 2009. The musical interface technology design space. *Organised Sound* 14, 2 (2009), 217–226.

[10] Nicola Privato, Victor Shepardson, Giacomo Lepri, and Thor Magnusson. 2024. Stacco: Exploring the Embodied Perception of Latent Representations in Neural Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Utrecht, Netherlands, 424–431. https://doi.org/10.5281/zenodo.13904899

[11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[12] Gilbert Simondon. 2017. *On the Mode of Existence of Technical Objects*. Univocal Publishing, Minneapolis. Originally published in 1958.

[13] Atau Tanaka. 2000. Musical performance practice on sensor-based instruments. *Trends in gestural control of music* 13, 389-405 (2000), 284.