Realtime Generation of Audible Textures Inspired by a Video Stream

Simone Mellace, Jerome Guzzi, Alessandro Giusti, Luca M. Gambardella
Dalle Molle Institute for Artificial Intelligence – Lugano, Switzerland

GitHub
git.io/fh1ng

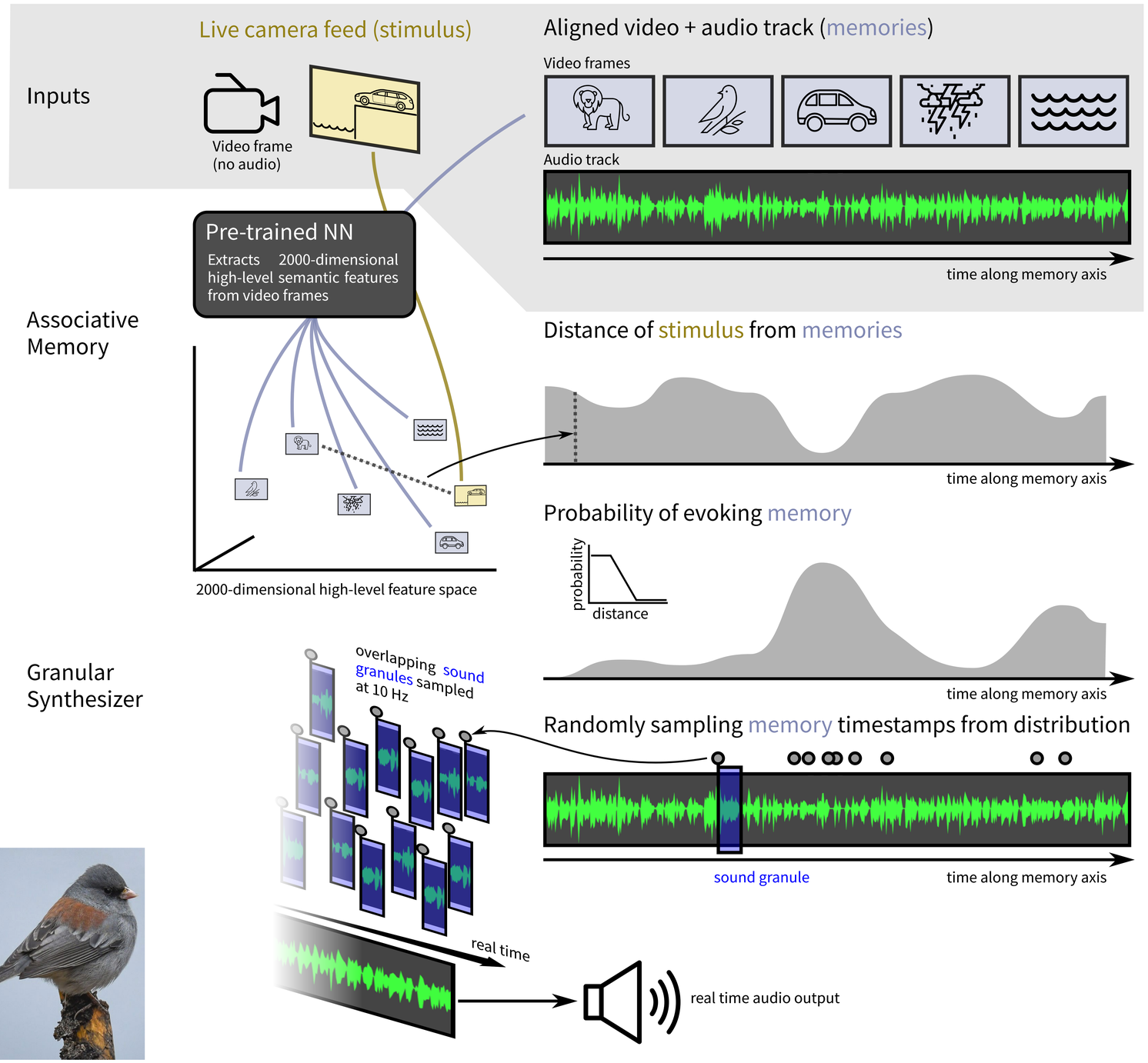
Abstract

We showcase a model to generate a soundscape from a camera stream in real time. The approach relies on a long training video with an associated meaningful audio track; a granular synthesizer generates a novel sound by randomly sampling and mixing audio data from such video, favoring timestamps whose frame is similar to the current camera frame; the semantic similarity between frames is computed by a pre-trained neural network.
The demo is interactive: a user points a mobile phone to different objects and hears how the generated sound changes.

Sound Generation

A granular synthesizer produces a continuous soundscape by randomly overlaying many short sound samples with a length between 1 and 1000 ms; each granule is too short to be perceived as an individual entity, but long enough to meaningfully contribute to the resulting sound.
Granules are sampled at random times from a source audio and played with soft attack and release transients, in order to better blend them in the overall texture. We use a granule length in the range between 100 and 1000 ms, and we start the playback of new granules at approximately 10-30 Hz (so that about 3 to 10 granules are overlayed at any given time): at these settings, the character of the resulting sound is pleasant and reminds the character of the audio from which the granules are sampled from, without the source being recognizable. **To affect the sound character, we manipulate in real time the probability distribution from which the synthesizer samples the starting points of the granules to be played next**.

Explore your environment! What do you hear?



This research was supported in part by the Swiss National Science Foundation through the National Centre of Competence in Research (NCCR).
Icons from the Noun Project: Sea by Daria Moskvina, Lion by Iconic, Bird by Laymik, Car by Vladimir Belochkin, videocamera by Ralf Schmitzer, family car by Setyo Ari Wibowo, loudspeaker by Stock Image Folio

robotics
Swiss National Centre of Competence in Research