Audio Tagging with Neural Networks

GA DSI Capstone Project

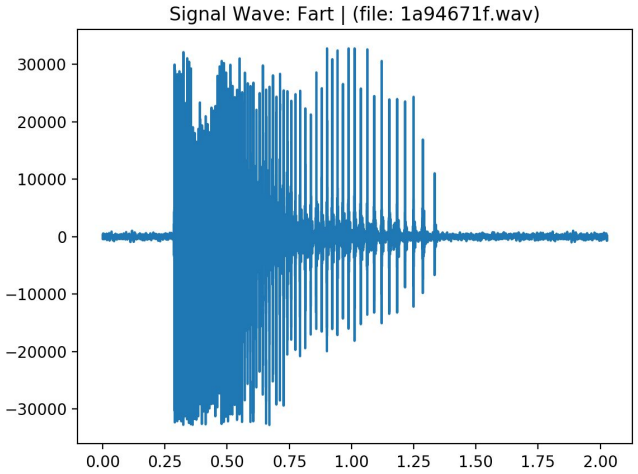
Kenny Evans

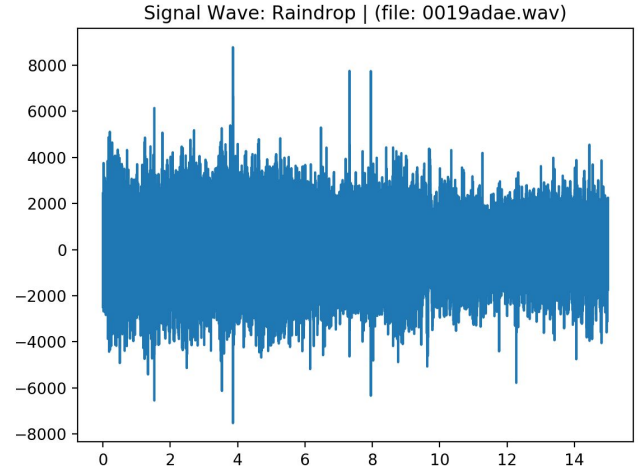
Project Overview

- Problem Statement
- Data Collection
  - 4,970 curated sound clips
  - 19,815 noisy sound clips
  - 80 categorical labels
- Computational Resources

Wave Signals

A quick comparison of a curated (top) versus a noisy (bottom) wave signal.

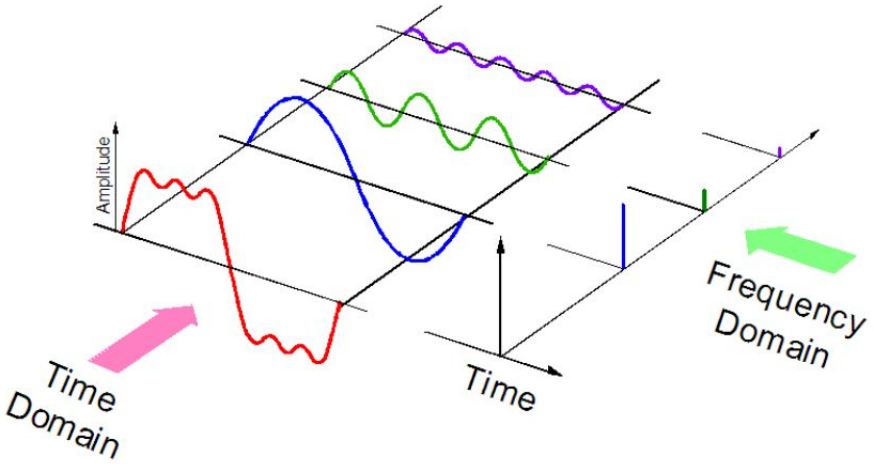




Preprocessing Options

- Amplitudes from raw wave signals
- Fourier transformations to decompose the signals into their constituent frequencies
- **Mel-frequency cepstral coefficients (MFCCs)**

