# Generative audio for music-making (GAMMa) workshop

October 10, 2025, 9:30am – 5:00pm
A-832/A-833 (8 floor) Elizabeth Wirth Music Building, 527 Sherbrooke O

## Workshop Schedule with Abstracts

### Part 1: Presentations (A832)

**09:30 —** *Welcome/Bienvenue* – Dominic Thibault

**09:45 —** *Generative Reverberation and the Future of 3D Music Production* – Jack Kelly

Generative AI has opened new possibilities for sound synthesis, yet its potential to transform spatial audio, particularly 3D reverberation, remains underdeveloped. This talk explores how AI might enable the creation of realistic, dynamic auralizations of space, not just as effects but as compositional tools. The ability to synthesize room impulse responses (RIRs) that reflect complex spatial acoustics could radically expand how we imagine and manipulate musical environments.

I will briefly survey current technologies, including diffusion models and neural reverb synthesis, that hint at these possibilities. However, significant obstacles remain. Chief among them is the lack of comprehensive spatial RIR datasets: most existing corpora are limited in scope, resolution, and diversity. Without rich real-world data, models struggle to learn the nuanced behavior of sound in space. Additionally, questions around data representation, how to encode spatiality, and model architecture present ongoing challenges.

Rather than presenting a technical solution, this talk invites reflection on the creative potential of generative reverberation and the infrastructural work needed to realize it. What kinds of musical thinking could emerge if we treated space as a generative domain? And what does it mean to author not just sound, but the acoustic conditions in which sound unfolds?

**10:15 —** *AI in 64Kb: can we do more with less?* – Philippe Esling

The research project led by the ACIDS group at IRCAM aims to model musical creativity by extending probabilistic learning approaches to the use of multivariate and multimodal time series. Our main object of study lies in the properties and perception of musical synthesis and artificial creativity. In this context, we experiment with deep AI models applied to creative materials, aiming to develop artificial creative intelligence. Over the past years, we developed several objects

aiming to embed these researches directly as real-time object usable in MaxMSP. Our team has produced many prototypes of innovative instruments and musical pieces in collaborations with renowned composers. However, The often overlooked downside of deep models is their massive complexity and tremendous computation cost. This aspect is especially critical in audio applications, which heavily relies on specialized embedded hardware with real-time constraints. Hence, the lack of work on efficient lightweight deep models is a significant limitation for the real-life use of deep models on resource-constrained hardware. We show how we can attain these objectives through different recent theories (the lottery ticket hypothesis (Frankle and Carbin, 2018), mode connectivity (Garipov et al. 2018) and information bottleneck theory) and demonstrate how our research led to lightweight and embedded deep audio models, namely
1/ RAVE in Raspberry Pi // 48kHz real-time embedded deep synthesis
2/ FlowSynth // a learning-based device that allows to travel auditory spaces of synthesizers, simply by moving your hand
3/ AFTER // A new approach to diffusion models for real-time generation of streaming audio

**10:45 —** *Coffee Break/Pause Café*

**11:00 —** *AI Collaborators in Scorch: Ethical and Creative Considerations for LLMs in Algorithmic Composition* – Norah Lorway

Scorch (https://scorch.live) is a music programming language, created by the authors, designed to be accessible for users without traditional coding experience. Originally developed as a MIDI-generating VST plugin, Scorch now supports live coding and broader media computing applications. With integrated AI features, including an AI collaborator inspired by projects like Autopia (Lorway et al 2021), Scorch enables human-AI musical co-creation.
This presentation explores how large language models (LLMs) such as ChatGPT can assist users in learning Scorch, generating ideas, and acting as virtual collaborators for composition and performance. It also examines the ethical implications of incorporating generative AI into creative coding environments, highlighting challenges and opportunities for human-computer collaboration in music and beyond.

