Datastore Design for Analysis of Police Broadcast Audio at Scale

Ayah Ahmad[1], Christopher Graziul[2] (Advisor), Margaret Beale Spencer[2] (Advisor)

[1]University of California, Berkeley [2]University of Chicago

National Institute on Minority Health and Health Disparities
Award Number R01MD015064

Abstract

With a growing desire to understand police interactions with civilians, a large corpus of data has remained excluded from analysis—police broadcast audio. With the intent of performing Speech Emotion Recognition (SER) on this audio, to characterize stress response via the emotions expressed in police communications, features must be extracted, clustered on, and fed into a machine learning model. Due to these various streams of input, we sought to create a database to make audio files available for easy interoperability with statistical methods for an unbiased large-scale analysis of police broadcast audio for SER.

Data

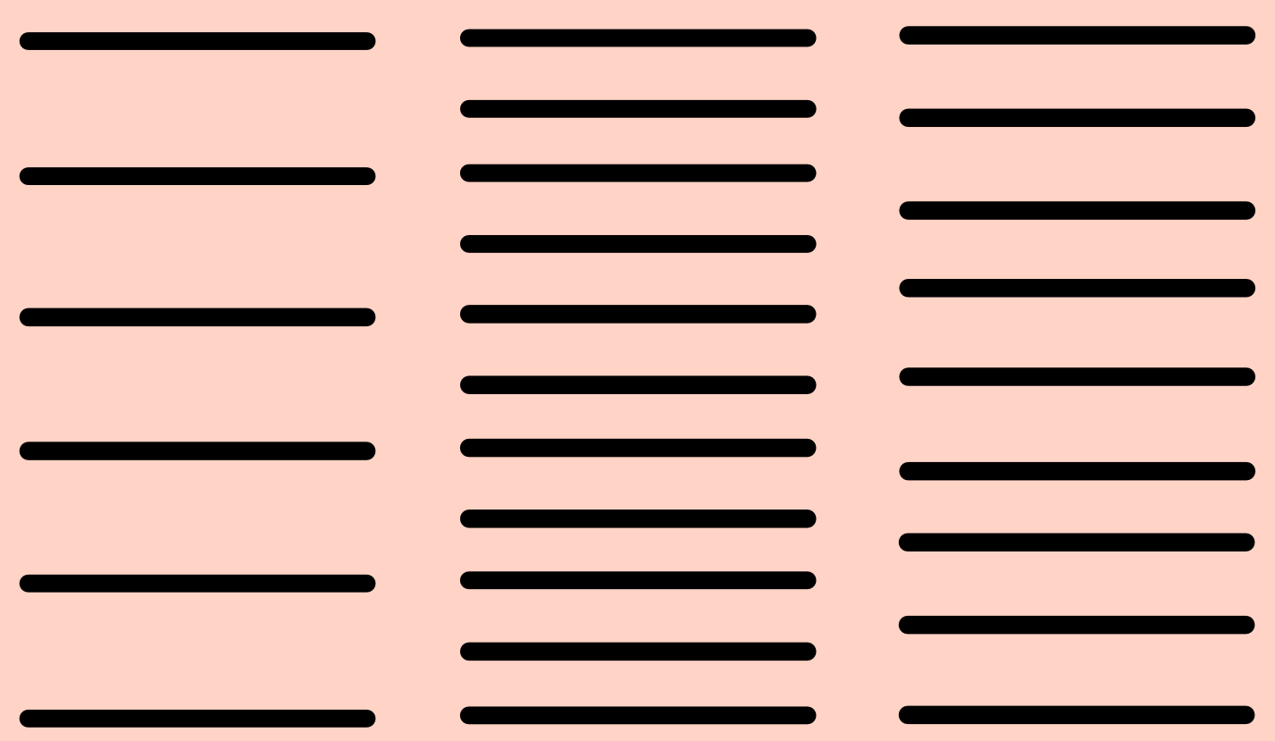
- Public archive of Chicago Police Department (CPD) Broadcast Police Communications (BPC)
- ~160,000 MP3 files, each ~30 minutes long and ~3.5 MB
    - ~4.8 million minutes (~80,000 hours) overall
    - ~560 GB overall
- Each MP3 file has two types of metadata
    - Extracted from the file name: dispatch zone, date, and timing
    - Extracted using Voice Activity Detection (VAD)—non-silent slices of audio, aggregate time of non-silent audio
- Each date, in a particular Zone, also has associated VAD metadata—completeness of data, aggregate time, total silence, silent files (boolean)

Challenges

Scale

- 6.4 trillion data points for all of the Raw audio files
- 65 billion data points for all Silence-removed audio files
- 878 billion data points for all of the Extracted features (GeMAPS & Praat-Parselmouth)
- 800,000 data points for all Metadata

R
S
E



Temporality

- Storage of temporal and non-temporal data
- Varying windowing functions and time sampling across applications used for extracting different features
- Features extracted at different time steps within individual audio files

Silence

- Silence of files can make intended clusters obsolete (e.g. clustering can lead to variance explained by one dimension--detecting silence vs. non-silence)

