

---

# LLMents I

Tina Tallon\*<sup>1</sup> and Ine Vanoeveren

<sup>1</sup>The Ohio State University [Columbus] – United States

## Abstract

question the ambiguities and limitations inherent in attempts to describe and represent music and other complex expressive sonic events using commonplace ontologies in audio classification systems and large language models. Through live audio processing and video generation, LLMents focuses on the ways in which music resists linguistic description, and questions the artistic utility of text-to-sound and text-to-music models. Each iteration of LLMents focuses on the overlap between live audio input and a specific musical instrument, with LLMents I focusing on the overlap in the expressive capabilities of both the flute and the human voice. An audio classifier is used to generate a textual description of the incoming audio signal in real time, which is then used to generate and modify elements of gesture-controlled live video such as color, texture, and shape through a combination of text-to-image models and stochastic systems. Not only is it difficult for many audio classification models to accurately identify both the physical sources of sounds and the actions that produce them with any degree of sophistication, the resulting slippage between action and object, between embodied sonic cognition and lexical fixity, yields an imprecision in image generation that ranges from humorous to grotesque. By focusing on the ways in which music resists the lexical descriptive framework at play in many of the generative tools being developed at present, we recenter the body as a site of knowledge production, storage, and transmission, particularly in artistic contexts.

**Keywords:** sound art, audio classification, installation art, generative AI, image generation, digital signal processing, large language models

---

\*Speaker