SoundMap

A Visualization Tool to Explore Multi-Attribute Sound Data

CPSC 547 Project
Elizabeth Reid, Mifta Sintaha, Nichole Boufford

Motivation

Standard techniques for audio visualization do not address several issues:

- Comparison of many sound files
- Non-audio attributes
- Support for novice users

SoundMap allows users to explore multi-attribute sounds files simultaneously
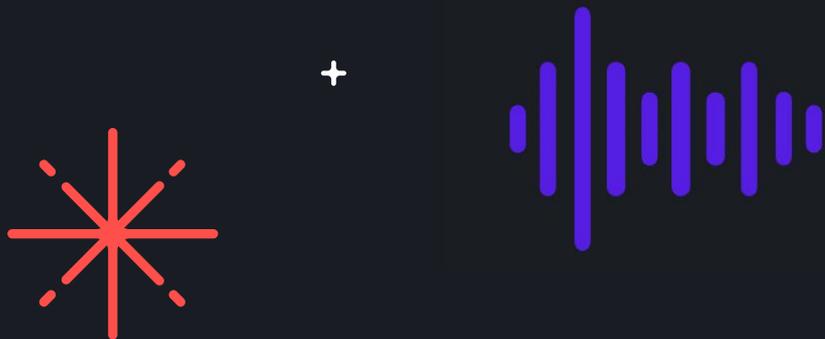
MULTI-FACETED

Visualize both audio and non-audio attributes

SCALABLE

Data from many sounds files can be visualized at once

USABLE

Designed for both novice and expert sound analysts

SoundMap allows users to explore multi-attribute sounds files simultaneously

3

Related Work

https://www.audacityteam.org/

https://dl.acm.org/doi/10.1145/1518701.1518896

https://www.cs.ubc.ca/~seifi/VibViz/main.html
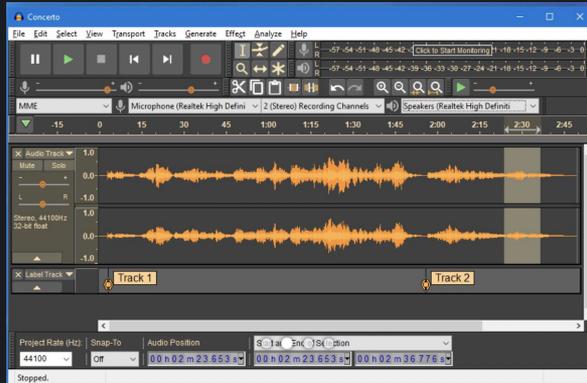
Sound File Visualization

Waveplots, spectrograms, and chromagrams are typically used
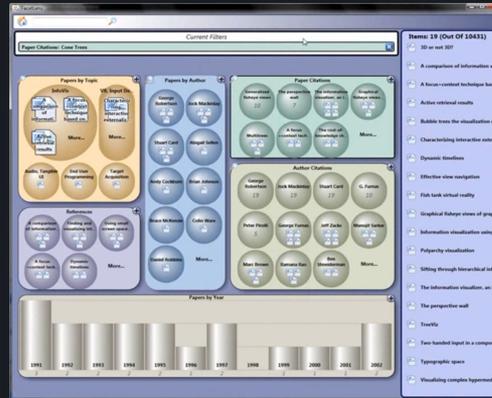
Faceted Browsing

FacetLens explores faceted data to compare and identify trends
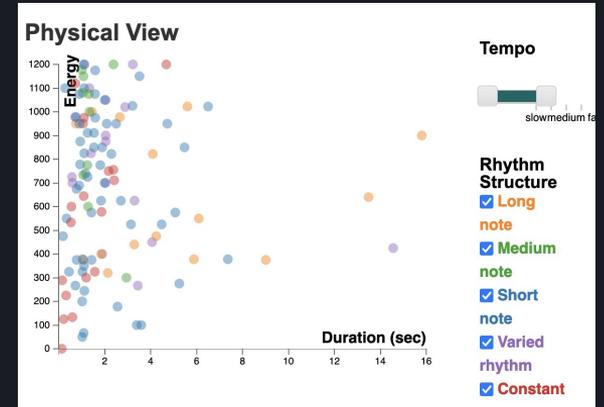
Waveform Visualization

Vibviz shows haptics visualization from waveforms, using faceted-browsing techniques

4

# Questions that SoundMap answers

- *Overview*: How do audio attributes like fundamental frequency or duration of each sound vary across different types of categories?

