

NEURIPS 2026 · CREATIVE AI TRACK

Windchime: an Audiovisual Installation with Audio-Language Model implementations

Trent Eriksen[1] · Celeste Betancur Gutierrez[2]

[1] Independent Researcher, San Francisco, USA
[2] CCRMA, Stanford University, Stanford, USA

NEURAL INFORMATION PROCESSING SYSTEMS

CCRMA

GRAY AREA

The work

A physical algorithm, recombined by spoken input

A wind chime is a physical algorithm.

Its tubular bells are fixed, and contact with the wind produces unbounded outcomes as an evolving melody.

Windchime extends this to a corpus of **artist-authored** audiovisual material, recombined by **spoken input**: music from **406 audio stems** and animation from **135 Three.js scenes**.



Two visitors at the Gray Area Cultural Incubator Showcase, San Francisco, August 2026.

The installation

Screen, microphone, monome grid and arc, vintage hi-fi





Gray Area Cultural Incubator Showcase, San Francisco, 5 to 7 August 2026. Right: a visitor at the monome grid while a second scene plays.

ONE VISITOR SESSION

intro
narrated

→

4 spoken inputs
15 s each, monome free

→

recap
latent-space animations

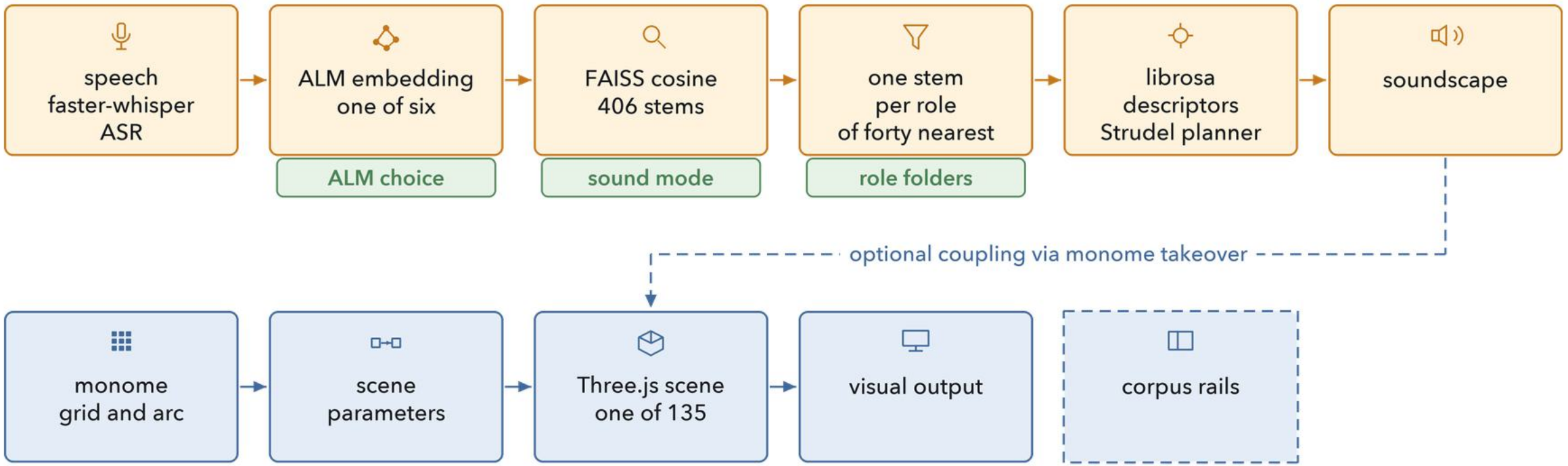
→

reset
next visitor

Spoken inputs range from descriptive to imagistic, performative and open phrasings. The installation then resets for the next visitor.

