

Simulation Intelligence in Musical Acoustics

From Sound and Vibration Measurements to ...

Cumhur Erkut | Aalborg University Copenhagen



AALBORG
UNIVERSITY
COPENHAGEN



ICSV32

32nd International Congress
on Sound and Vibration

05 - 10 JULY 2026

I S T A N B U L

The annual congress of the International
Institute of Acoustics and Vibration (IIAV)



Multisensory Experience Lab
<https://www.melab.dk>

OVERVIEW

TALK ROADMAP

1 - The Classical Foundation of Simulation Intelligence

- DSP and physical modeling as the bedrock of musical acoustics

2 - The Machine Learning Turn

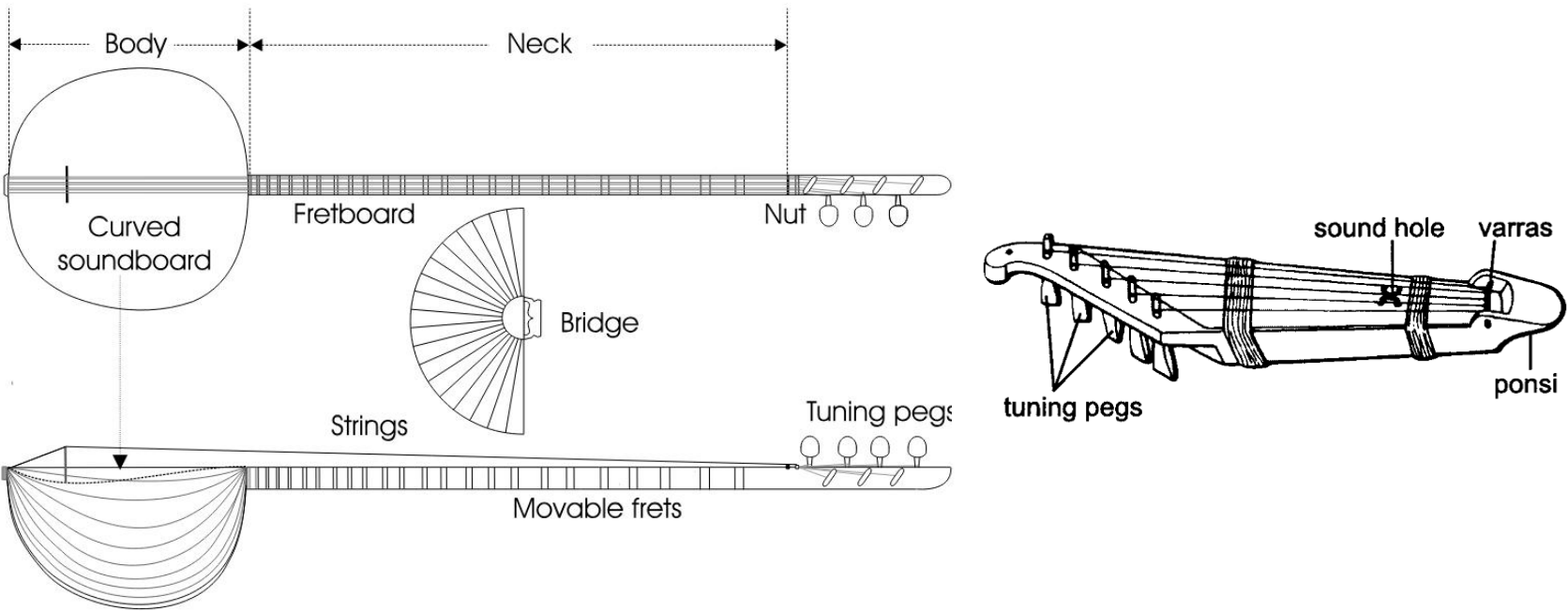
- How data-driven methods are transforming the pipeline

3 - Two Demonstrators

- Measurement-driven studies + Physical modeling in XR

4 - The AI Audio Engineering Framework

- Harnessing the current practice in Agentic AI for Audio



Example 1: Resurrecting Tromba Marina.

Also Flamenco, Nordic and Middle Eastern traditions.



Example 2: Generative Choreographies combining

sound, music, and movement. Also therapy.



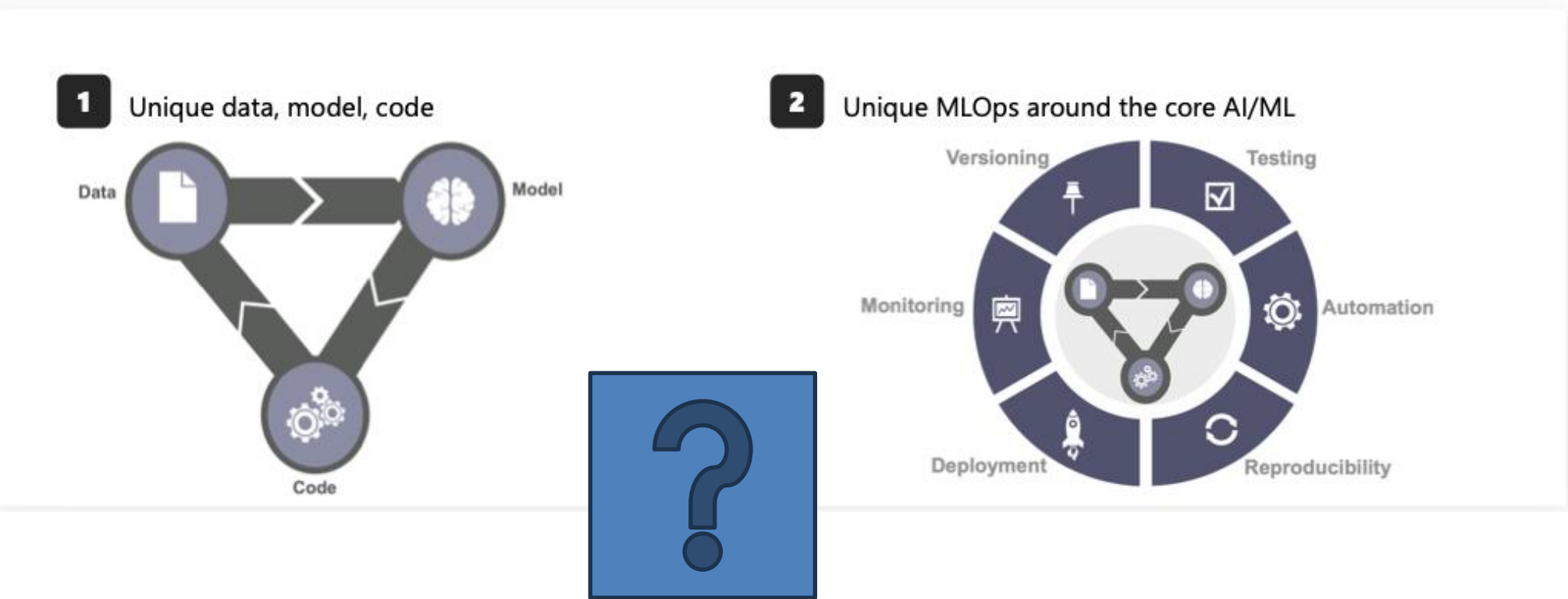
Methods

Plugins

Real-time inference

Differential Digital Signal Processing

Generative Models



SIMULATION INTELLIGENCE — THE NINE MOTIFS OF SCIENTIFIC OS

Building blocks of the SI “operating-system” stack — a roadmap merging scientific computing, simulation & AI (Lavin et al., 2021)



Motifs are interconnected & interdependent — like layers of an operating system; coordinated, they accelerate scientific discovery in silico.

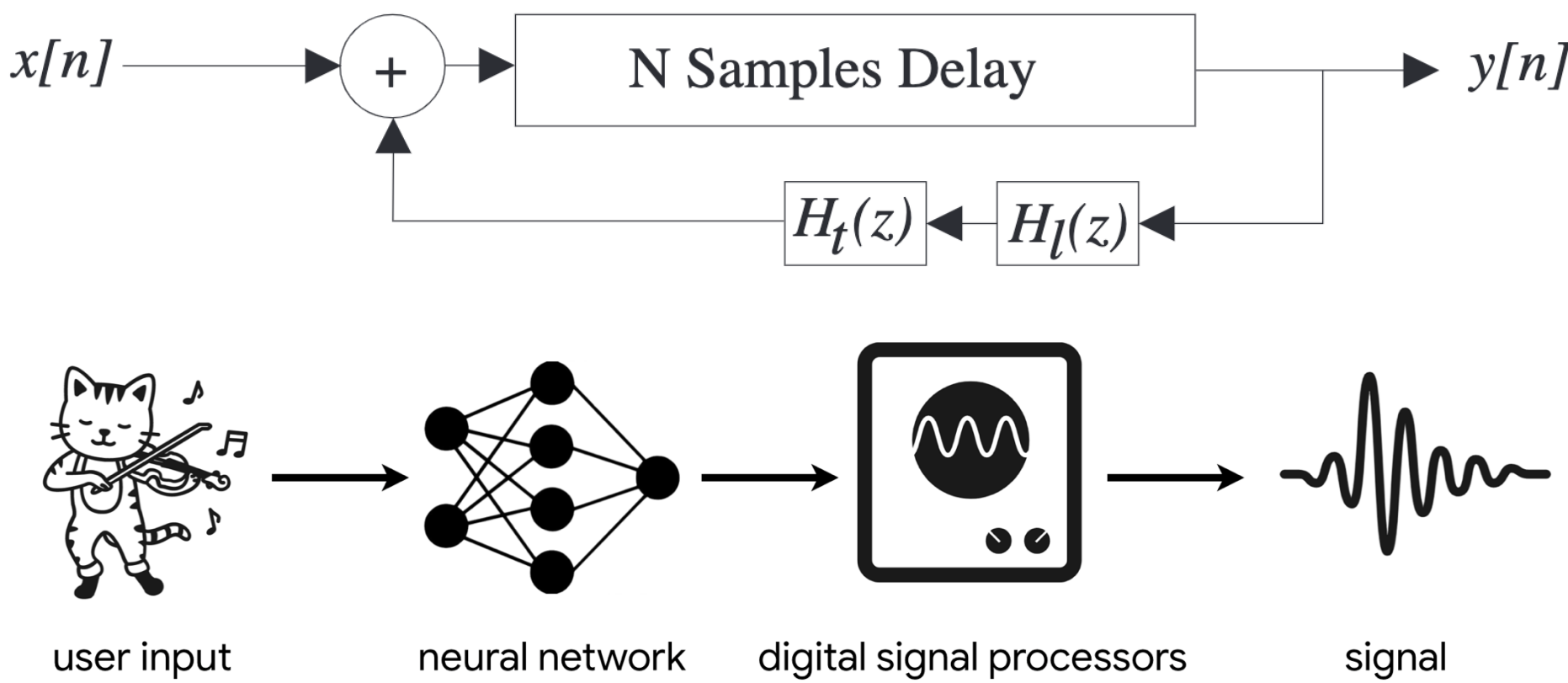
PHYSICAL MODELING: A MATURE DISCIPLINE FOR SIMULATION INTELLIGENCE

A MATURE SYNTHESIS & SIMULATION TECHNOLOGY

- Vibro-acoustic systems simulated in real time
- Digital prototyping for traditional and unconventional instruments
- Perceptually meaningful control interfaces support expressivity
- Computational tractability with physical interpretability

