

# Beni Airlines

---

Incorporation of speech emotion recognition in client support calls

- Dynamic feedback
- Helps to list best practices
- Identify the level of satisfaction
- Clients free of lengthy questionnaires

**Gabriela Fichtner**

# RAVDESS dataset

---

One of the most complete databases

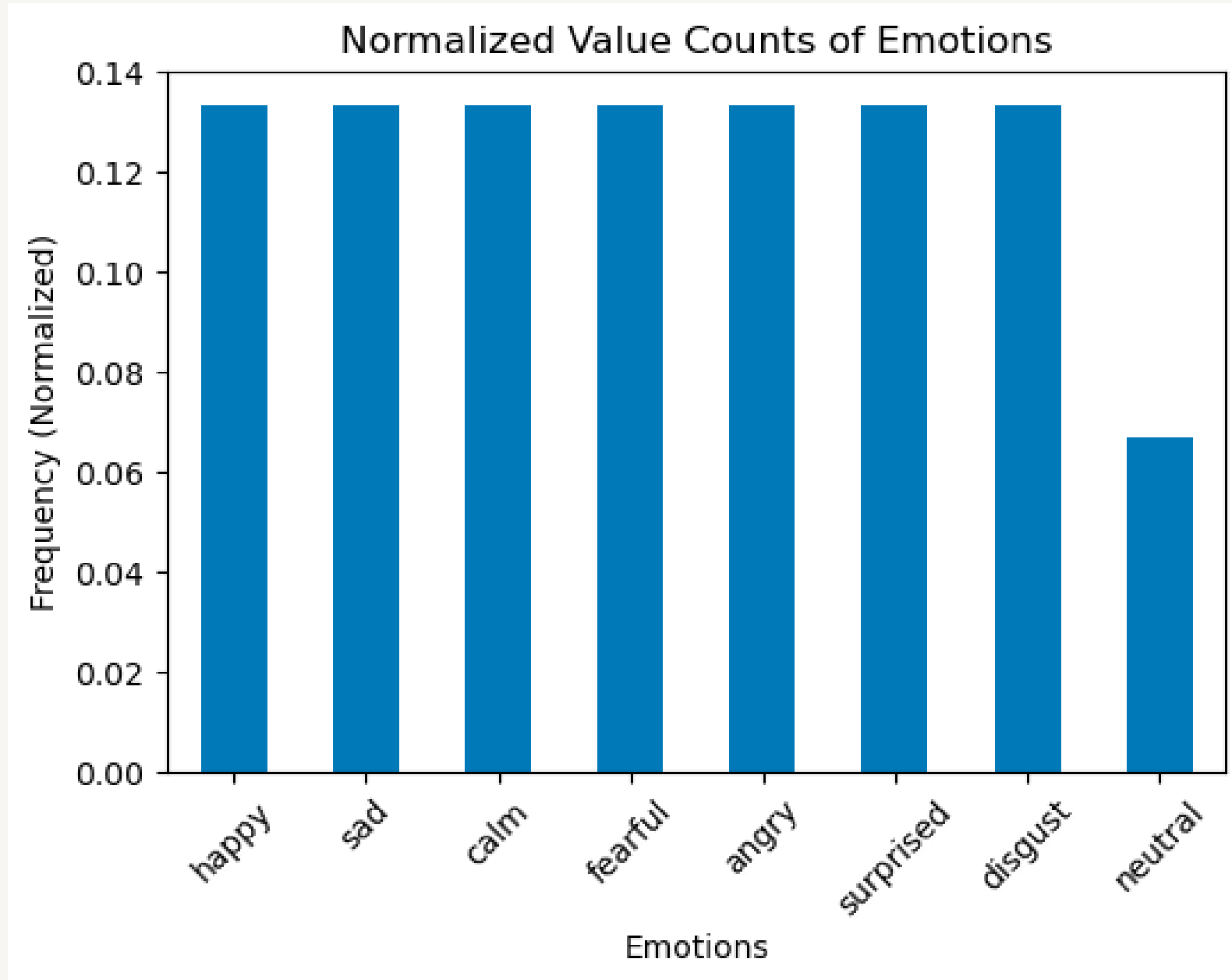
Source: [zenodo.org](https://zenodo.org/record/1258591/files)