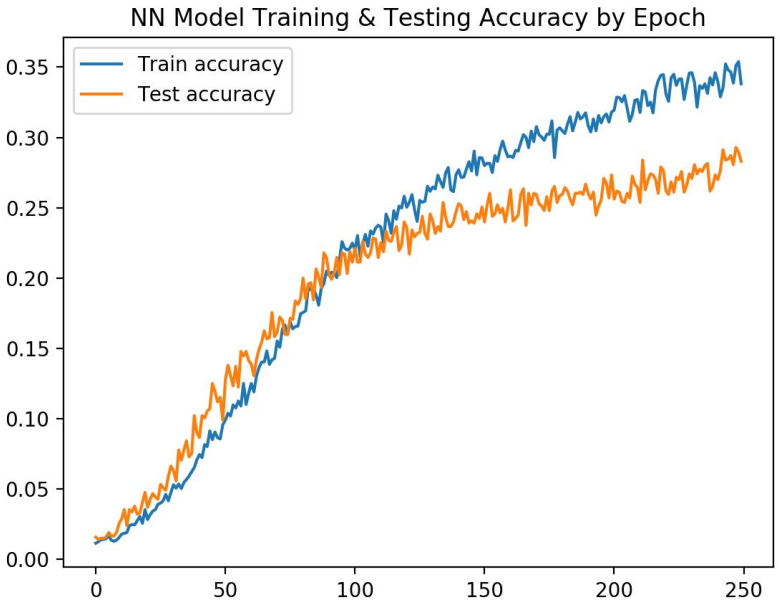
MFCCs Explained



1. Take the Fourier transform of the signal
2. Map powers of the spectrum obtained onto the mel scale
3. Take the logs of the powers at each of the mel frequencies
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal
5. The MFCCs are the amplitudes of the resulting spectrum

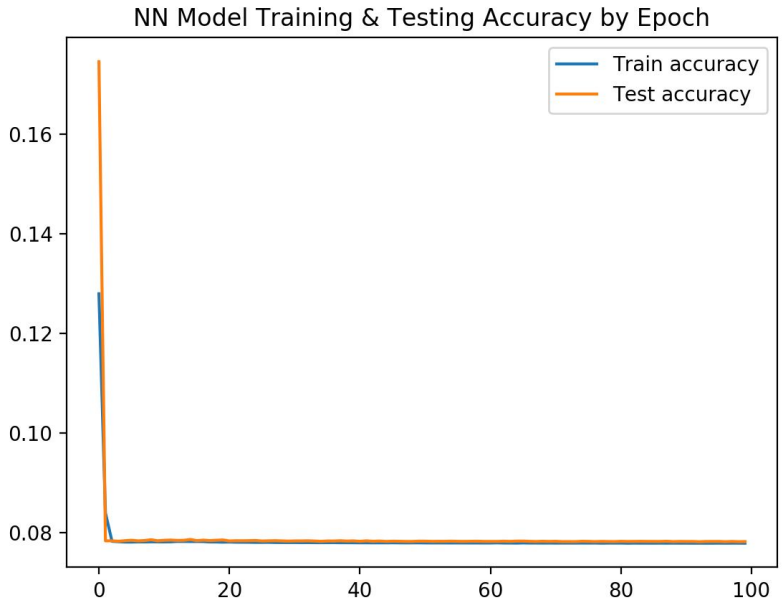
Neural Network Model I

- Sequential Neural Network
- Single-labelled targets
- Output activation: ‘softmax’



Neural Network Model II

- Sequential Neural Network
- Multi-labelled targets
- Output activation: ‘sigmoid’