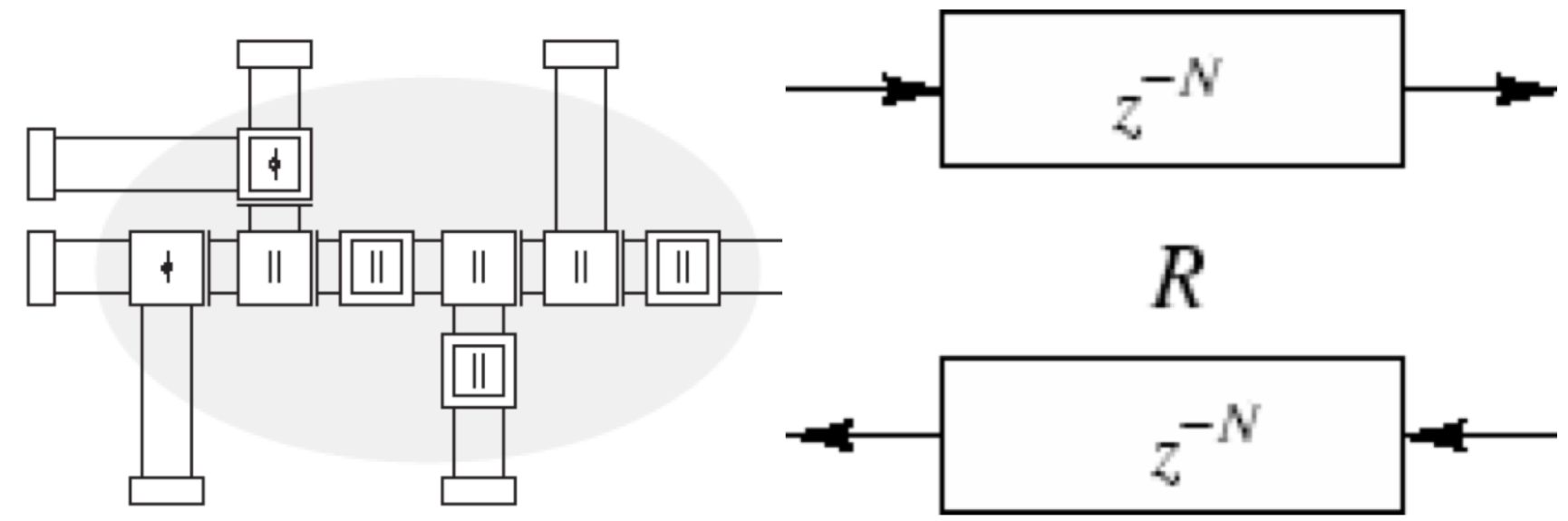
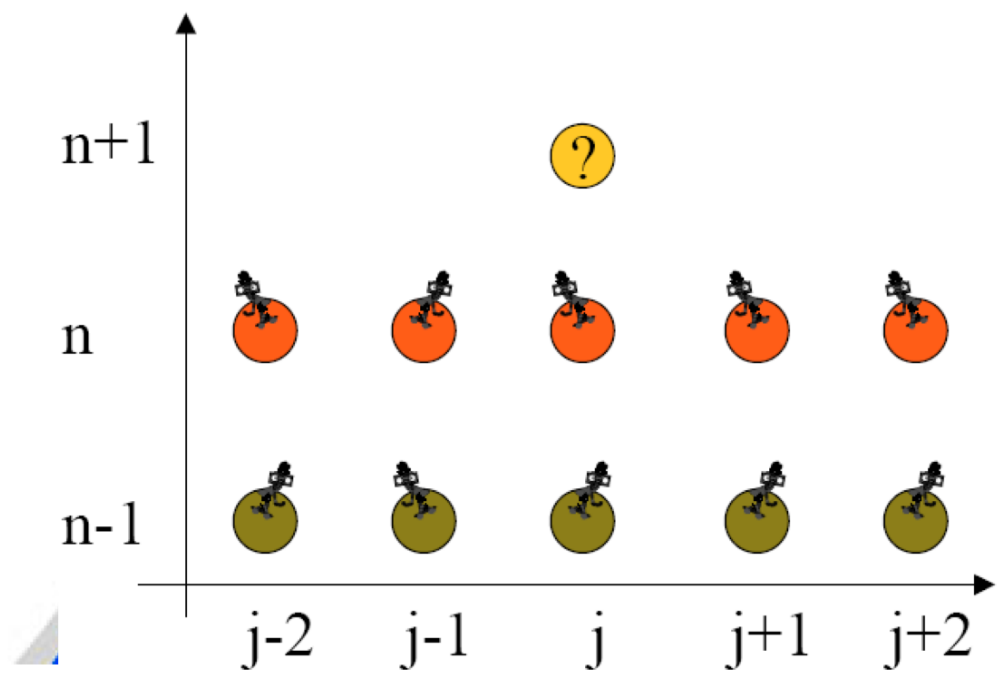
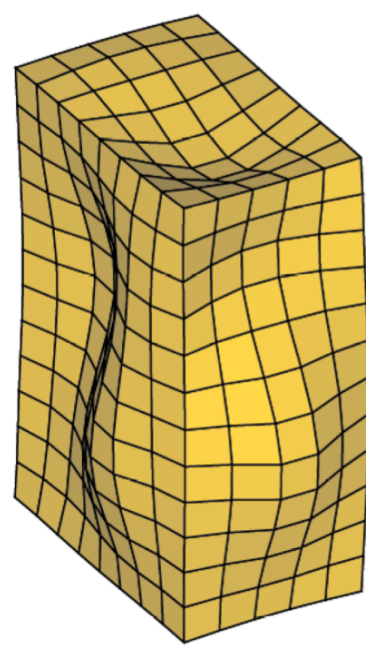
KEY PARADIGMS

- Finite-difference time-domain (FDTD) simulation
- Digital waveguide synthesis
- Modal synthesis and spectral modeling
- Coupling vibro-acoustic models with haptic/gestural control



$$K\mathbf{P} + i\omega C\mathbf{P} - \omega^2 M\mathbf{P} = i\rho\omega\mathbf{F}$$