24 actors x 60 audios = 1440 files

8 different emotions

2 statements

# Dataset



Baseline: 13.3%

## Measurements per second:

Sample Rate: 22050

## Length:

Longest Audio: 5.27 s (~116203 values)

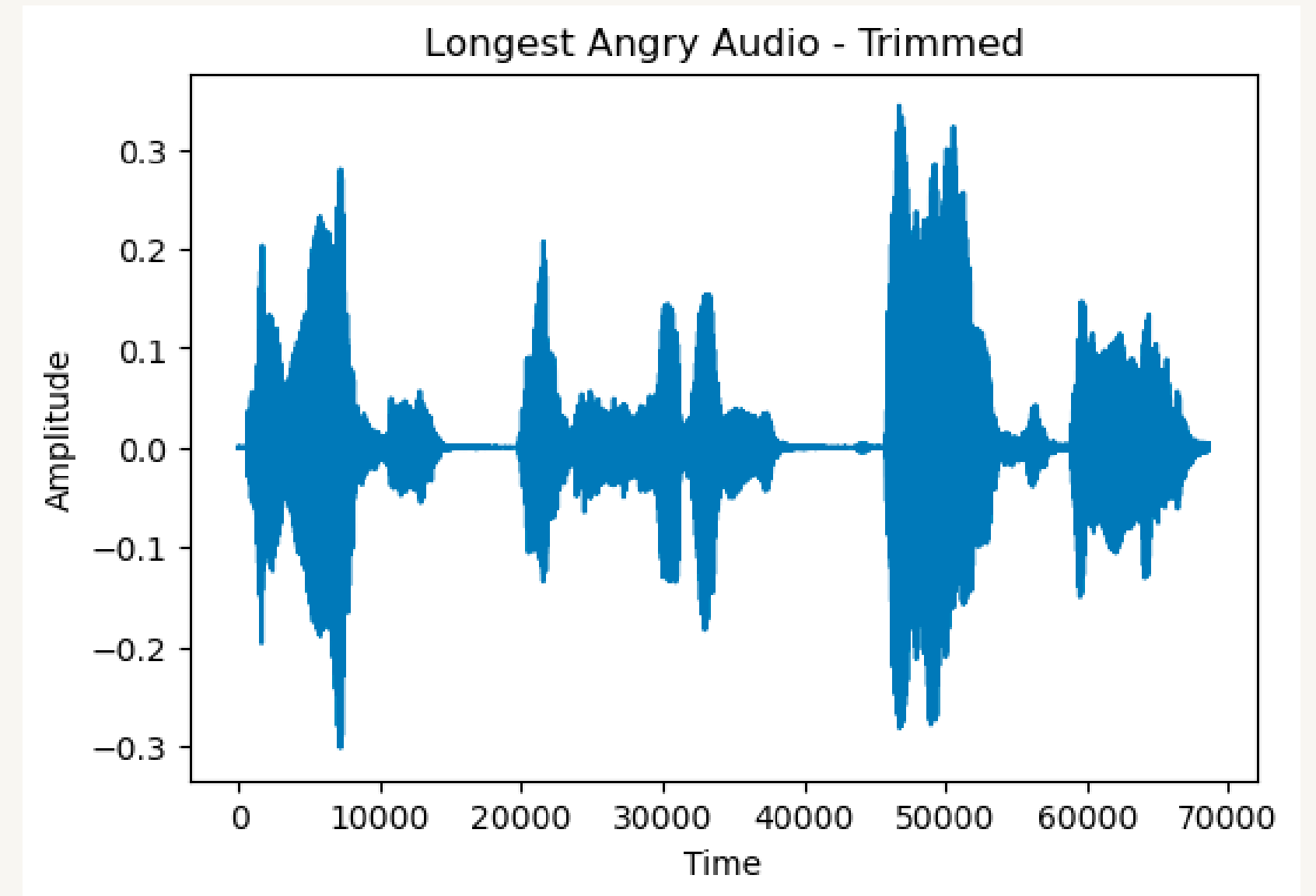
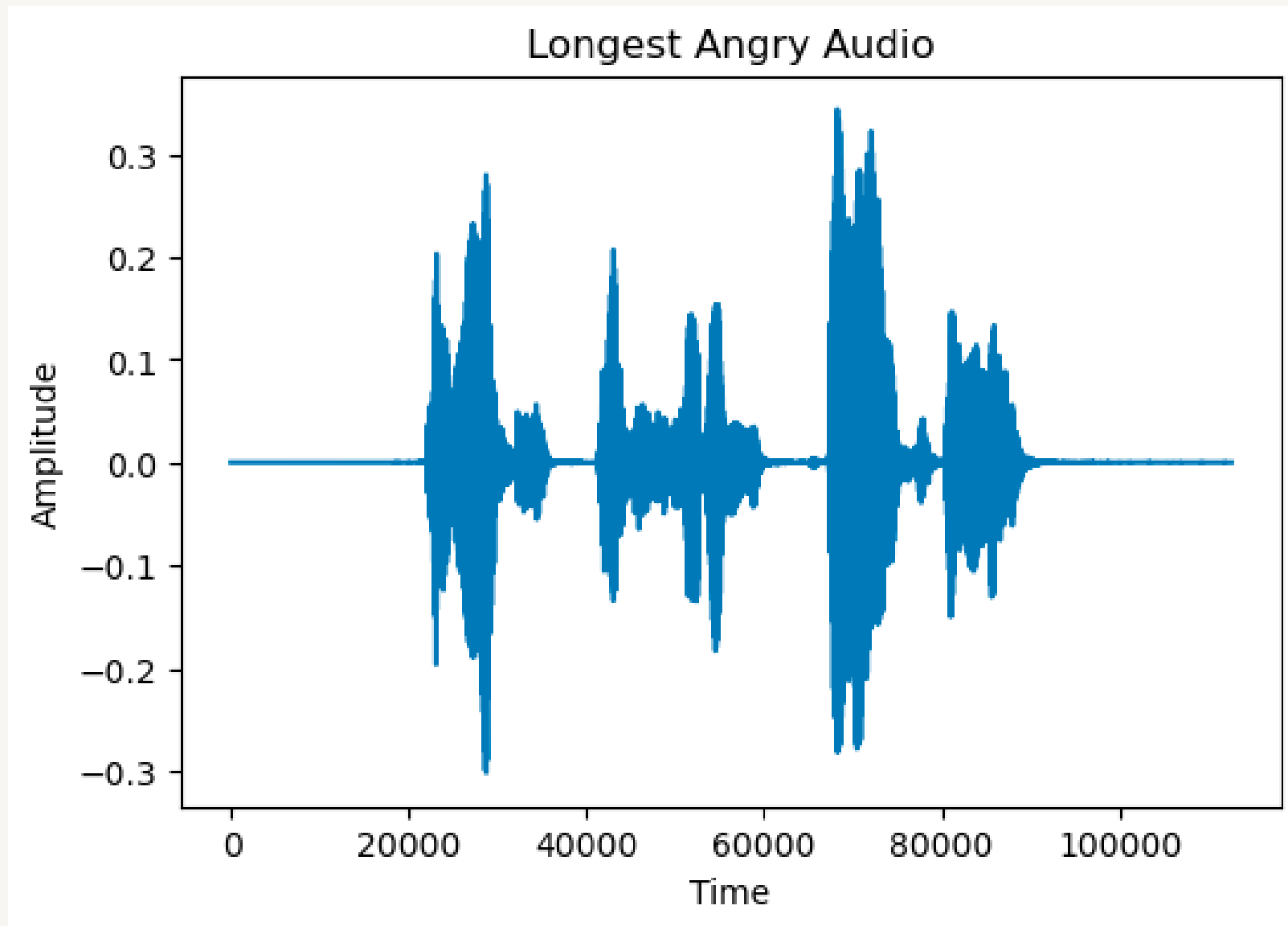
Shortest Audio: 2.94 s