- *Detailed View*: How does the frequency/amplitude in a particular sound file spectrogram or waveplot look like?

- *Summary View*: What is the average duration or fundamental frequency of each sound file?

Domain - Audio Analysis

Traditional vis techniques:

Spectrograms, waveplots, chromagrams

Sound attributes: Fundamental

frequency, chroma features, MFCC's,

duration, and many more

Metadata: Contextual information, e.g.,

what or who was recorded, song genre,

language
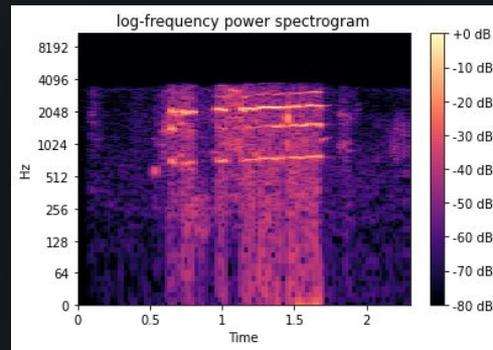
Spectrogram


Waveplot

Datasets

CatMeows

440 sound files of cat vocalizations with attributes such as sex, scenario stimulus, breed

FreeMusicArchive
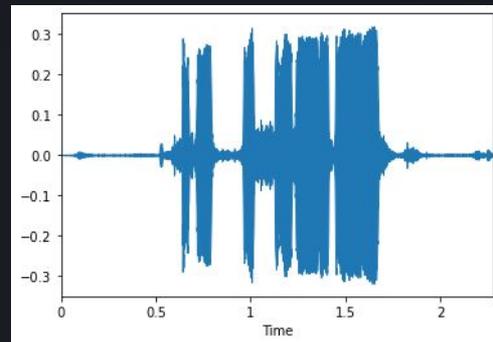
Over 100 000 songs with precomputed sound attributes, reduced to a subset of 98 songs for analysis
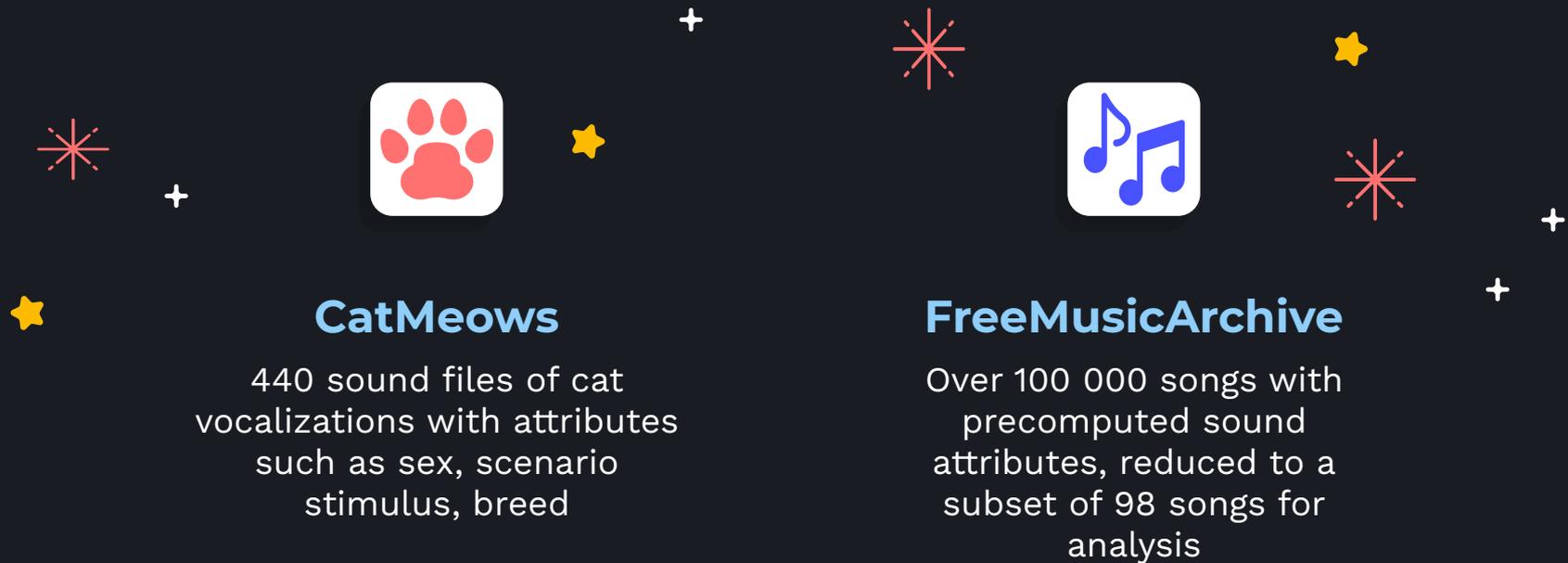
Data and Task Abstraction