M, C, and K are mass, damping, and stiffness matrices



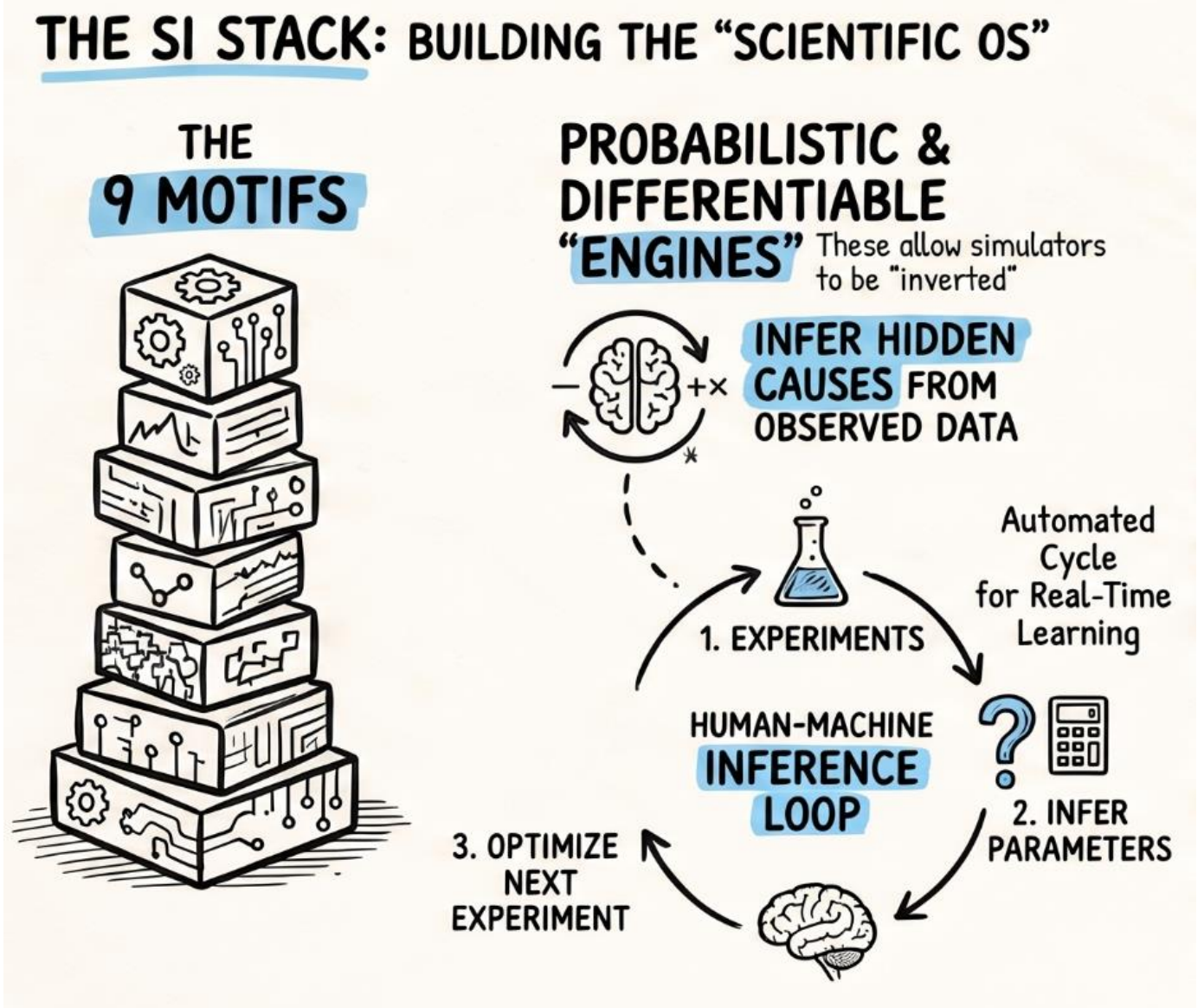
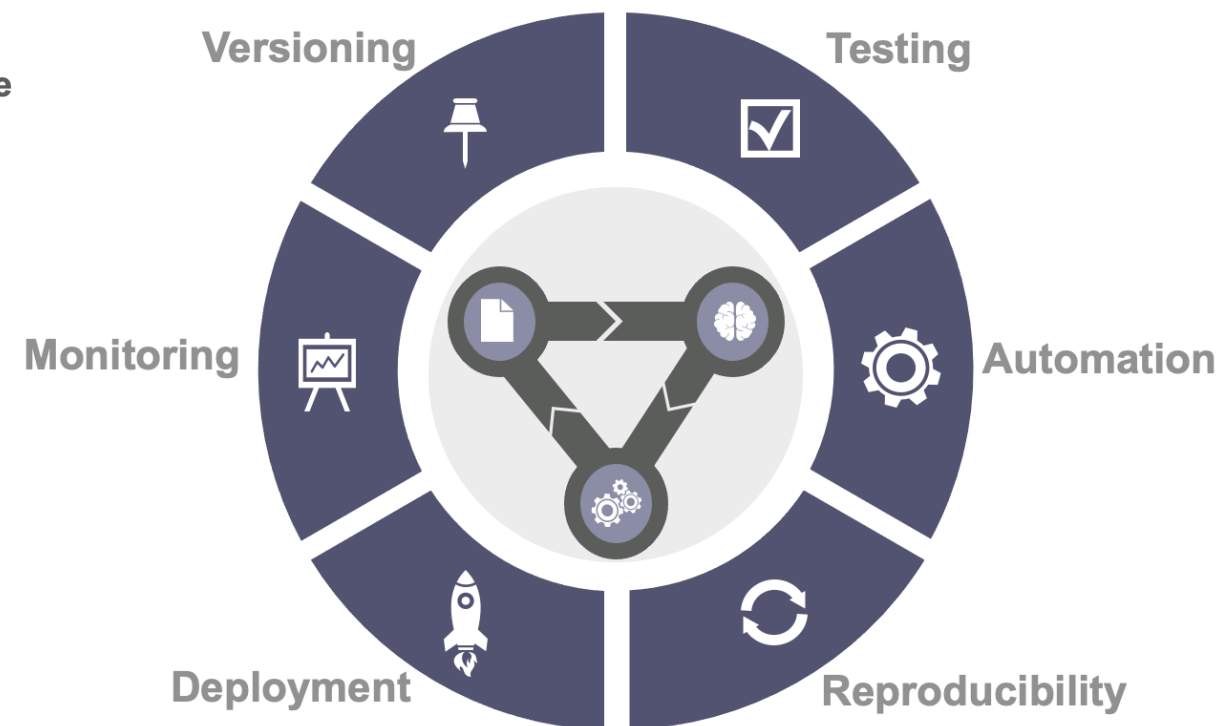
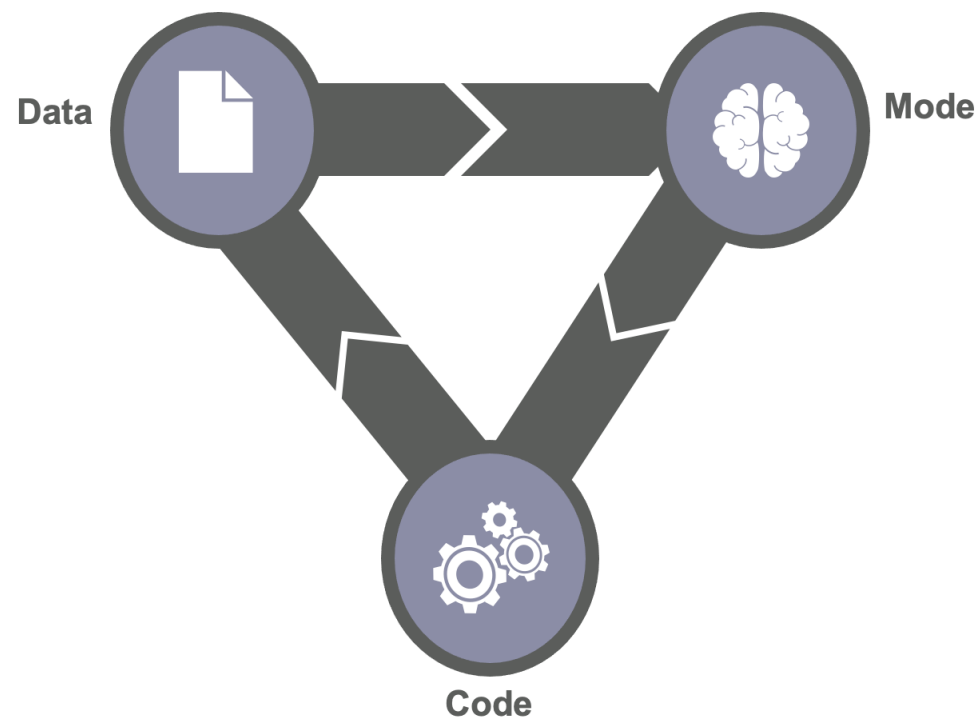
AI AUDIO ENGINEERING IN SIMULATION INTELLIGENCE

A PARADIGM SHIFT

Not just new algorithms — new workflows, evaluation practices, and deployment contexts

THE SHIFT IN PRACTICE

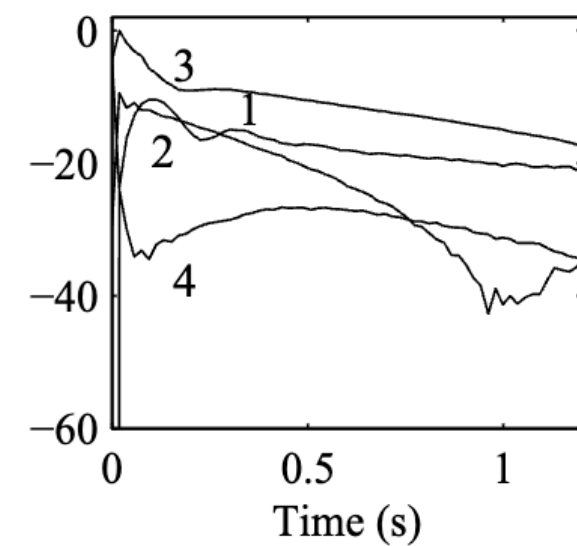
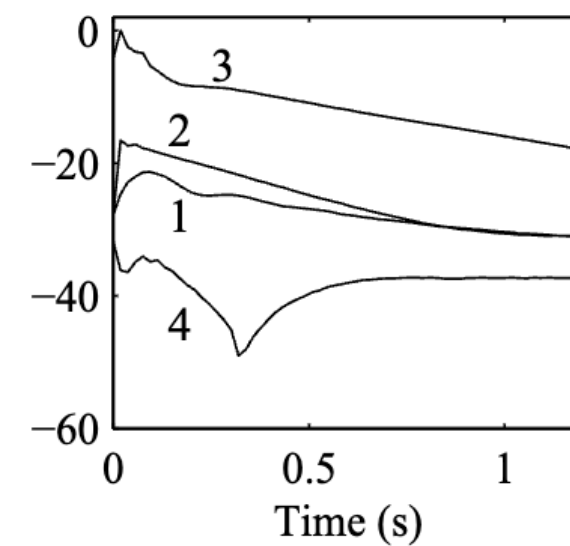
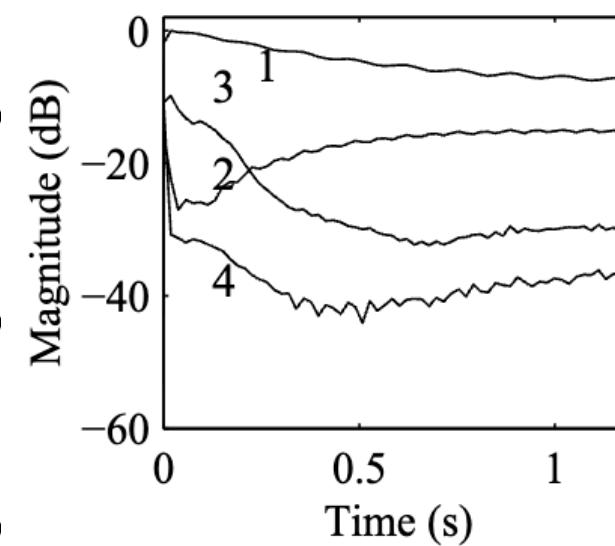
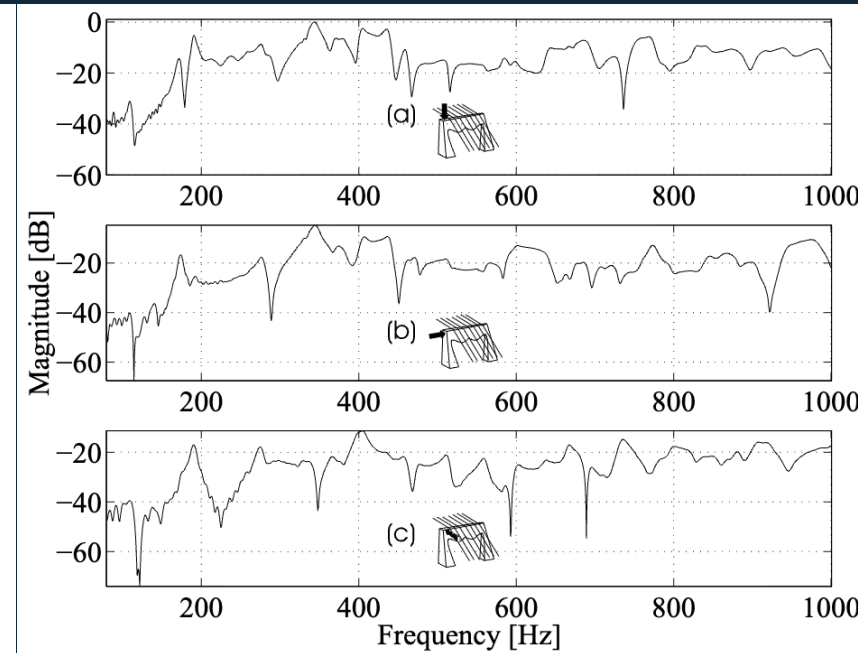
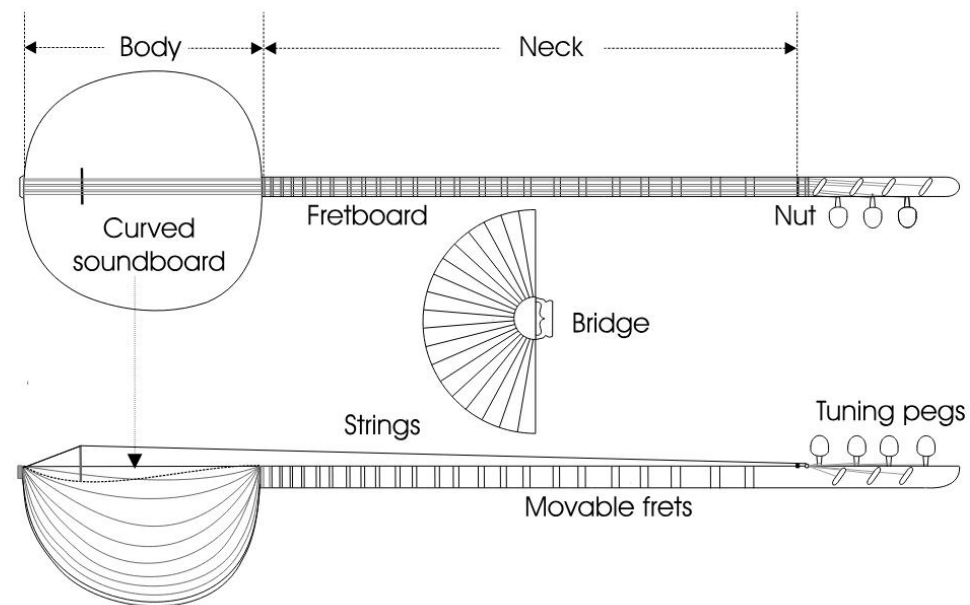
- Classical DSP: algorithms terminate at release, AI Audio Engineering assumes post-deployment evolution
- Perceptual feedback integration and dataset governance
- Realism and expressivity judged by performer experience