How it works

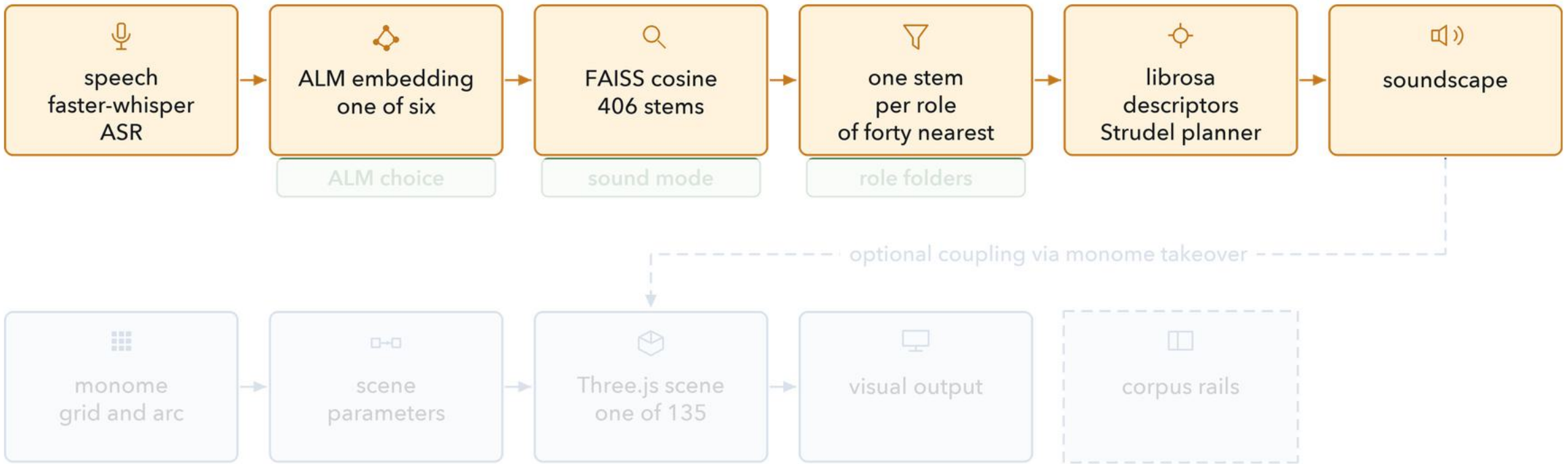
Machine learning operates at three points in the system



Amber: the audio path from speech to soundscape. Green tabs: the three choices the artist makes. Blue: the tactile and visual path controlled by monome hardware. Dashed: the optional coupling and the corpus rails.

How it works

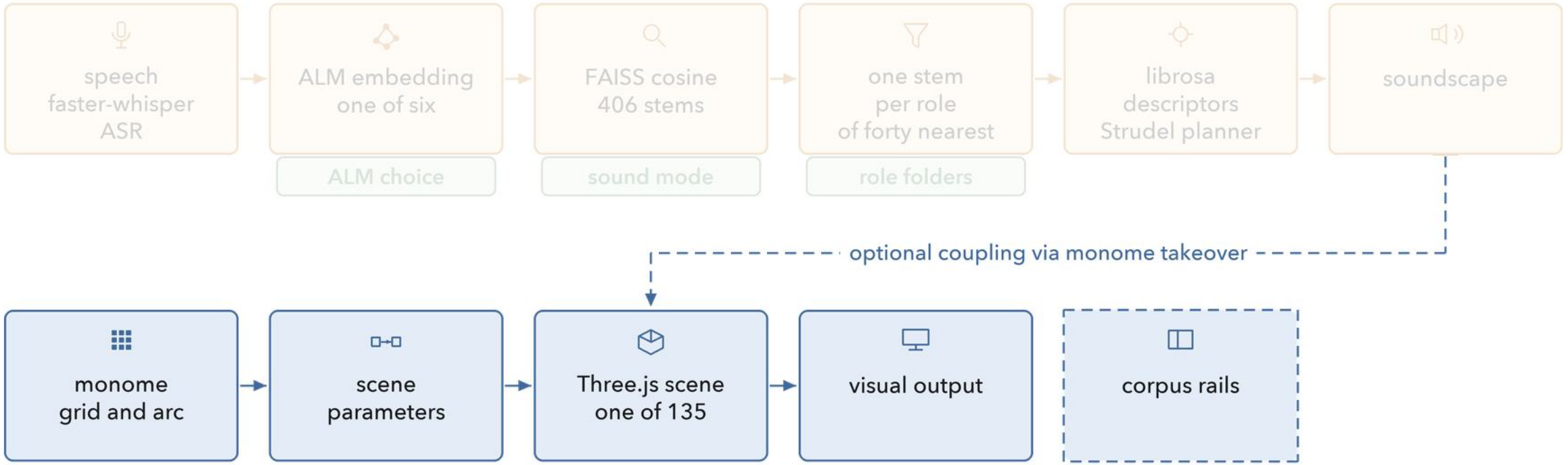
The audio path: speech to soundscape



faster-whisper transcribes the prompt. The selected ALM embeds it, FAISS ranks 406 stems by cosine similarity, and at most one stem per instrument role is admitted among the forty nearest. librosa descriptors inform a planner: hosted language model, then local model, then a fixed template.