	MeowAnalysis	SongAnalysis
Categorical	5	3
Quantitative	5	28

+File data and sample rate for each sound file

MeowAnalysis Dataset

TABLE I
CLASSIFICATION OF ATTRIBUTES IN THE MEOWANALYSIS DATASET.

Attribute	Description	Type	Items/Range
Cat ID	Unique ID corresponding to each cat	Categorical	21
Owner ID	Unique ID corresponding to each cat's owner	Categorical	12
Stimulus	Situation in which the meow was recorded. Situations included brushing, isolation in an unfamiliar environment, and waiting for food.	Categorical	3
Breed	Cat breed, either Maine Coon or European Shorthair	Categorical	2
Sex	Specifies whether the cat is male or female, and has been spayed or neutered	Categorical	4
Mean Fundamental	Mean of the estimated fundamental frequency in Hz (pitch) for the entire meow.	Quantitative	$124.589 - 1122.891$
Max Fundamental	Max value of the estimated fundamental frequency in Hz (pitch) for the meow.	Quantitative	$456.570 - 2205.000$
Min Fundamental	Min value of the estimated fundamental frequency in Hz (pitch) for the meow.	Quantitative	$21.554 - 26.957$
Peak Frequency	Estimated value for peak frequency in Hz (loudest frequency that occurred during the meow). Note this may occur at an overtone higher than the fundamental frequency.	Quantitative	$450.000 - 6960.000$
Duration	Total meow time in seconds, not including any silences before, after, or during the meow.	Quantitative	$0.006 - 1.847$
			Total: 440 meows

SongAnalysis Dataset

TABLE II
CLASSIFICATION OF ATTRIBUTES IN THE SONGANALYSIS DATASET.

Attribute	Description	Type	Items/Range
Album	Album the song belongs to. Some songs are single tracks, in which case the album is described as N/A.	Categorical	10
Artist	Song artist.	Categorical	7
Genre	Genre of the song. Note that some songs have not been assigned a genre, in which case the genre is described as unknown.	Categorical	7
Mean STFT Chroma	Chromagram values calculated over the short time Fourier transform for the entire song. Note that there are 12 separate chromagram attributes, one for each semitone note name (C, C#, D, D#, E, F, F#, G, G#, A, A#, B). The range for each attribute is approximately the same.	Quantitative	$0.1 - 0.6$
Mean Tonnetz	Projection of chromagram values onto a 6-dimensional basis, representing harmonic relationships (perfect fifth, minor third, and major third). There are therefore 6 separate tonnetz attributes, one for each basis element. The range for each attribute is approximately the same.	Quantitative	$-0.03 - 0.04$
Mean Spectral Bandwidth	Difference between the highest and lowest frequency in the spectrum.	Quantitative	$461 - 2020$
Mean Spectral Contrast	Contrast in energy between the top quantile (peak energy) to that of the bottom quantile (valley energy). This is calculated over 7 frequency bands (from low to high frequency). Each of the seven attributes has approximately the same range.	Quantitative	$13.0 - 27.0$
Mean Spectral Rolloff	The center frequency for a spectrogram bin where at least 85% of the energy of the spectrum is contained in this frequency bin and the bins below.	Quantitative	$449 - 3370$
Mean Zero Crossing Rate	The rate at which a signal (e.g., as a waveform) crosses the line $y = 0$.	Quantitative	$0.0172 - 0.0859$
			Total: 98 songs