DEMONSTRATOR 1

MEASUREMENT-DRIVEN STUDIES

From high-resolution instrument data through model constraints to Simulation Intelligence



DATA AS MODELING CONSTRAINT for SIMULATION INTELLIGENCE

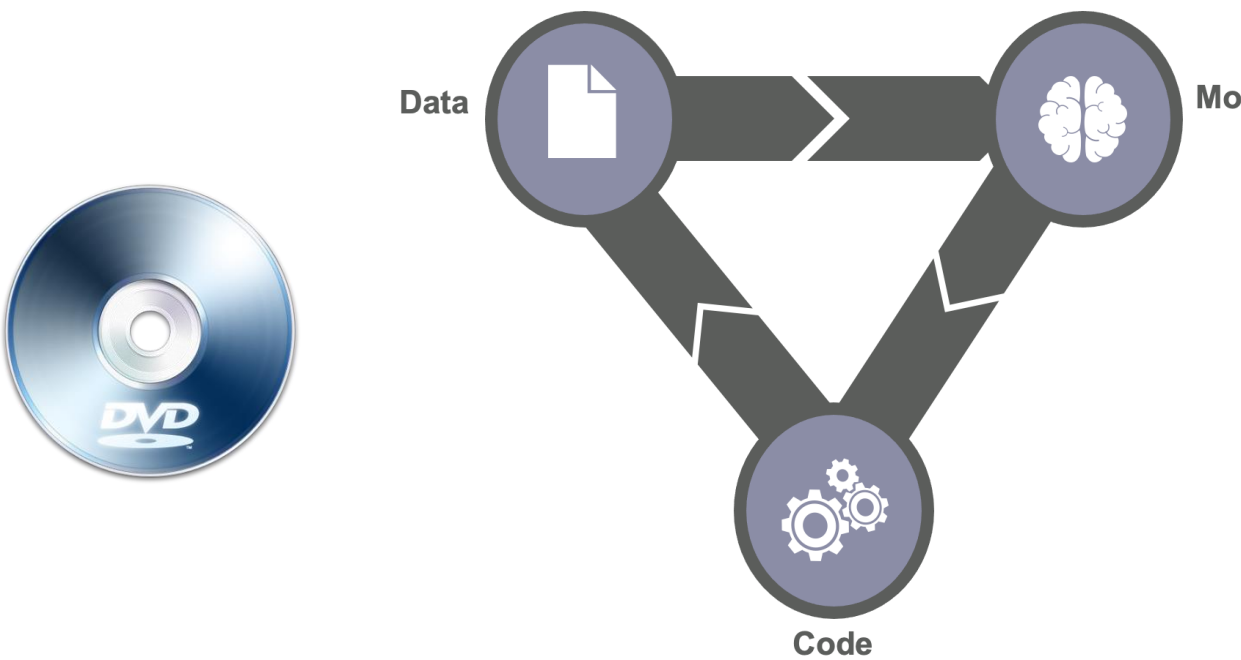
MEASUREMENT EXTRACTS

- Partial frequencies and decay rates
- Linear and nonlinear modal features
- Structural coupling and radiation patterns

→ CLASSICAL MODEL

Physical models parameterized by measurement-derived priors

- Reduced parameter search space
- Constrained model architecture



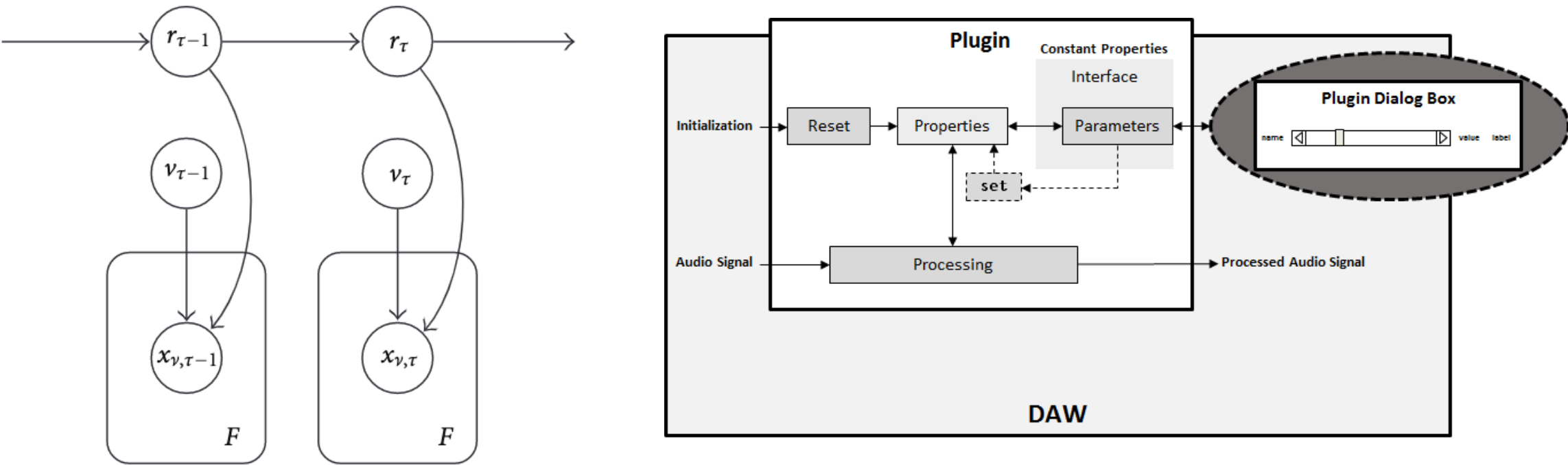
THE HYBRID APPROACH

- Analytical model captures the physics
- Neural surrogate learns residuals
- Residuals = behavior not captured analytically

KEY PRINCIPLE

Data as a modeling constraint for simulation intelligence, not as an end-to-end solution

- Measurement narrows the hypothesis space
- Models remain interpretable and physically grounded



Şimşekli, Umut, Antti Jylhä, Cumhuri Erkut, and A T Cemgil. “Real-Time Recognition of Percussive Sounds by a Model-Based Method.” *EURASIP Journal on Advances in Signal Processing* **2011** (1): 291860.

CASE STUDY: HISTORICAL INSTRUMENT ACOUSTICS

CASE STUDY: TURKISH TANBUR / FINNISH KANTELE

A historical plucked-string instrument with rich timbral complexity

METHODOLOGY

- High-resolution acoustical and vibration analysis
- Linear and nonlinear vibration features extracted from measurement data
- Calibrating nonlinear digital waveguide against measurement
- Neural surrogates + transfer learning for related instruments