How it works

The tactile and visual path



Each spoken input also retrieves one of 135 Three.js scenes. Every scene has its own monome grid and arc mappings with LED feedback. A visitor can couple audio and animation through a monome takeover.

Sound modes and corpus

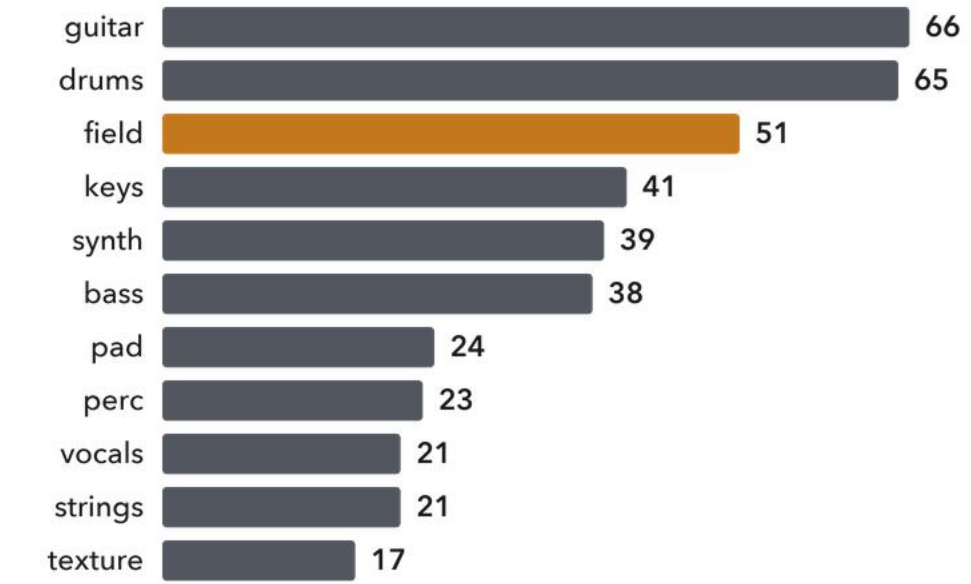
Which part of each retrieved stem becomes audible

FOUR SOUND MODES, IN PUBLIC PLAY ORDER

Peak	starts at the most energetic 30 seconds of the stem
Soundscape	plays from the beginning
Immediate	starts at the first audible section
Settled	starts at the earliest 30 seconds at 60 percent of peak energy

Stems vary in length and content, so the same retrieval sounds different in each mode.

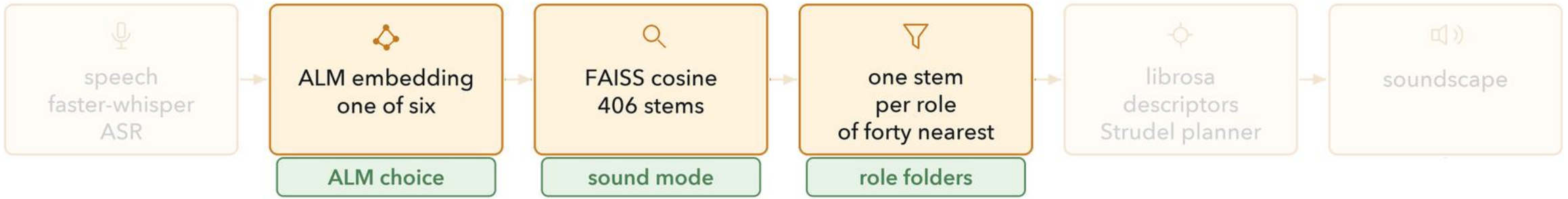
406 STEMS, ELEVEN INSTRUMENT ROLES



Field recordings from Japan, Korea, the European Union and the United States, in urban and rural settings.

Agency

Choosing the model is an artistic decision



ALM

maps words to sounds. Six configurations: LAION-CLAP (default, re-indexed copy, music checkpoint), MS-CLAP 2022 and 2023, and the multilingual CLaMP 3.

Sound mode

decides which parts of the retrieved stems become audible.