Tasks: Analyze

- ***Audio and metadata analysis:*** The user can visualize all sound files with respect to a single audio attribute and a single metadata attribute.

- ***Detailed individual analysis:*** By selecting a single sound within the larger overview, the user can see more detailed audio information for the sound file.

Tasks: Search

- ***Browse an area:*** With a known location in mind with respect to either metadata or audio data, the user can investigate sounds within that area, opening detailed views if desired.
- ***Outliers:*** The user looks for unexpected outliers within their dataset.

Tasks: Query

- ***Visualize multiple audio attributes:*** Users can compare multiple audio attributes at the same time, and track the location of a single sound or group of sound files across all attribute displays.

- ***Visualize multiple metadata attributes:*** Users can compare multiple metadata attributes at the same time, with respect to one or more audio attributes.

- ***Filter:*** Users may filter both metadata and audio attributes.
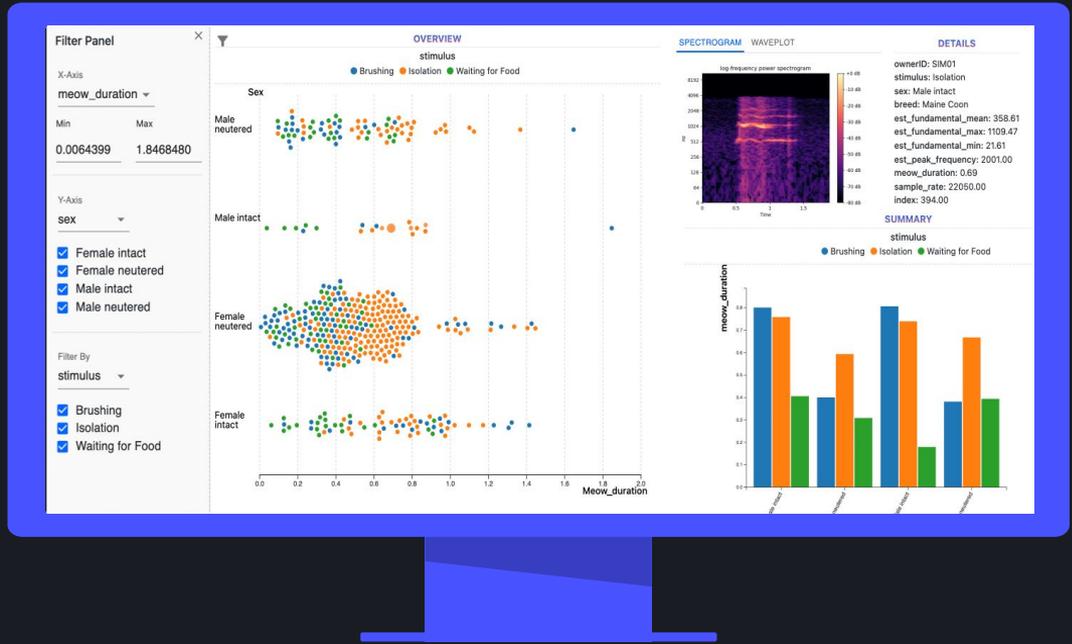
Solution