OUTCOME

Measurement-derived priors inform both classical modeling and neural surrogates

Robustness to measurement noise evaluated; remaining challenges identified

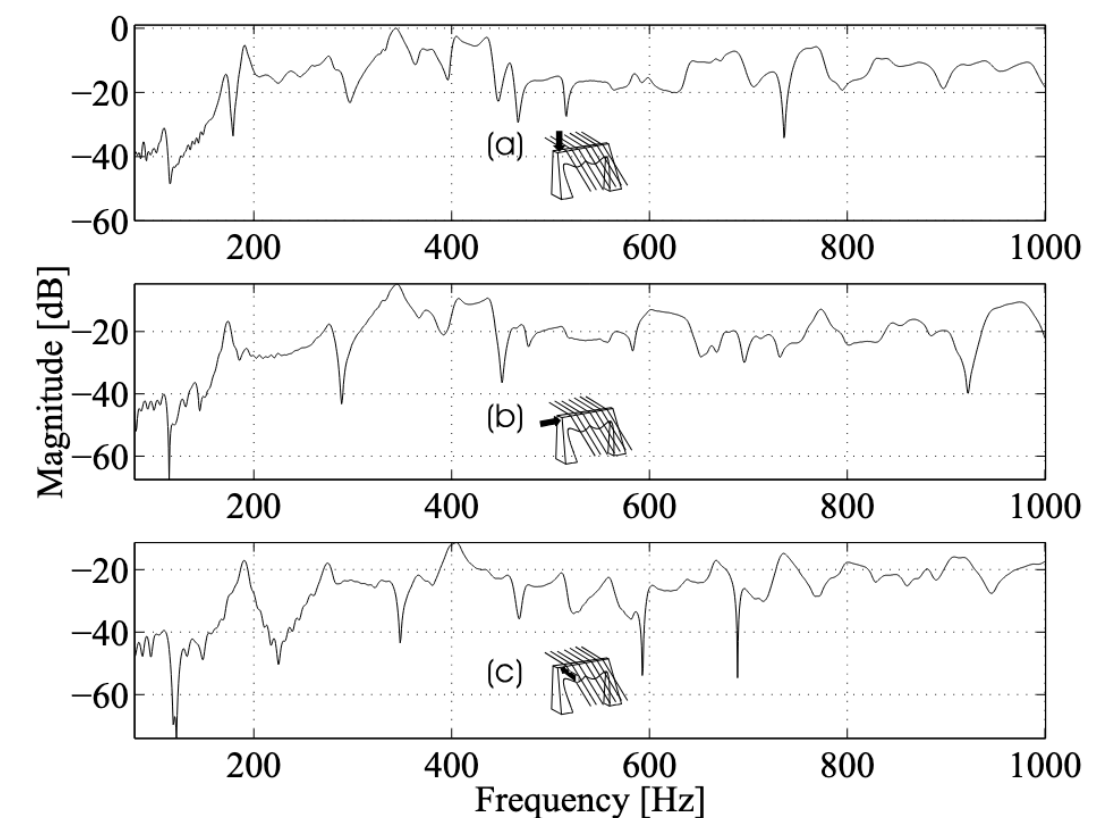
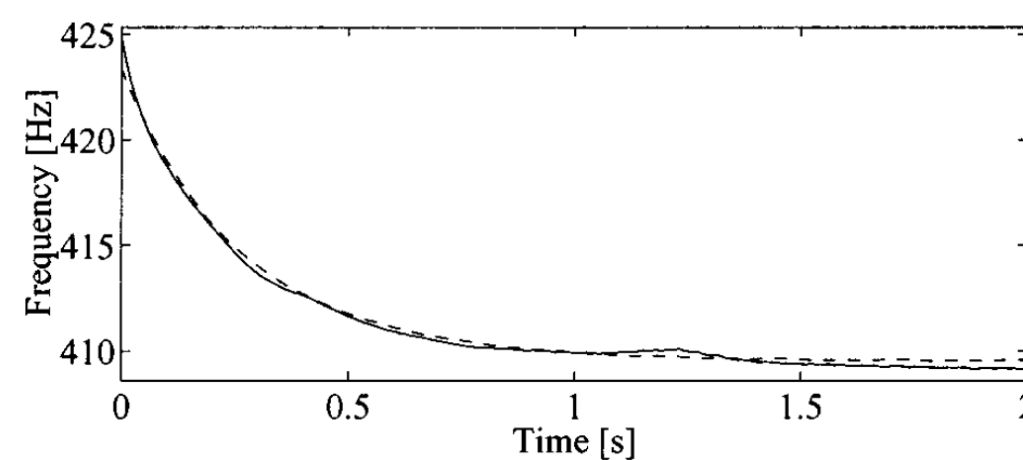
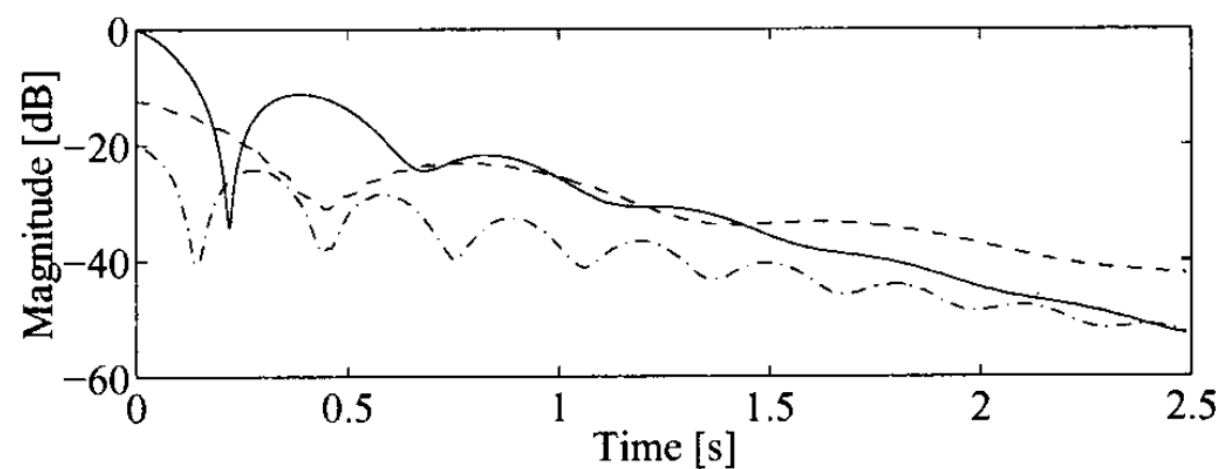
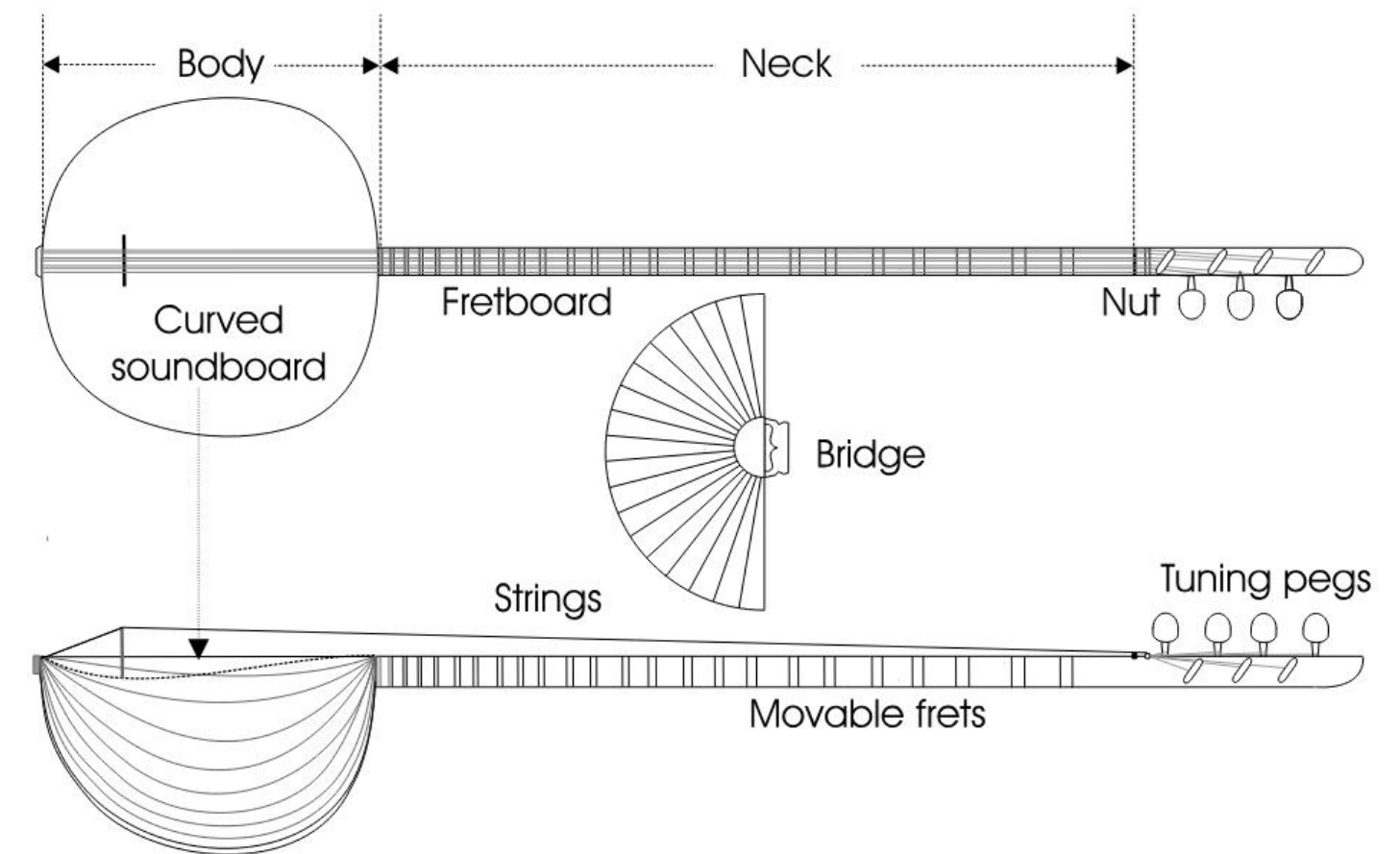
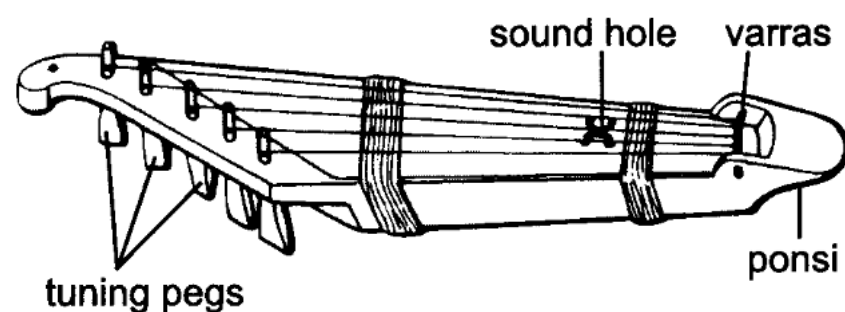


FIG. 15. Variation of the fundamental frequency in a kantele tone (obtained from the modified kantele) resulting from a fortissimo pluck (solid line) and a synthetic tone (dashed line).