## Loudness:

Maximum amplitude: 1.010

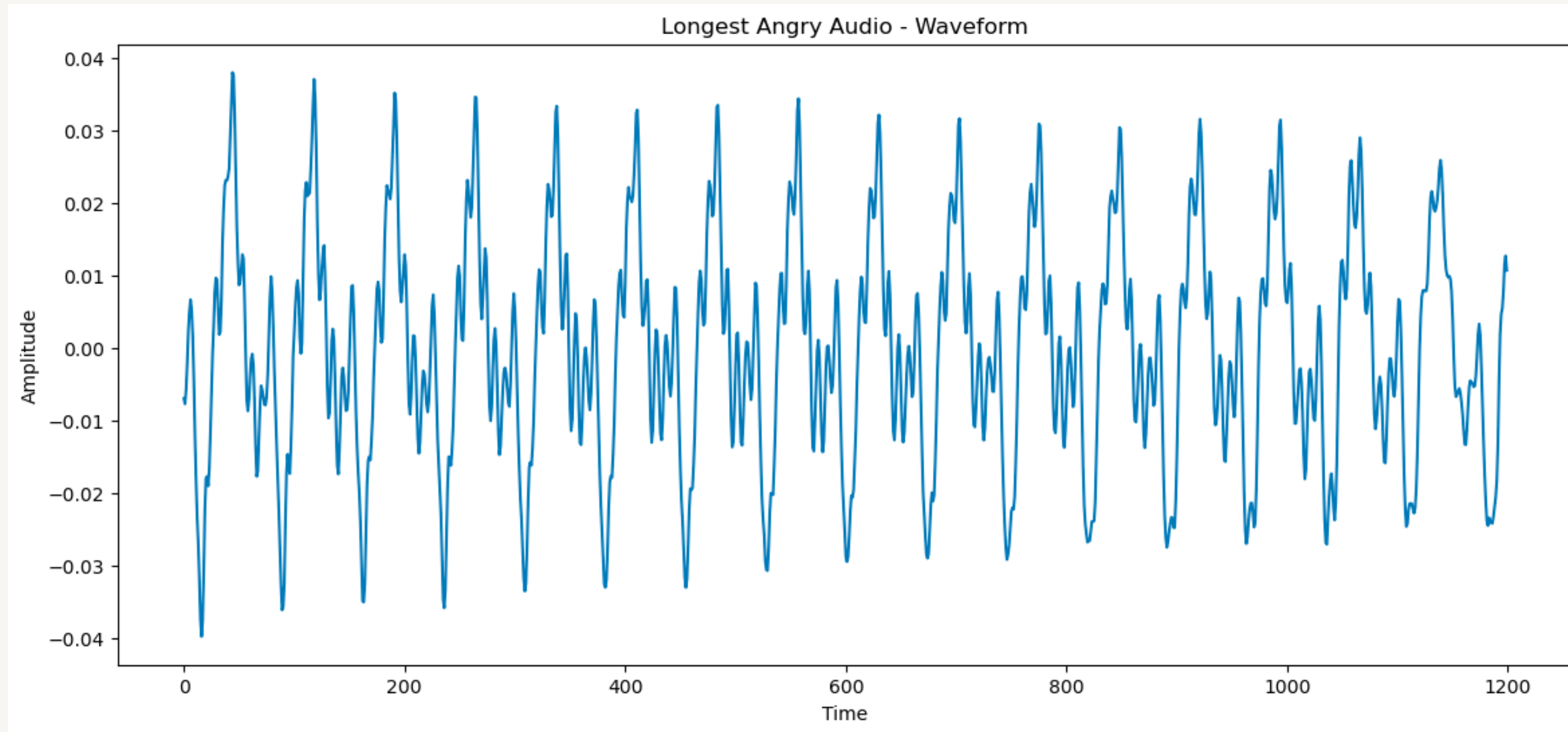
Minimum amplitude: 0.004

# Preprocess



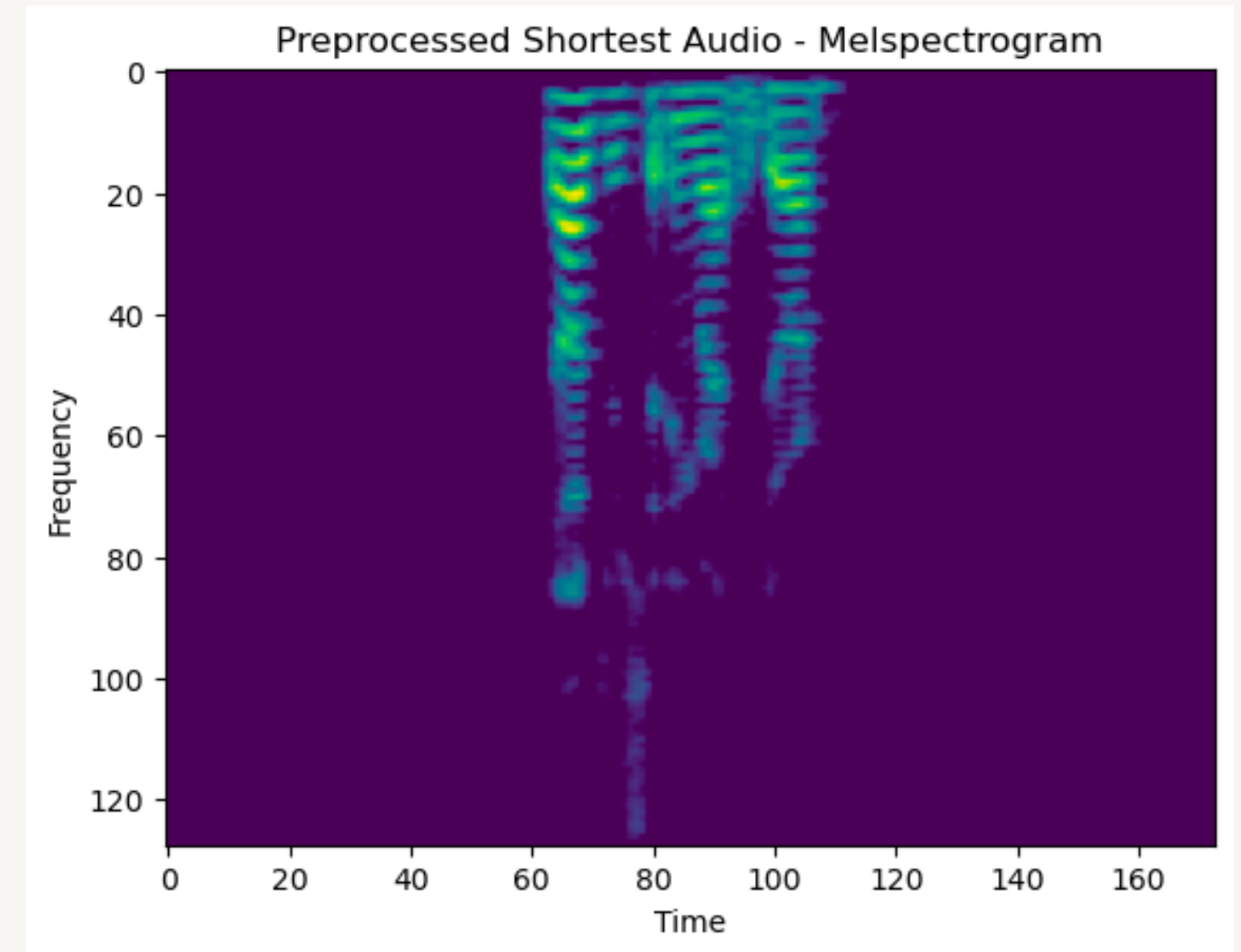
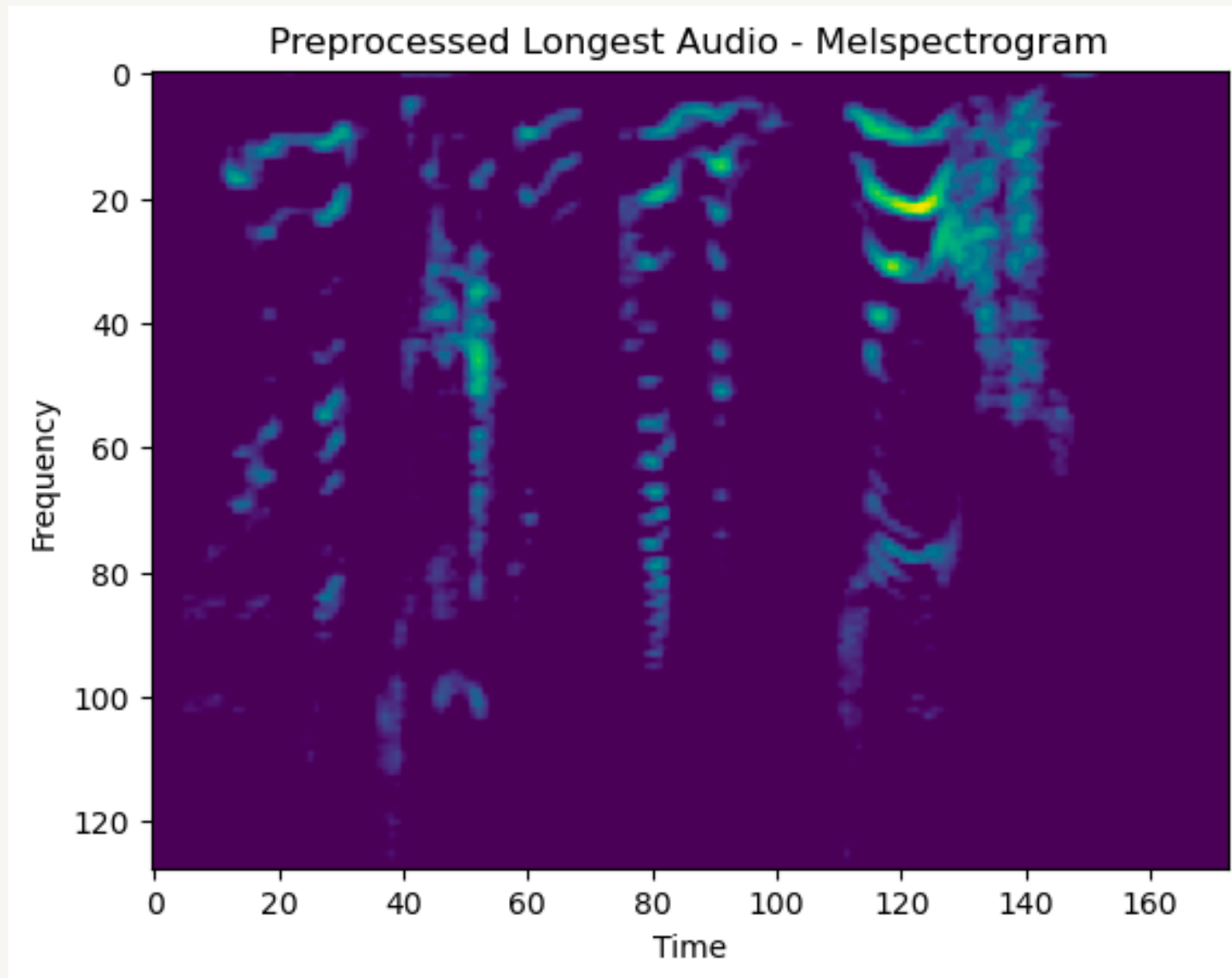
- Setting a threshold of 30db below reference to be considered as silence  
Benefit: smaller inputs, model runs faster
- Set same size for all - longer audios were trimmed, shorter were padded

# Preprocess



Sounds are many frequencies with different amplitudes varying along time

# Melspectrogram



Melspectrogram: takes into account how humans perceive frequency

Amplitude to dB: how humans perceive loudness

# Model

## Convolutional Neural Network

- Normalization Layer
- 2 Convolutional 2D Layers
- MaxPooling
- Dropout
- Flatten
- Dense
- Dropout
- Dense

Accuracy: 60% in validation data

# Limitations and Recommendations

---

- Further incorporation of different datasets
- Further incorporation of different emotions
- There are differences even when considering only words and excluding keywords, showing their concerns are different
- Subjective evaluation of emotions - people express and sense emotions differently



**Any questions?**