Overview
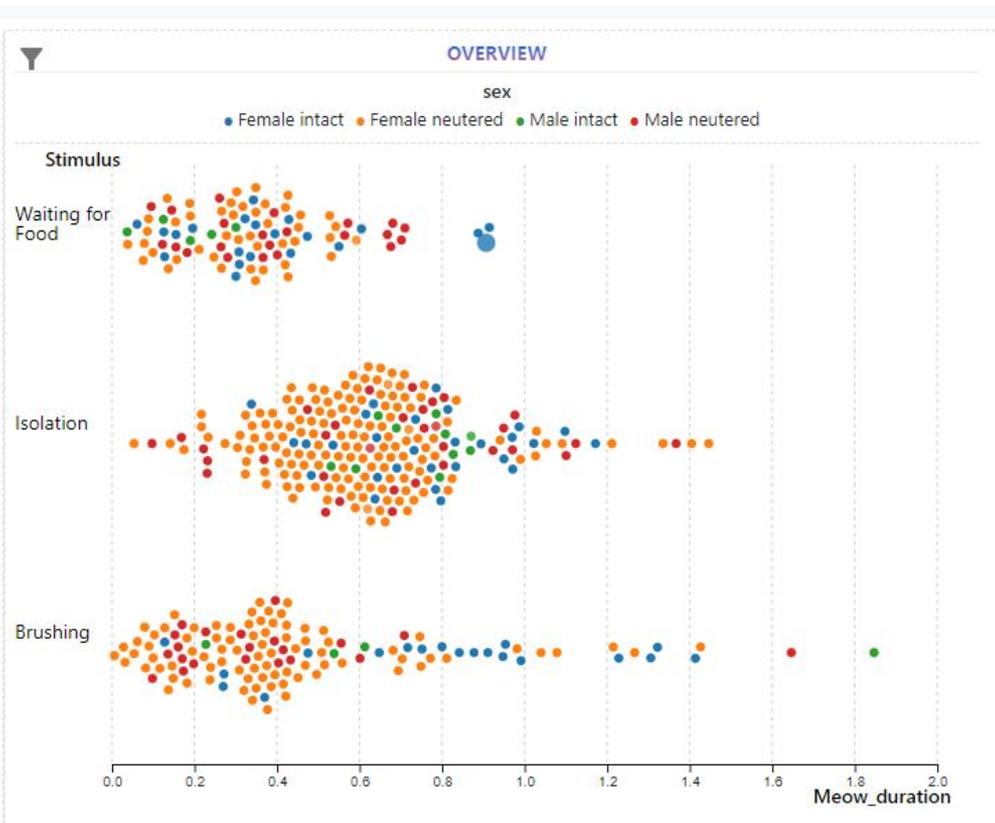
Key Question: How do audio attributes like fundamental frequency or duration of each sound vary across different types of categories?

Supported Tasks:
- Audio and metadata analysis
- Browse an area
- Outliers
- Filter

Summary View

SUMMARY

sex

● Female intact ● Female neutered ● Male intact ● Male neutered
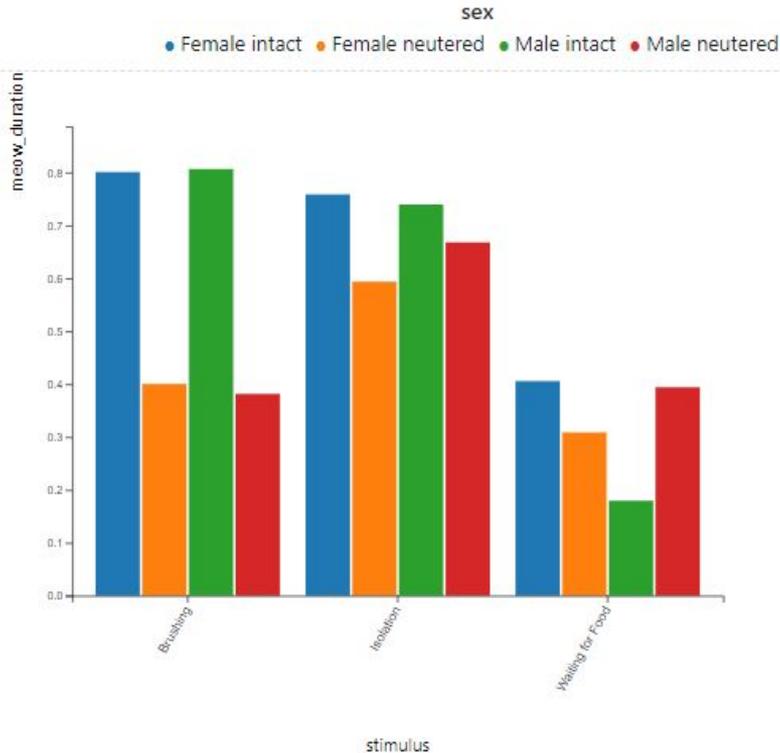
Key Question: What is the average duration or fundamental frequency of each sound file?