DEMONSTRATOR 2

PHYSICAL MODELING INTELLIGENCE IN XR

Virtual instruments for experimentation, pedagogy, and performance

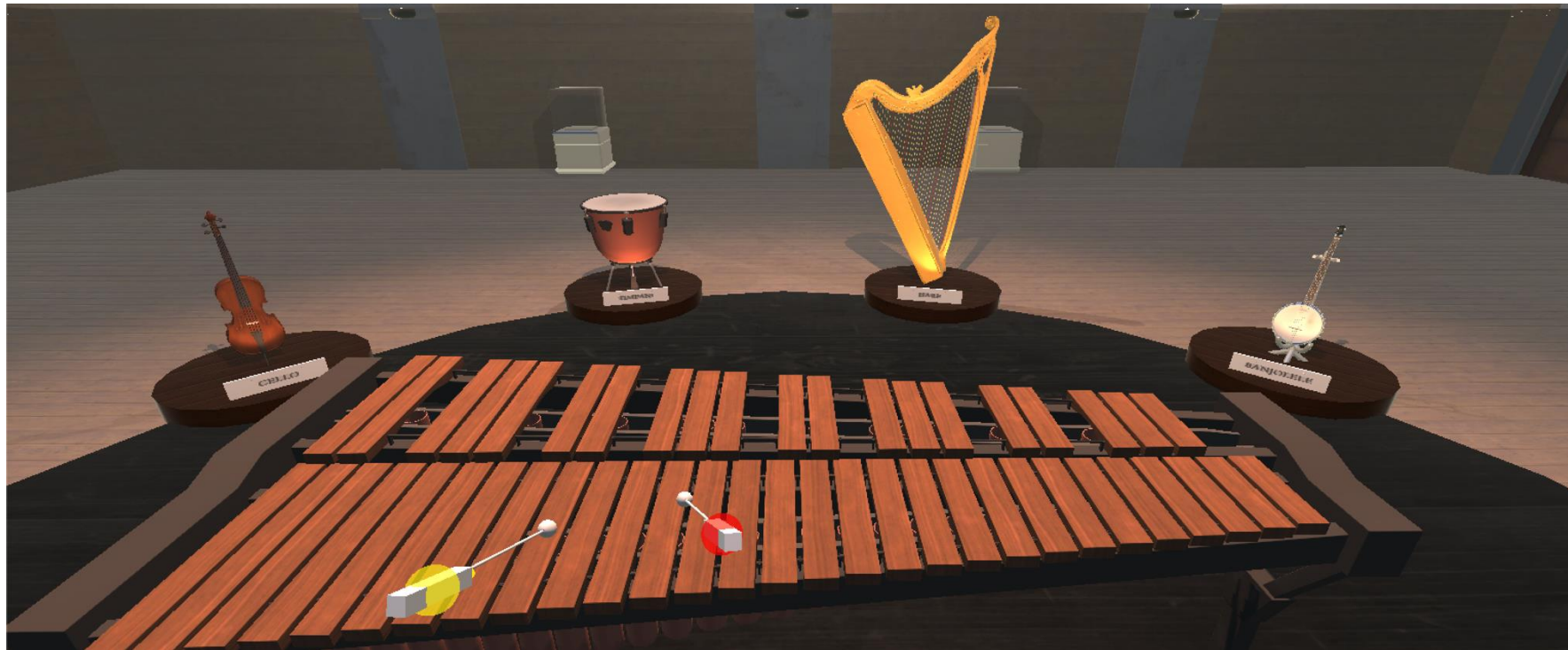
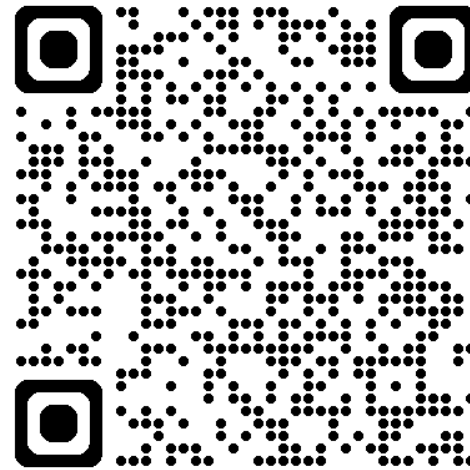
VIRTUAL INSTRUMENTS IN EXTENDED REALITY

THE XR INSTRUMENT SYSTEM

- Virtual strings, membranes, plates
- Driven by gesture or haptic interfaces
- Physically-informed or machine-augmented synthesis
- Binaural and room acoustic spatialization

HAPTIC CONTROL

- Perceptual feedback loop with the player
- Spatial audio creates convincing sonic interactions
- Ecological validity approximating real playing



XR ENABLES

- Immersive instrument experiences
- Novel pedagogical scenarios
- Performance and composition tools
- Rapid prototyping of instrument designs

REAL-TIME PIPELINE

Gesture detection → physical model → synthesis → spatialization

- Latency-critical: DSP handles time-sensitive stages
- Neural components: auxiliary inference tasks



<https://github.com/SilvinWillemsen/SIVEtoolkit>

XR AS PERCEPTUAL EVALUATION PLATFORM (<https://anders-bargum-cv.github.io/projects>)



XR AS A PERCEPTUAL DATA COLLECTION PLATFORM

Beyond ecological validity – XR instruments actively generate evaluation data

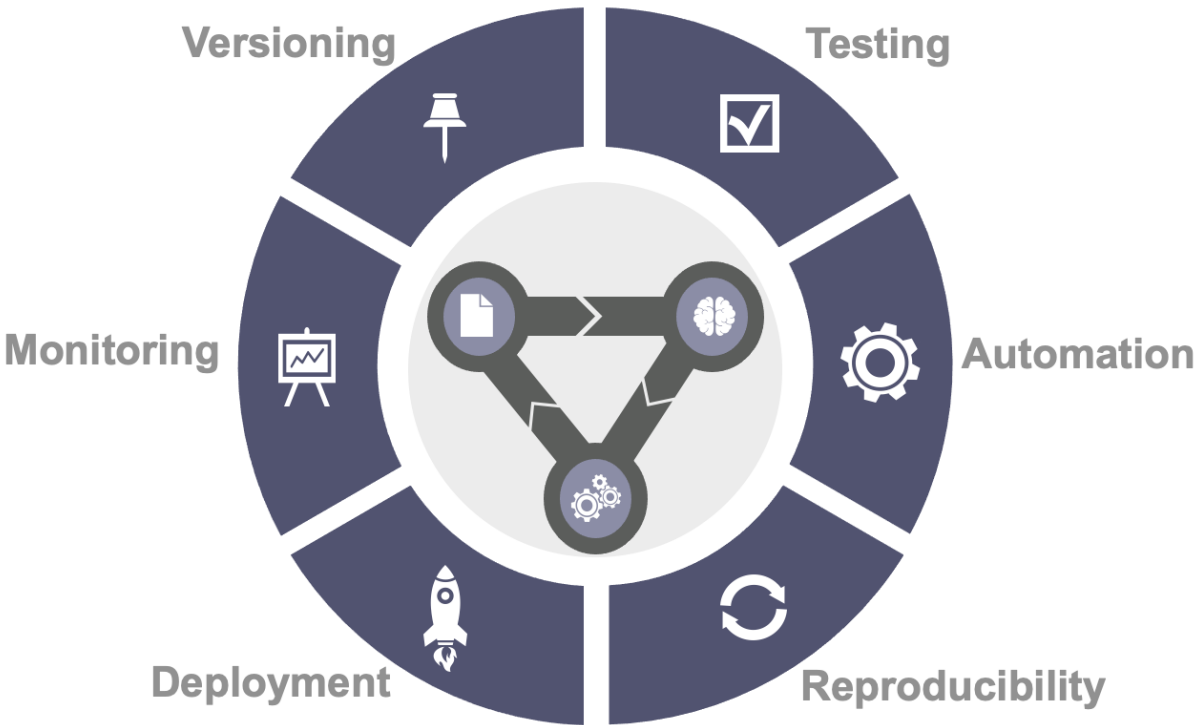
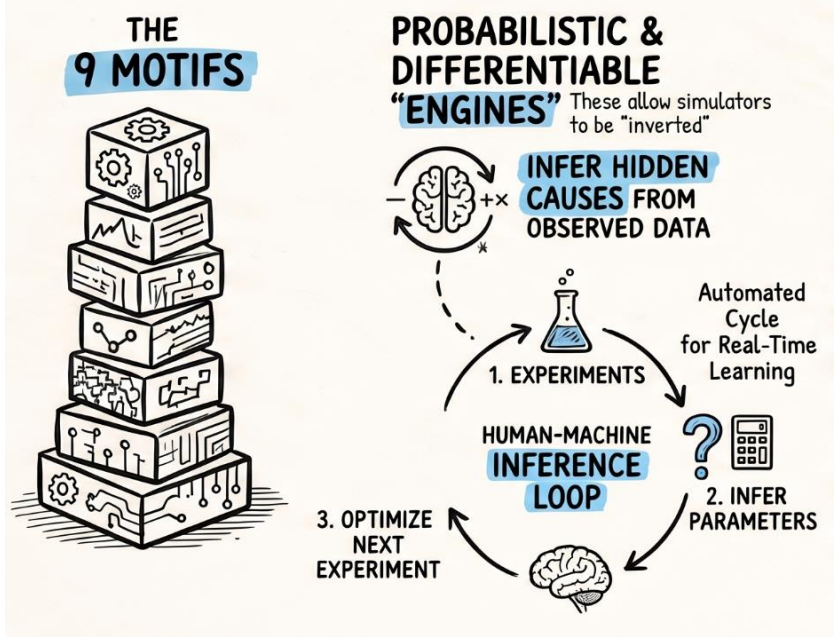
CAPABILITIES

- In-context subjective testing (vs. isolated lab listening)
- Preference modeling and perceptual rating at scale
- Player-in-the-loop adaptation of synthesis parameters
- Performance capture: gesture, timing, expressive nuance

WHY THIS MATTERS

These tasks are historically expensive or impractical to run at scale
XR lowers the barrier while increasing validity and sample diversity
Perceptual data feeds directly back into model refinement

THE SI STACK: BUILDING THE “SCIENTIFIC OS”



Deep Learning Based Voice Conversion for Avatar Therapy Voice Design in Low Latency and Streamable Environments

Anders R. Bargum^{1,2}, Cumhur Erkut¹ and Simon Lajboschitz²
¹Multi-Sensory Experience Lab Aalborg University Copenhagen², Khora & Heka VR

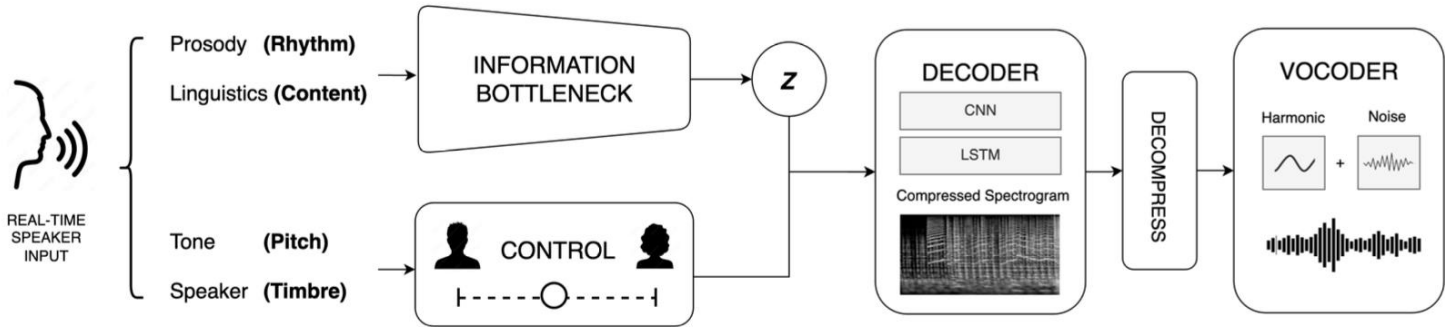


Figure 1: Controllable voice conversion with real-time optimizations.

Objectives

This project researches interpretable, streamable and real-time possibilities in deep learning based voice conversion (VC) pipelines. The goal of the project is to create an audio engine allowing voice control and design for AVATAR therapy - a new form of virtual reality treatment that through therapeutic sessions allows individuals who hear voices to have a dialogue with a digital representation of their presumed persecutor.

Conventional Voice Conversion

Voice conversion seeks to render an utterance from one speaker to sound like that of a target speaker. State-of-the-art VC pipelines most commonly consist of autoencoders (AE's) treating the conversion problem as a style transfer problem. The pipeline traditionally contain three stages:

- ▶ 1) Different speech features are disentangled into different embeddings through an information bottleneck (encoder).
- ▶ 2) The features one wish to convert are substituted with an embedding from the target speaker and decoded into an acoustic representation.

