A Human-Annotated Dataset of Electroacoustic Sound Objects: Spectromorphological Labels for Generative Sound

Maurizio Berta

1 Scope of the project:

The aim of this thesis project is to create the foundation for a dataset targeted at experimental artists to augment creative possibilities in *generative artificial intelligence*. This is done in contrast with the tendency of creating datasets with the purpose of using the data for *Generative AI* that emulates certain styles of popular music, like *AudioLM* by Google[1].

1.1 A Human-Annotated Dataset:

A dataset is a collection of data, and in this case, labeled data. This means that for each data point, there will be a little piece of information describing what is in that data. In the context of Machine Learning, this is known as supervised learning. To teach a machine how to recognize the sound of a bird we first need to provide information about what birds sound like, similar to teaching a baby, usually done by giving the machine an audio of the bird plus a text file containing the information "bird".

1.2 Electroacoustic Sound Objects:

The project requires each participant to submit a number of sound objects with lengths between 5 and 10 seconds (following Pierre Schaeffer's terms). Using the given app, you will describe some characteristics of these sounds. To create a proper dataset, a significant amount of sounds is needed, aiming for 2000 sounds. That means that ideally each user is going to insert and label around 20/30 brief segments of audio.
As a comparison, MusicLM[1] by Google contains 5291 samples of 10 seconds each, but their approach involved downloading segments of YouTube music videos without granting attribution to the creators. In this project, the audio files must be created by the users who are annotating them. This aspect is crucial, as each audio segment and the dataset as a whole will be released under a Creative Commons CC-By licence. All contributors will be attributed on a page, following the guidelines provided by Creative Commons[2].

2 Usage of the app:

Upon opening the provided zip file, you will find a folder containing a max patcher called "labelling app" and some other folders. Open the max patcher to access the interface, which is divided into three sections. The first section allows you to set your username and load the audio file you want to tag. Then, you can turn on the audio and proceed to the next section.
In section 2, you are provided with the waveform of the sound that you loaded, and three subsections for labeling the data. Starting from the the left it is required to type a caption describing the audio in its integrity. Moving to the right there are checkboxes where you can choose some labels describing acoustic and spectromorphological characteristics, I kindly ask you to provide at least the morphological informations. Lastly there is text field to manually insert labels that the user consider necessary.
The last section is used for accepting the terms and conditions and for submitting the audio and labels to a cloud folder.

3 Licensing:

Participants will be requested to license their sound contributions under the CC-By licence, as mentioned before. This licence allows sharing, copying, and redistributing the material in any medium or format. Additionally, it permits remixing, transforming, and building upon the material for any purpose, even commercially. However, it requires appropriate credit to be given, providing a link to the licence, and indicating if any changes were made. The decision not to include a Non-Commercial (NC) licence was made to enable the release of music using the dataset, allowing for commercial use with proper attribution.

References

[1] "Music-lm research webpage `https://google-research.github.io/seanet/musiclm/examples/`," 2023.

[2] "Creative commons attribution rules https://creativecommons.org/use-remix/attribution/," 2023.