Supported Tasks:
- Audio and metadata analysis

Detailed View

SPECTROGRAM   WAVEPLOT

DETAILS

log-frequency power spectrogram
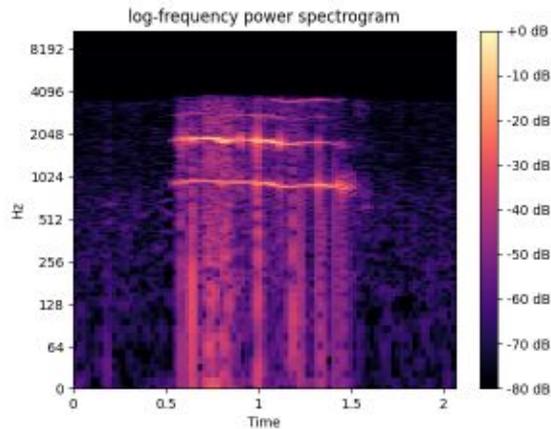
catID: WHO01
ownerID: SIM01
stimulus: Waiting for Food
sex: Female intact
breed: Maine Coon
est_fundamental_mean: 586.63
est_fundamental_max: 1932.55
est_fundamental_min: 22.28
est_peak_frequency: 3829.00
meow_duration: 0.91
sample_rate: 22050.00

*Key Question:* How does the frequency/amplitude in a particular sound file spectrogram or waveplot look like?

Supported Tasks:
- Detailed individual analysis

Multiple/Compare Views
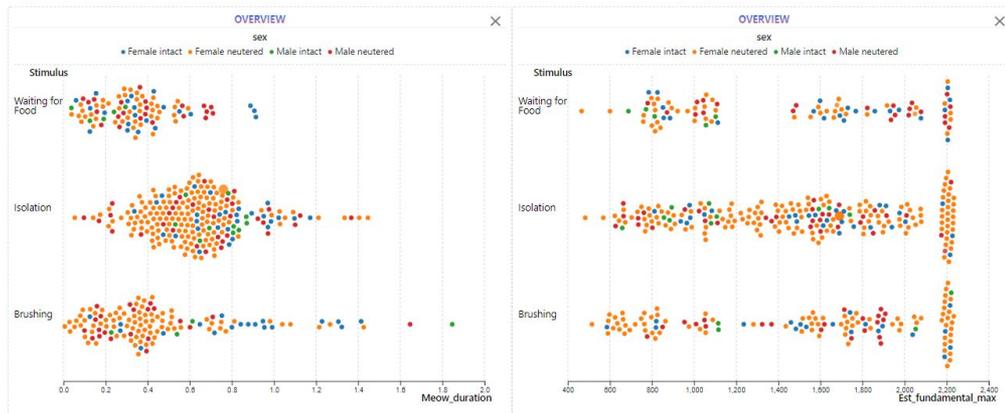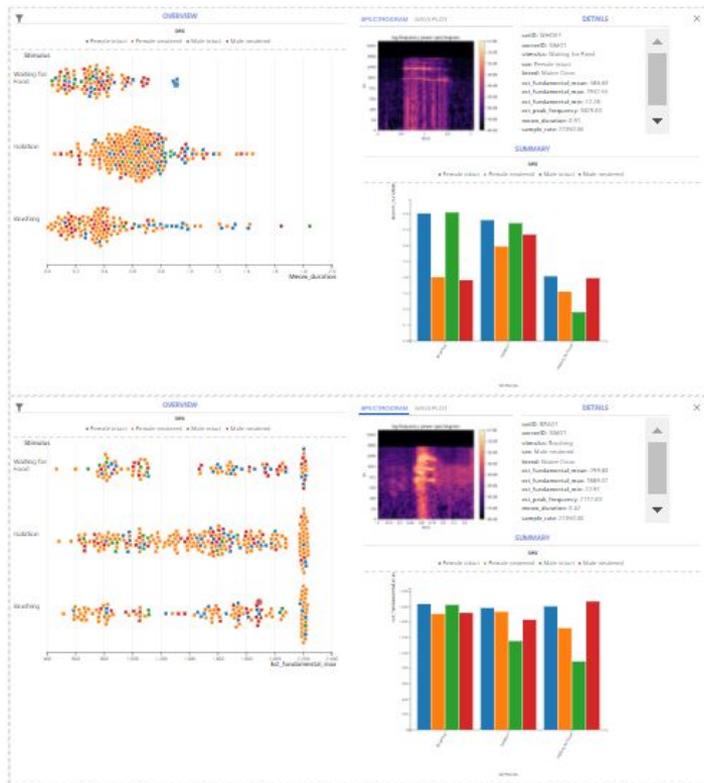
Supported Tasks:
- Visualize multiple audio attributes
- Visualize multiple metadata attributes