**11:30 —** *Neural Audio Synthesis for Low-Latency Interaction* – Franco Caspe

The impact of architectural choices on the latency of deep learning models for audio remains largely unexplored in the Neural Audio Synthesis (NAS) literature, even though it is an essential component of natural and intimate instrumental interaction.
In this talk, we first present sources of latency and jitter typically found in interactive NAS models. This analysis is then applied to the task of timbre transfer using RAVE, an autoencoder for audio waveforms introduced by Caillon et al. in 2021. Finally, we present an iterative design approach for optimizing latency, culminating in a model we call BRAVE, which is low-latency and exhibits better pitch and loudness replication while showing timbre modification capabilities similar to

RAVE. We then present a proof-of-concept audio plugin compatible with audio signals from musical instruments. We hope the challenges and guidelines described in this talk can support and inspire NAS researchers in designing models for low-latency inference from the ground up, enriching the landscape of possibilities for musicians. See related paper at: https://fcaspe.github.io/brave/

**12:15 —** *Lunch/Dîner*

**13:15** — *Quigital: Performing Corporate Culture* – Elliot Britton

Quigital is a data-driven audiovisual performance framework that co-opts commercial aesthetics to both engage audiences and critique the logics of consumer culture and surveillance capitalism. Developed through iterative deployment across three works—Quigital: Digital Edition (2020), Home Comfort Advisor (2021), and the recently premiered Corporate Retreat (summer 2025) the project assembles code, music, text, and design assets into a credible "corporate fiction." Audiences encounter familiar interaction patterns (a phone tree mediated by an AI agent; a product-launch website; in-person QR workflows) that are strategically defamiliarized by artistic interventions: a choir sings machine-learning–generated text trained on smarthome manuals; automated follow-ups return audience voice recordings, transformed and recontextualized.

Participants will experience animated excerpts from past works and a preview of Quigital: Corporate Retreat, exploring how the framework fuses score, text, and code into hypersignifying gestures that read as both plausible branding and subversive theatre. We will discuss design principles, share performer and audience responses, and demonstrate how Quigital's strategies for co-opting corporate aesthetics can be adapted by others for generative composition and critical media performance.

**13:45 —** *Adding Sound to Street Views: A Generative AI framework for Street-resolution Audio Generation* – Feng Fei

While cities can be visualized with street-level resolution on digital maps, the auditory dimensions remain largely absent, hindering large-scale urban studies on sound. City-scale audio recording is not feasible, which leads to the generation of a large collection of audio data. This project proposes a generative AI-based methodology framework to create a street-resolution audio map, addressing the limitation of physics-based model in complex environmental sound synthesis.
The methodology involves three phases that focus on integrating location, time, and human activity as contextual inputs for sound generation. We will curate a novel multimodal dataset of 800h+ street walk videos from 100+ cities, annotated with geographical, time, and activity information using vision language models. A unified generative model will be developed to generate audio from a Streetview image and its context inputs. The results will be validated through similarity metrics, subjective listening tests, and comparison with field recordings.
This research introduces a new paradigm for environmental acoustics modeling and a novel problem for the machine learning community. By creating a "Google Street

View for the ears," this work will provide a powerful tool for planners, researchers, and the public, enabling the design of more acoustically pleasing, vibrant, and healthy cities.

**14:15 —** *Coffee Break/Pause Café*

## Part 2: Demos (A832 and A833)

**14:30 (A832) —** *Latent Bending: Creative Misuse of Pre-Trained Generative Audio Models* – David Piazza

Demonstration on how to use generative audio models like Stable Audio Open locally using the ComfyUI software. Specifically, I would present a custom node made for ComfyUI that allows users to intercept the latent vectors at the output of the model and perform operations on them before they are decoded back to sound, and discuss the specific effects and affordances of these operations.

**14:30 (A833) —** *Realtime latent space synthesis* – Charles Bicari

This presentation focuses on the embedded AI demo for the 3rd ArtAI Symposium presented at SAT in August 2025. It seeks to demonstrate how many models can be run on the same platform whilst keeping CPU capacity under control. Briefly, it runs four models simultaneously on a rPi5 and the CPU stays under 60%.
The demo runs a mediapipe script for hand recognition, sending 63 OSC [x,y,z] points into Wekinator in order to train an interpolated mapping with 16 OSC outputs directed into the nn~ operators in PureData. Two audio models are running in PureData thanks to the nn~ operator, exposing 16 dimensions of their latent space in which we then stimulate with our trained mapping in Wekinator.