Database Design & Implementation

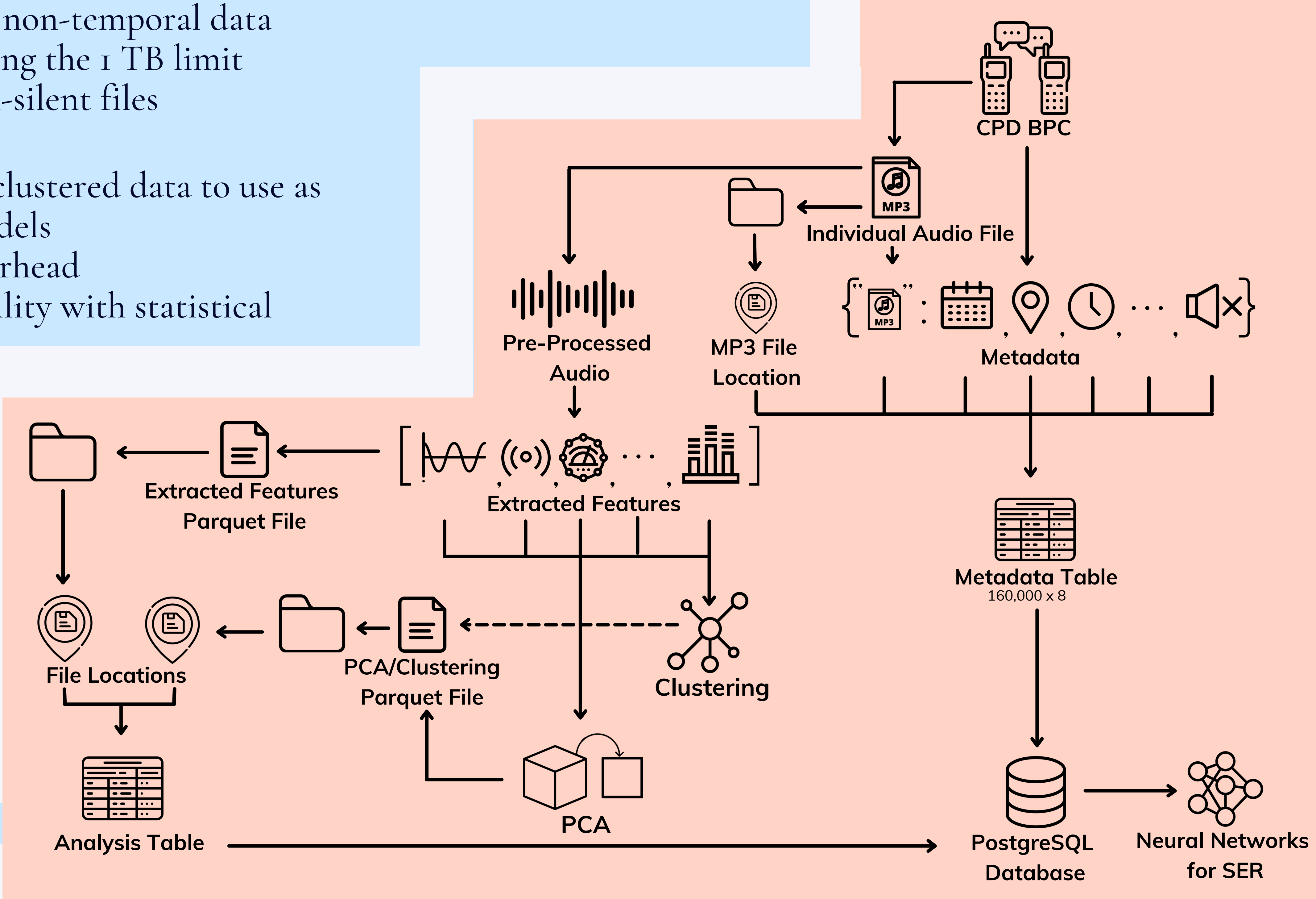
DBMS Desired Features: extensibility, scalability, ACID-compliance, high levels of concurrency
⟶ PostgreSQL database, 1 TB constraint

Design

- Datastore, with raw and extracted features stored in specified locations
- Each location would be added to, and accessible from, the database
- Would allow for storage of:
    - Temporal & non-temporal data
    - Data exceeding the 1 TB limit
    - Silent & non-silent files

Outcomes

- Organization of clustered data to use as input to SER models
- Low memory overhead
- Easy interoperability with statistical methods



Flow of data, informing database design, for the enablement of SER

Conclusion

- Created a framework that enabled easy interoperability with statistical methods for an unbiased large-scale analysis of police broadcast audio for SER
- Completed large-scale pre-processing and feature extraction using datastore framework
- Completed 5-dimensional Principal Component Analysis (PCA) on GeMAPS features

Related Literature

Akçay, Berkehan & Oguz, Kaya. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication. 116. 10.1016/j.specom.2019.12.001.

M. Chen, X. He, J. Yang and H. Zhang, "3-D Convolutional Recurrent Neural Networks With Attention Model for Speech Emotion Recognition," in IEEE Signal Processing Letters, vol. 25, no. 10, pp. 1440-1444, Oct. 2018, doi: 10.1109/LSP.2018.2860246.