18

Implementation

Client / React JS
Frontend framework

Server / Flask
Backend framework

D3 (JavaScript)
Chart library

Librosa (Python)
Data wrangling and sound charts
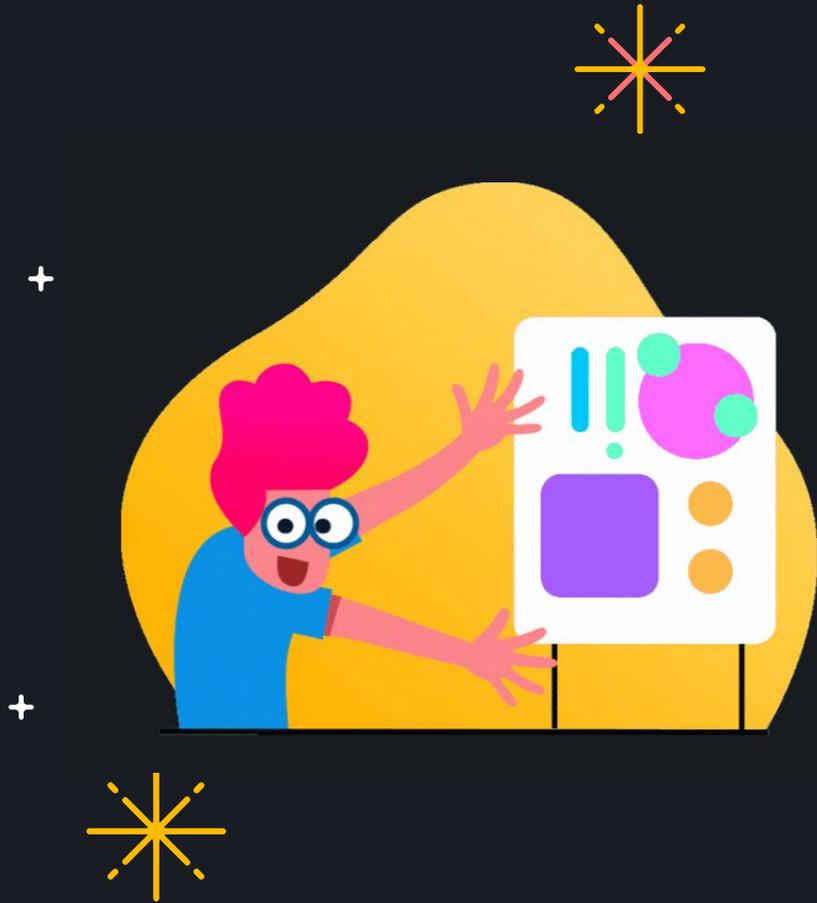
Demo

Limitations
+
Future Work

Interactivity

Audio plots with zoom or pan interactivity

Audio Preview

Ability to glimpse into the audio file through quick play

Data processing

Ability to preprocess within the tool

High Cardinality

Bar charts cannot handle high cardinality. Use filters or pan/zoom

Questions?