**15:00 (A832) —** *TISSERV (Tool for interactive sound space exploration with reactive visuals by classification)* – Josue Medina Rueda

An interactive system for AI-informed sound-space exploration that maps a corpus of recordings onto two continuous axes and enables real-time, descriptor-driven navigation using concatenative synthesis (MuBu/CataRT in Max). Combined with a FLUCOMA-trained MLP that classifies incoming audio from MFCC features into categories (bird calls, insects, voices) and sends OSC messages to another software (TouchDesigner) to trigger reactive visuals or select video assets. Supports non-linear sequencing, live timbral manipulation, and audiovisual interaction for research-creation, performance, and installation contexts.

**15:00 (A833) —** *Formachron - A tactile grid-based polymetric, polytuplet, polyphasic sequencer* – Adam Tindale

Formachron is a software instrument that addresses the limitations of traditional grid sequencers which typically require a predefinition of a single subdivision per

sequence. This system aims to facilitate the creation of polyrhythmic, polymetric, and polyphasic patterns. By defining regions on the grid where rows represent subdivisions and columns represent beats, users can directly create rhythms that mix tuplets. The notation also offers efficient grid space utilization, as only cells describing an event are used, allowing multiple rhythms or sequences on the same grid simultaneously. The project prioritizes making the expression and editing of complex rhythms a first-order operation, enhancing performance possibilities and potentially expanding musicians' rhythmic vocabulary. This approach focuses on a new method of using grids, applicable to existing hardware, rather than inventing new devices. Formachron has been implemented using an Ableton Push, though more devices are planned for future support.

In this demo the expression of sequences with multiple subdivisions, lengths, and phases. Short demos will explore triggering events in Ableton Live and driving a modular synthesizer. The demo will culminate in a polyrhythmic nightmare created with audience participation.

**15:30 (A832) —** *Using Generative AI for Adaptive Learning in Music* – Alberto Acquilino

While music educators strive to provide personalized learning paths, existing educational resources often lack the flexibility to adapt to diverse student needs. This project is addressing this gap by integrating Generative AI into a modular, web-based platform that empowers teachers to design customized exercises and adjust lessons based on individual progress. The interface adapts exercises according to musical parameters and user prompts, with outputs available in JSON, piano roll, and MIDI formats. Although in its initial phase, the project highlights significant potential for enhancing inclusivity and accessibility in music education, with future directions including support for diverse notations and musical traditions.

**15:30 (A833) —** *On Using IRCAMs RAVE with Acoustic Instruments* – James Ford

I will improvise explorations of a rave latent space using an acoustic instrument controller (digeridoo). I will make the metaphors of my gestural acquisition methods for navigating the latent space the subject of my talk. (trumpet valves for discrete events, formant shifts for latent space quadrant shifts)

**16:00 (A832) —** *Latent Manipulation in Virtual Game Environments* – Austin Tecks

By mapping player interactions within an Unreal Engine sketch to the latent space of a generative audio model, the project examines how movement through virtual space can be both a driver for sonic exploration as well as a source of inspiration for virtual environment design. A short live demo will be provided as well insights into the primary tools (RAVE, Unreal and Flucoma) and methods used to create the sketch. Additionally, the presentation will explore how this workflow can serve as a

powerful prototyping tool for immersive installation outside of the traditional game environment.

**16:00 (A833) —** *Composition et transduction créative : Du motif naturel à l'œuvre sonore. Le cas de "Pekuakami" comme alternative à la génération par IA* – Raphaël Néron

La musique générative contemporaine est largement dominée par les paradigmes de l'intelligence artificielle (réseaux de neurones, modèles de langage, etc.), souvent perçus comme la frontière ultime de l'innovation. Cette proposition offre un contrepoint critique à cette tendance en présentant une approche alternative et délibérément ""low-tech"" : une composition générative par algorithmie déterministe, dont la source créative première est l'observation et la modélisation des motifs et systèmes présents dans la nature.
À travers le prisme de mon projet Pekuakami, je détaillerai mon processus de ""transduction créative"" : la transformation de motifs d'écorce de bouleau en partitions à travers un algorithme de transduction d'image en fichiers midi. Je discuterai de comment, pour moi, des systèmes naturels (e.g., la croissance fractale d'une fougère, le rhythme erratique de la pluie, l'activité sismique, etc.) sont d'une richesse et d'une complexité infinies, et qui peuvent fournir une matrice compositionnelle bien plus organique, imprévisible, innovante et conceptuellement fertile que le traitement statistique de vastes corpus de données musicaux.

Démonstration musicale
Présentation d'extraits de la pièce Wikwas.