Sound Dataset Annotation Tool: *Crowdsourcing Timbre and Emotion Labels*

Andrew Young Patrick J. Donnelly

Oregon State University

Motivation

Machine learning models that recognize the **timbre** and **emotion** of music need large datasets of labeled audio. Expert labeling is slow and costly, and words for sound are subjective, so one listener's labels are not enough.

Our idea: a website where many volunteers listen to short music snippets and pick the words that describe them. **Game-like** rewards keep them coming back, and every answer becomes training data.

Goal: a General Study Tool

Rather than one fixed experiment, the tool lets researchers set up **separate studies**:

- **Upload** royalty-free music snippets
- **Choose** the descriptors and how they are presented
- **Export** every annotation for analysis (CSV)

Listeners use a mobile-friendly website; a backend serves the samples and stores the annotations.

Descriptors

The starting vocabulary comes from Carron et al. [1], who reviewed 45 studies of how people describe sounds. Paired opposites make quick, comparable choices.

Basic	high/low, loud/weak, noisy/tonal, short/long, natural/artificial, near/far, dynamic
Temporal	continuous/discontinuous, constant/fluctuating, ascending/descending, crescendo/decrescendo, slow/fast attack
Timbre	mate/resonant, rough/smooth, bright/dull, nasal, rich, round, warm, metallic, strident

Studies can add their own descriptors, including **mood and emotion** words.

Six Annotation Styles

Showing every descriptor at once would overwhelm listeners, so each study picks how descriptors are offered. Varying the style also keeps the task interesting.



Best Match

Pick the one descriptor that fits best.



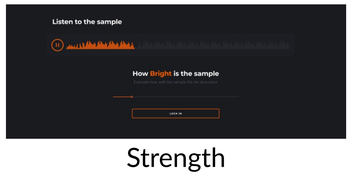
Pick All

Select every descriptor that fits.



Compare

Which of two samples fits one descriptor?



Strength

Rate how well one descriptor fits.



Opposites

Choose between a paired opposite.

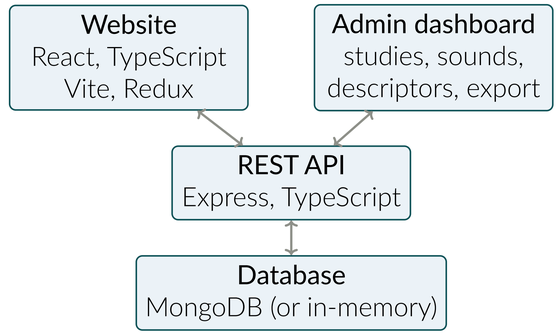


Forced Match

Drag four descriptors onto four samples.

Studies also choose **when to ask**: after the whole sample, after each musical section, or at any time while listening.

System



Built in Summer 2024

The minimum viable product works end to end:

- Listeners **register and log in**
- Each study serves a **random sample**, plays it and offers descriptors in its chosen style
- **Annotations are stored** and researchers can **export them as CSV**
- Admins browse studies, sounds, users and annotations, and upload sounds

Next Steps

- Detail and edit pages for studies, sounds and descriptors
- Per-study intervals and profile pages with progress
- **Gamification**: levels and rewards for annotating
- Stretch goal: play MIDI files

Daniel Lounsbury continued the tool in 2025.

References

[1] Maxime Carron, Thomas Rotureau, Françoise Dubois, Nicolas Misdariis, and Patrick Susini. **Speaking about sounds: a tool for communication on sound features.** *Journal of Design Research*, 2017.