Proposed Pipelines

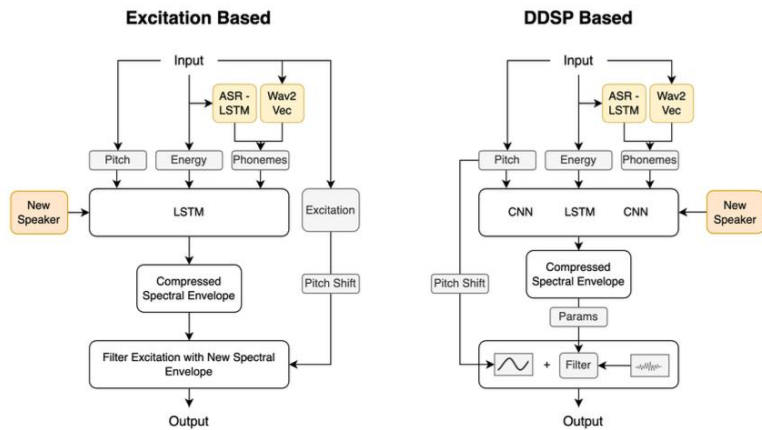
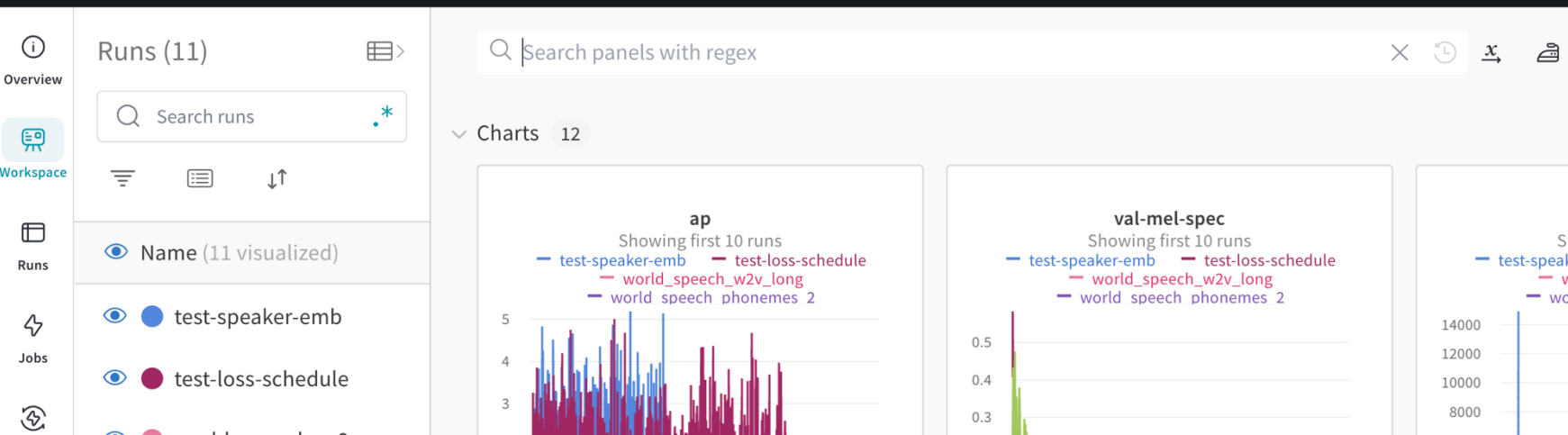
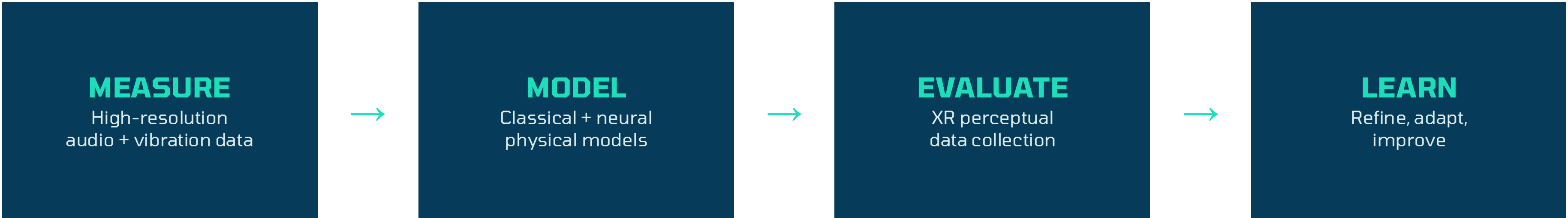


Figure 2: Two different real-time efficient voice conversion pipelines.

abargum > Projects > ddsp-world synthesis



THE CONTINUOUS WORKFLOW LOOP



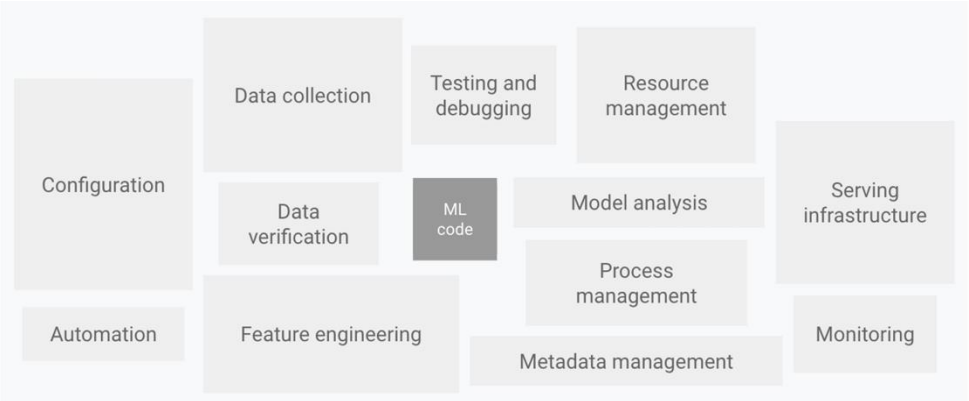
Measurement constrains model space • Models generate real-time audio • XR collects perceptual data • Learning refines models

MinIO

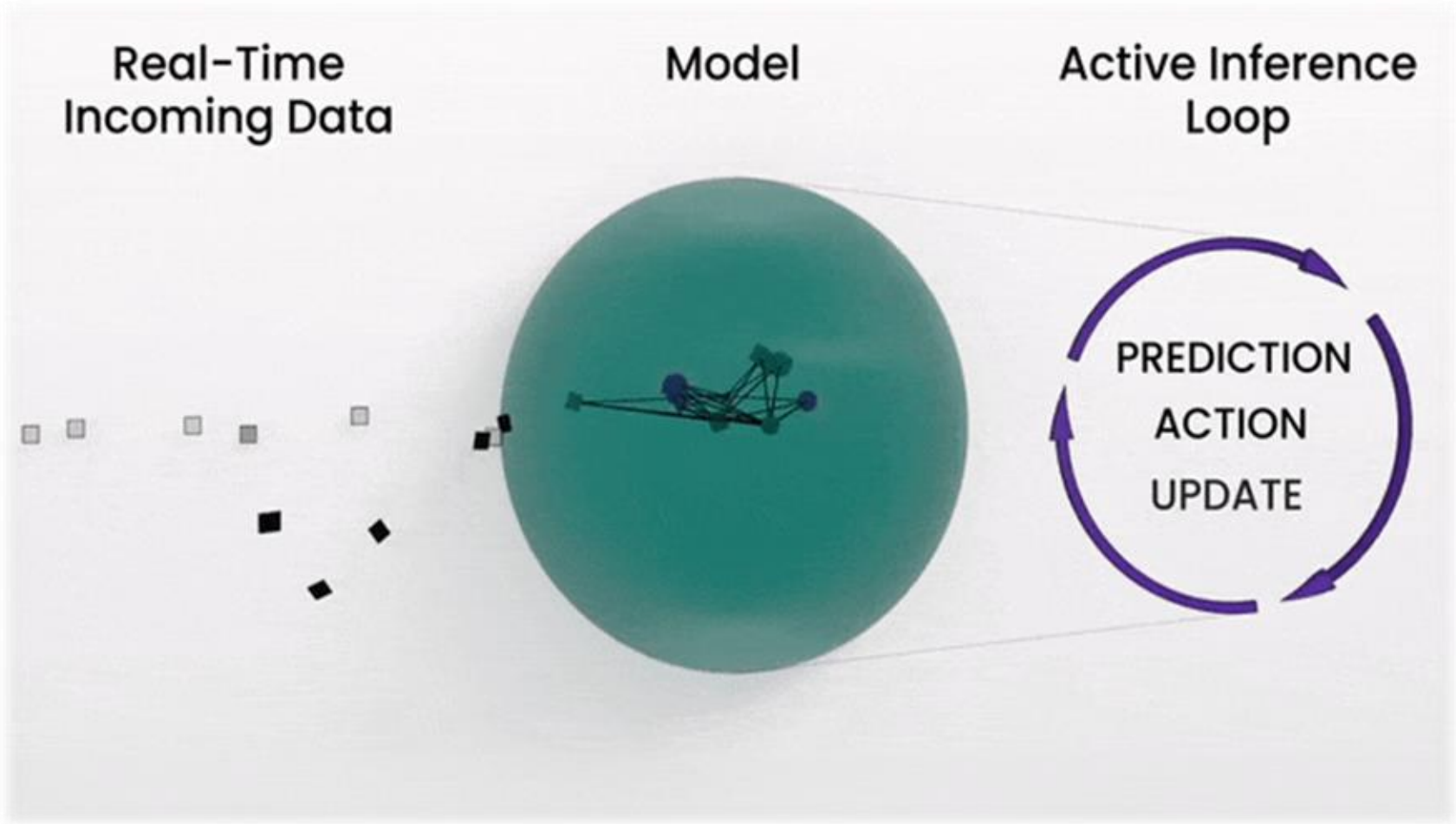
MinIO is a software-defined object store that runs on any cloud or on-premises infrastructure. It is built for large scale data workloads and supports Kubernetes, encryption, replication, immutability, and data lifecycle management. MinIO offers bucket-level granularity and supports both synchronous and near-synchronous replication.

▶ Open application

Many tools available for all researchers / students in Denmark
See <https://hpc.aau.dk> Also in EU through LUMI HPC.



But beware the **technical debt**.