Neural Network Model III

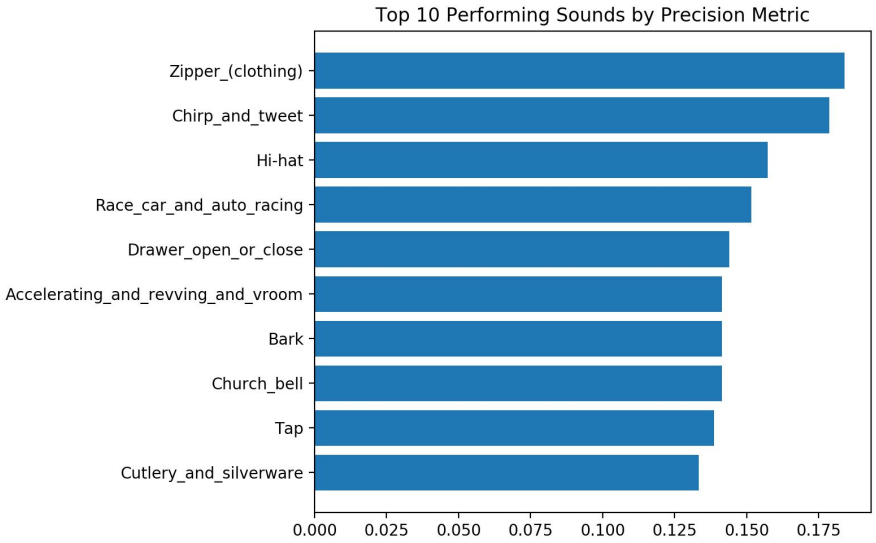
- Sequential Neural Network
- Multi-labelled targets
- Output activation: customized ‘sigmoid’

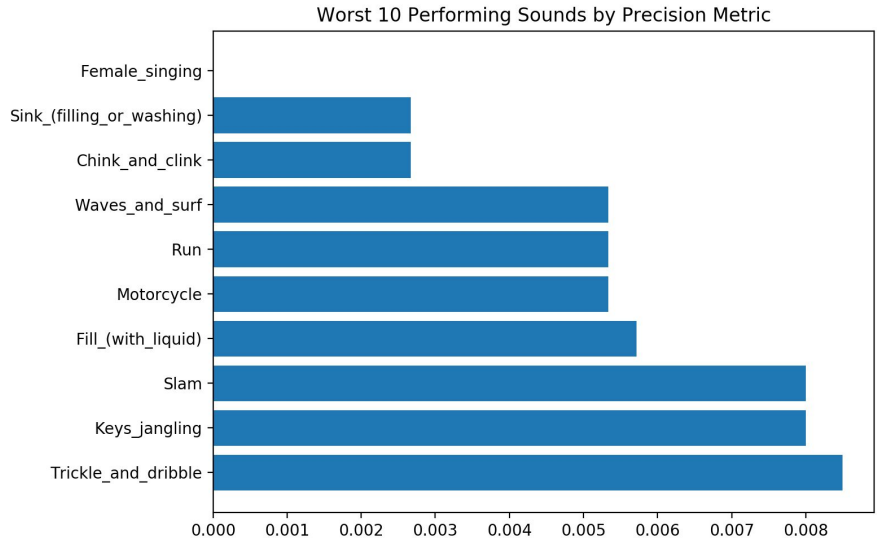
Precision -> .055

Baseline precision -> .002

27.5 times better than baseline!

Takeaways


Top 10 Performing Sounds by Precision Metric
Zipper_(clothing)
Chirp_and_tweet
Hi-hat
Race_car_and_auto_racing
Drawer_open_or_close
Accelerating_and_revving_and_vroom
Bark
Church_bell
Tap
Cutlery_and_silverware
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175


Worst 10 Performing Sounds by Precision Metric
Female_singing
Sink_(filling_or_washing)
Chink_and_clink
Waves_and_surf
Run
Motorcycle
Fill_(with_liquid)
Slam
Keys_jangling
Trickle_and_dribble
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008

Future Steps

- Improve loss function for neural network
- Apply different transformations to wave signals
- Apply background noise reduction techniques