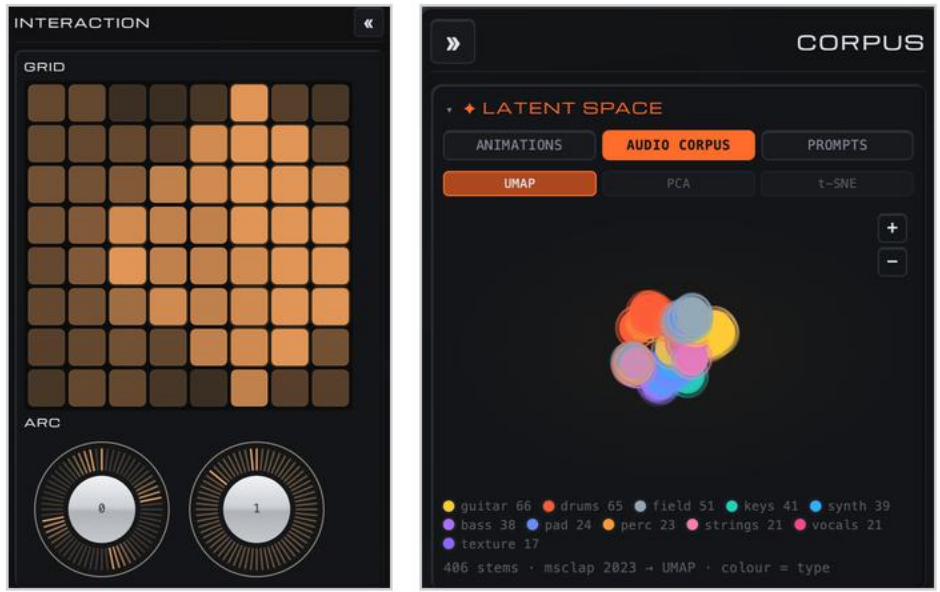
Role folders

the stem categories the selection policy draws from.

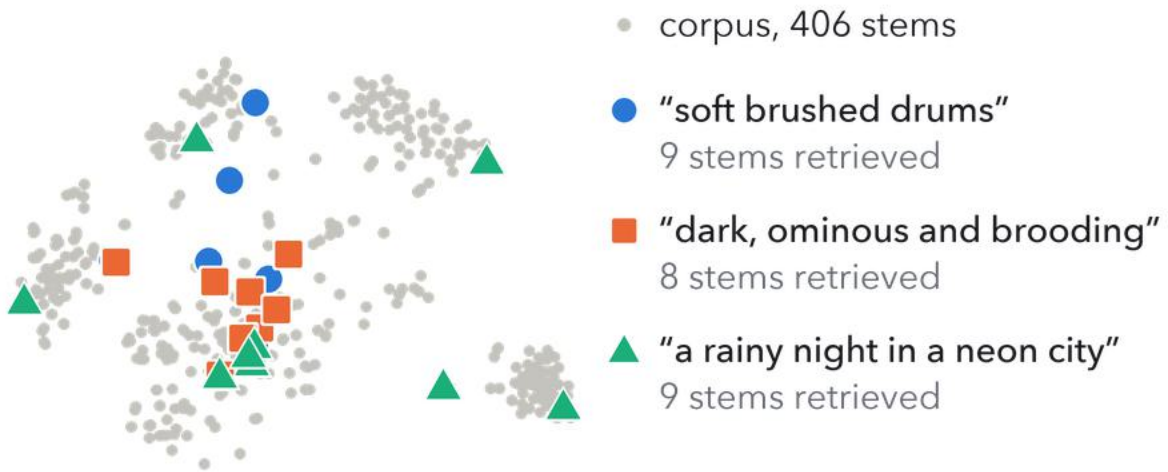
The choice of ALM conditions which parts of the corpus a prompt can reach, as some models retrieve more distinct stems than others. It takes some agency from both artist and visitor, and it produces surprising retrievals from voice input.

Legibility

Rails and latent maps show what the model is doing



Left: the interaction rail mirrors the monome hardware as a digital twin. Right: the corpus rail places the stems in the latent space of the active ALM, coloured by instrument role.



t-SNE of the 406 stem embeddings of the default LAION-CLAP model, with the stems three example prompts retrieve.

TAKEAWAY

In Windchime agency is intentionally distributed.

The artist authors the corpus and sets the configuration. The visitor initiates every retrieval, and the model settles the correspondence between spoken utterances and the artist's material.

Gray Area event page
grayarea.org/event/cultural-incubator-showcase-2026

Documentation
grashopr-888.github.io/current-projects

Thank you. Trent Eriksen and Celeste Betancur Gutierrez.

NEURAL INFORMATION PROCESSING SYSTEMS

GRAY AREA