Timbral Analysis

Alec Krsek

Oregon State University

soundbendor lab

Abstract

This research project establishes a tool for computing timbral descriptive words for any sound file. Currently there are no existing datasets with timbral labels. This project aims to automatically label audio datasets with timbral labels so that we can train a machine learning network to automatically label timbre directly from an audio file. For computing the timbral descriptors, we created a Python tool aided by ChatGPT that extracts 15 different acoustic parameters including spectral centroid, harmonics-to-noise ratio (HNR), pitch variations, and temporal envelope features. These quantitative measurements are used to map sounds into meaningful textual descriptors that are organized in five general categories of timbre: tonal, dynamic, rhythmic, textural, and brightness characteristics.

Data

The data used for our work is the `ESC-50` dataset. This dataset is a sample of 2,000 environment audio files. Each audio file is tagged with a category and they range from airplane sounds to people coughing.

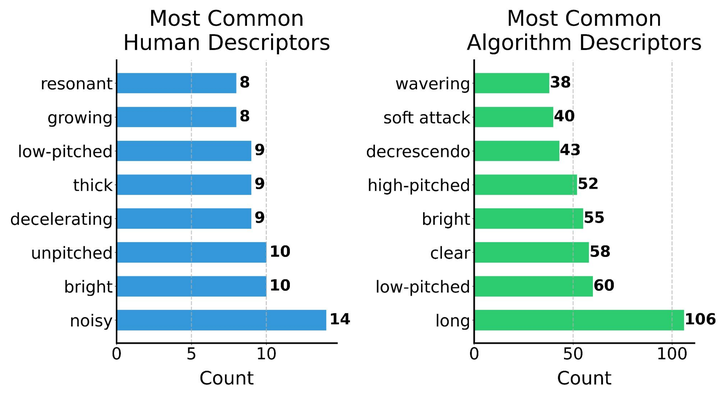

Fig. 1: Comparison of human and algorithmic descriptors.

Methods

Our system uses ChatGPT to analyze audio and find thresholds for audio features that differentiate between timbral descriptors. This process extracts 15 acoustic features using the library tools `librosa` and `Praat`:

- Spectral: centroid, flux, alpha ratio , ...
- Harmonic: HNR, jitter, shimmer, ...
- Temporal: duration, RMS, tempo var., ...

These features are mapped to linguistic descriptors using adaptive thresholds.

Example

high spectral centroid (>4000Hz) with low HNR (<-3)	**strident**
moderate jitter (>0.06) with pitch slope (>3)	**wavering**
low shimmer (<0.15) with low spectral flux (<1.2)	**clean**

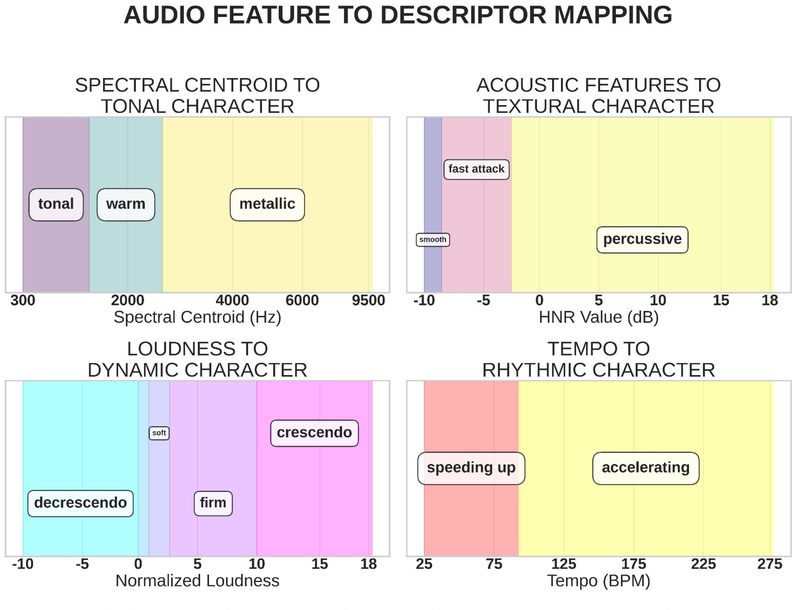

Fig. 2: Illustration mapping audio feature to descriptors.

User Study

Our user study received only 254 responses, in our attempt to label 200 stratified audio files. Each user was asked to listen to a random sound file and select the best choice from a dropdown menu showing descriptive words for an acoustic feature.

Results

Given that our user study was small and our response rate very low, coupled by the fact that many respondents may not have a background in music, we were only able to achieve a 19.9% weighted match agreement rate and 18.5% raw match rate. As future work, we will design a larger, rigorous study.

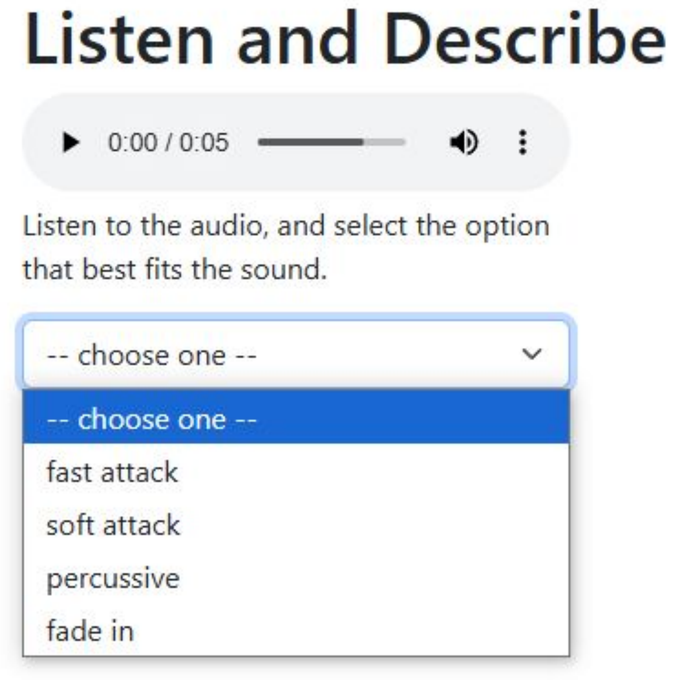

Fig. 3: Example illustrating the task of the user study.