CONCLUSION

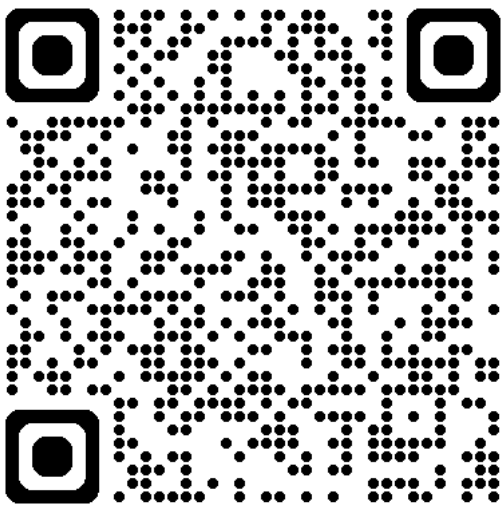
AN OPEN INVITATION TO COLLABORATE

This framework invites collaboration in Simulation Intelligence across:

- Acoustics • Digital Signal Processing • Spatial Audio • AI & Machine Learning

OPEN QUESTIONS FOR THE COMMUNITY

- How do we evaluate “realism” in Simulation Intelligence?
- What dataset governance standards apply to trained audio models?
- How should perceptual feedback be integrated into training pipelines?
- Can simulation intelligence extend to non-musical vibro-acoustic domains?

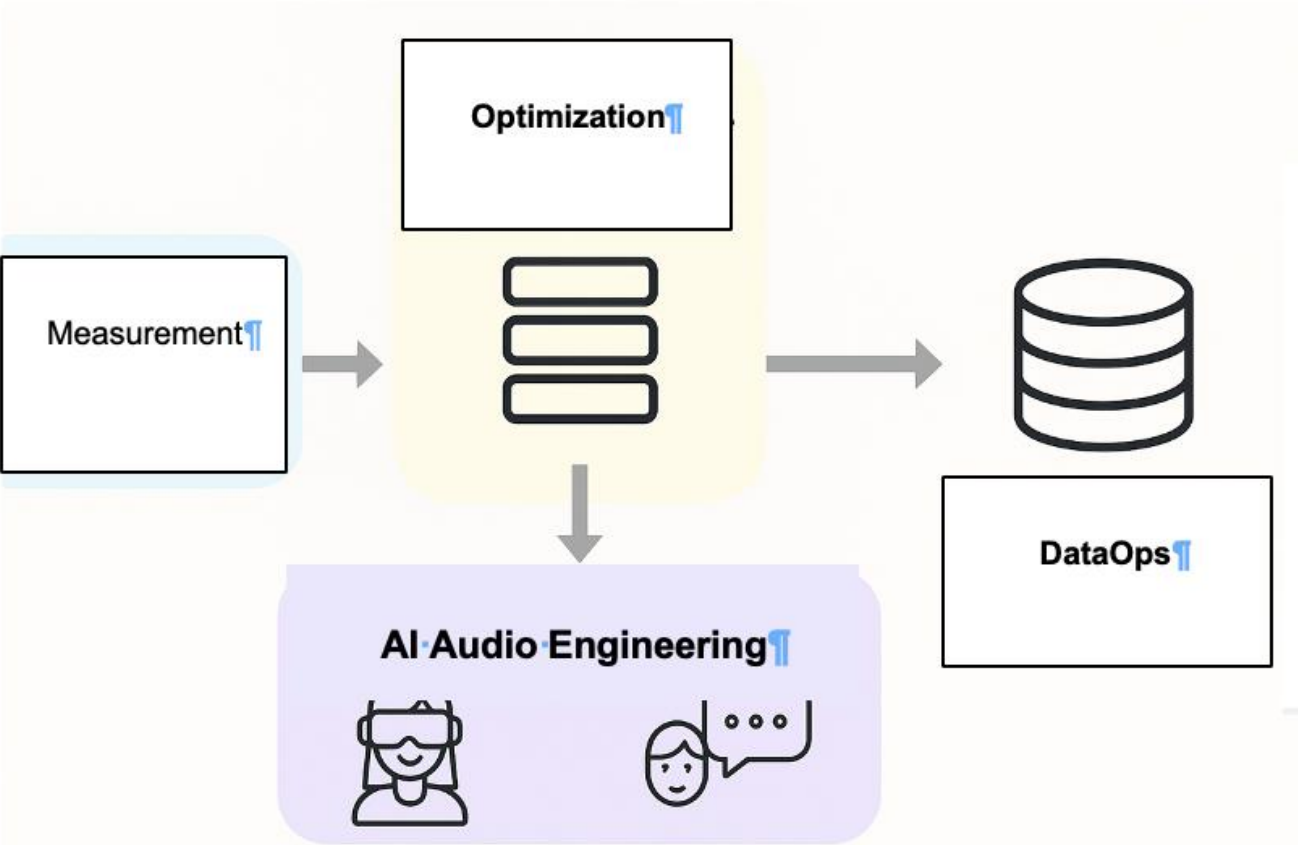


Thank you • Cumhur Erkut • Aalborg University Copenhagen • AES Europe 2026

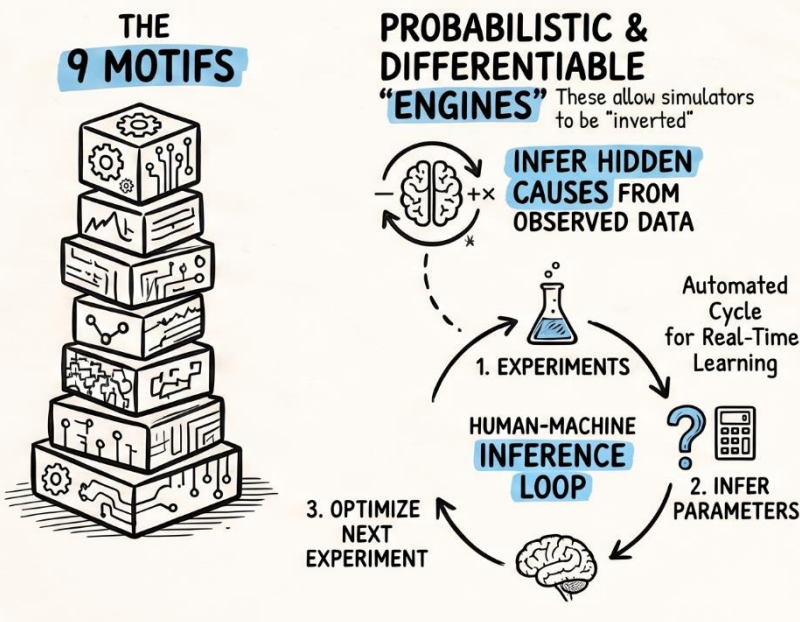
<https://moduler.aau.dk/course/2026-2027/TECHSPA126>

FACTS ABOUT THE MODULE

Danish title	Signalbehandling for AI-ingeniører
Module code	TECHSPA126
Module type	Course
Category	Separate module
Semester	November 2026. Copenhagen, Building A, 2.0.023 04.11 08:15 - 11:45 11.11 08:15 - 11:45 18.11 08:15 - 11:45 25.11 08:15 - 16:15
ECTS	5
Language of instruction	English
Location of the lecture	Campus Copenhagen
Responsible for the module	Markus Löchtefeld, Cumhur Erkut, Razvan Paisa, Nicola Walker



THE SI STACK: BUILDING THE “SCIENTIFIC OS”



AI Engineering In Audio Signal Processing

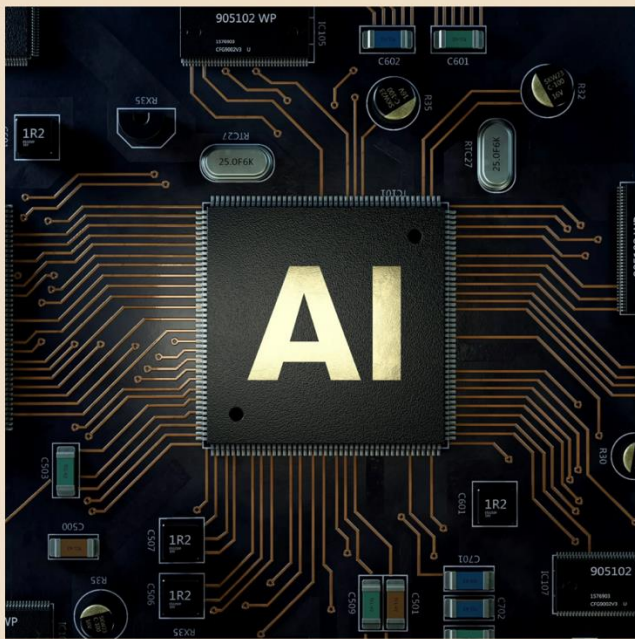
SOUND DISCOVERY: A forum for sound exploration

AI in audio moves rapidly from research prototypes to deployed real-world systems. 2026 is coined as the year of Voice AI Interaction, and Large Language Models (LLM) can now be inferred from small, embedded devices like mobile phones, hearing aids, and headphones.

However, the growing size of LLM knowledge caches are still too heavy for these platforms to support, restricting the full potential of Generative and Agentic AI to big GPU clusters.

What if we would transfer the best AI Engineering practices of LLMs to the domain of audio signal processing, without the LLMs themselves?

Join us for a free-for-members tasting session that introduces participants to the core ingredients of modern AI audio engineering, starting with Data, Models, and Algorithmic Code, and expanding toward the full lifecycle management used in production environments.



ANTICIPATED QUESTIONS

OBJECTION 1

“Where’s the empirical evidence?”

- This talk proposes an architectural framework — quantitative benchmarking is the natural next phase
- Tanbur/kantele demonstrator has published measurement-derived priors with validated nonlinear waveguide calibration (Şimşekli et al., EURASIP 2011)
- XR demonstrator (SIVE toolkit, open-source) has been used in peer-reviewed perceptual studies
- We invite collaborative benchmarking across the AES community

OBJECTION 2

“How do you handle real-time latency?”

- Strict architectural separation: DSP runs the synthesis loop at sub-5 ms latency: Pruning (Sudholdt et al, 2023)
- Neural components operate asynchronously on parameter inference — not in the audio callback
- SIVE toolkit demonstrates this split on commodity VR hardware (Meta Quest) with stable performance
- Neural surrogates handle offline tasks (transfer learning, residual modeling) — not sample-by-sample synthesis

OBJECTION 3

“Isn’t this just MLOps for acoustics?”

- Standard MLOps assumes stationary data and objective metrics — audio has neither
- Integrates simulation intelligence (9 motifs, Lavin et al.) with perceptual-in-the-loop evaluation — absent from generic MLOps
- Domain-specific: measurement-constrained model search, physics-informed priors, XR ecological evaluation
- Perceptual quality is subjective, context-dependent, and performer-specific — no parallel in standard ML pipelines

JAM WITH AI

AutoMLOps

The next evolution of MLOps

AutoResearch

overnight experiment loop
agent + scalar metric

+

MLOps

feature store + pipelines
eval harness + monitoring

=

AutoMLOps

harness engineering
for production ML