Audio Classification for Smart Home Audio Systems

Springboard Capstone 3

The Fully Connected Home

Systems

Lights, temperature, <u>audio</u>, energy, visual media, & more...
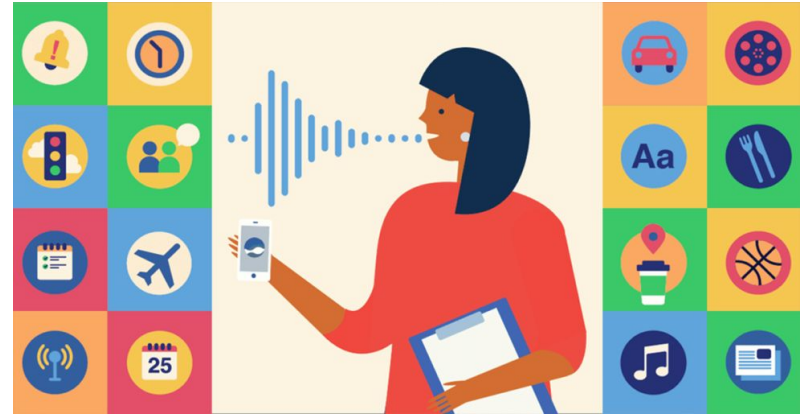
Control interfaces

App, voice.

Fluid Responsiveness

The smart home ecosystem increases functionality, but falls short of offering a new control paradigm

- App controller: hands on, multiple apps, touch control, 3rd party
- Voice controller: few providers, active, specific sequences necessary

Requirement: smart home should have smart input

Automatic Experiences

With environmental detection through sound, new experiences are opened up for smart home audio system customers:

- **Dog barking:** auditory nudge to let out
- **Knock on door:** nudge alerting arrival
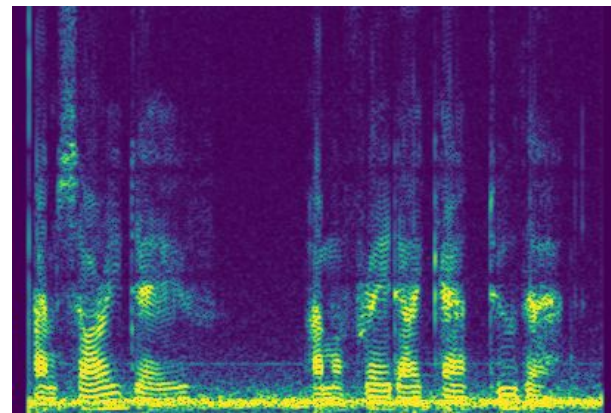- **Laughter:** cue party playlist

Recognizing Audio Data

5684, 5326, 5435, 5109, 4306, 3636, 3156, 3765, 4001, …

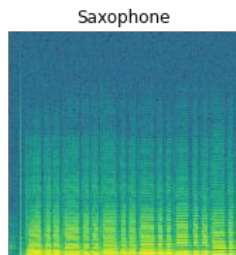waveform   array of amplitudes over time   spectrogram

Methodology

model.compile(
    optimizer=tf.keras.optimizers.Adam(),
    loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=**True**),
    metrics=['accuracy'],
)

1) Digital File

'Saxophone'

41 labels:

```
array(['Oboe', 'Bass_drum', 'Saxophone', 'Chime', 'Electric_piano',
       'Shatter', 'Bark', 'Acoustic_guitar', 'Scissors', 'Double_bass',
       'Knock', 'Telephone', 'Violin_or_fiddle', 'Gunshot_or_gunfire',
       'Burping_or_eructation', 'Clarinet', 'Computer_keyboard', 'Flute',
       'Cello', 'Tambourine', 'Drawer_open_or_close', 'Snare_drum',
       'Fart', 'Meow', 'Trumpet', 'Fireworks', 'Bus', 'Keys_jangling',
       'Applause', 'Harmonica', 'Cough', 'Gong', 'Glockenspiel',
       'Tearing', 'Writing', 'Squeak', 'Microwave_oven', 'Laughter',
       'Finger_snapping', 'Hi-hat', 'Cowbell'], dtype=object)
```

2) Spectrogram

Saxophone

3) Convolutional Neural Network (training on patterns)
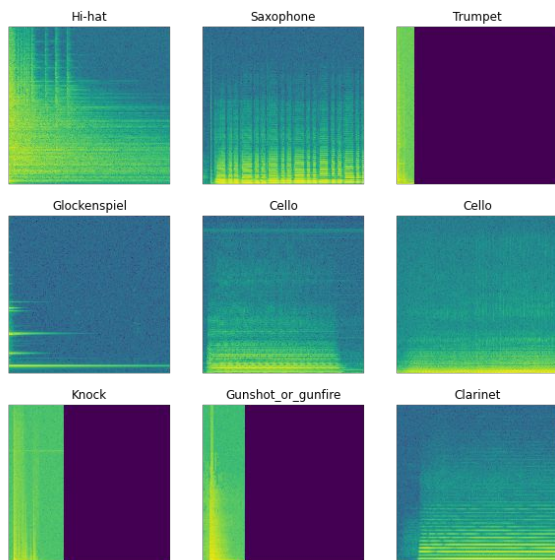

feature extraction   classification
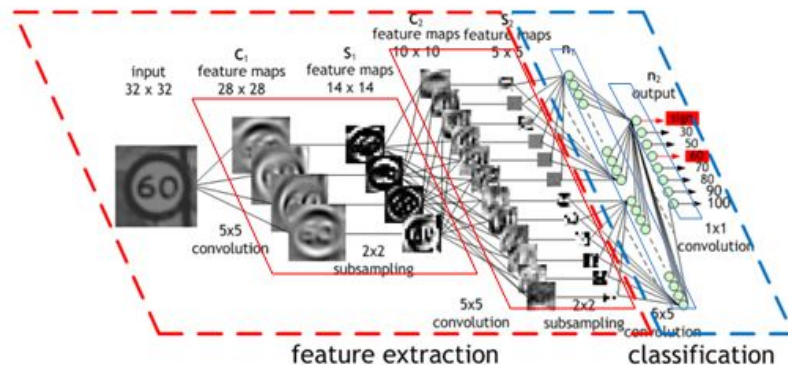
TensorFlow:
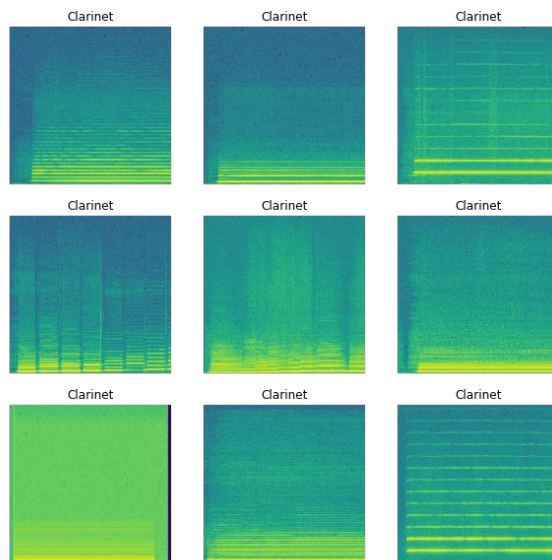
audio, _ = tf.audio.decode_wav(contents=audio_binary)

spectrogram = tf.signal.stft(equal_length, frame_length=555, frame_step=343)

Spectrograms: A Closer Look
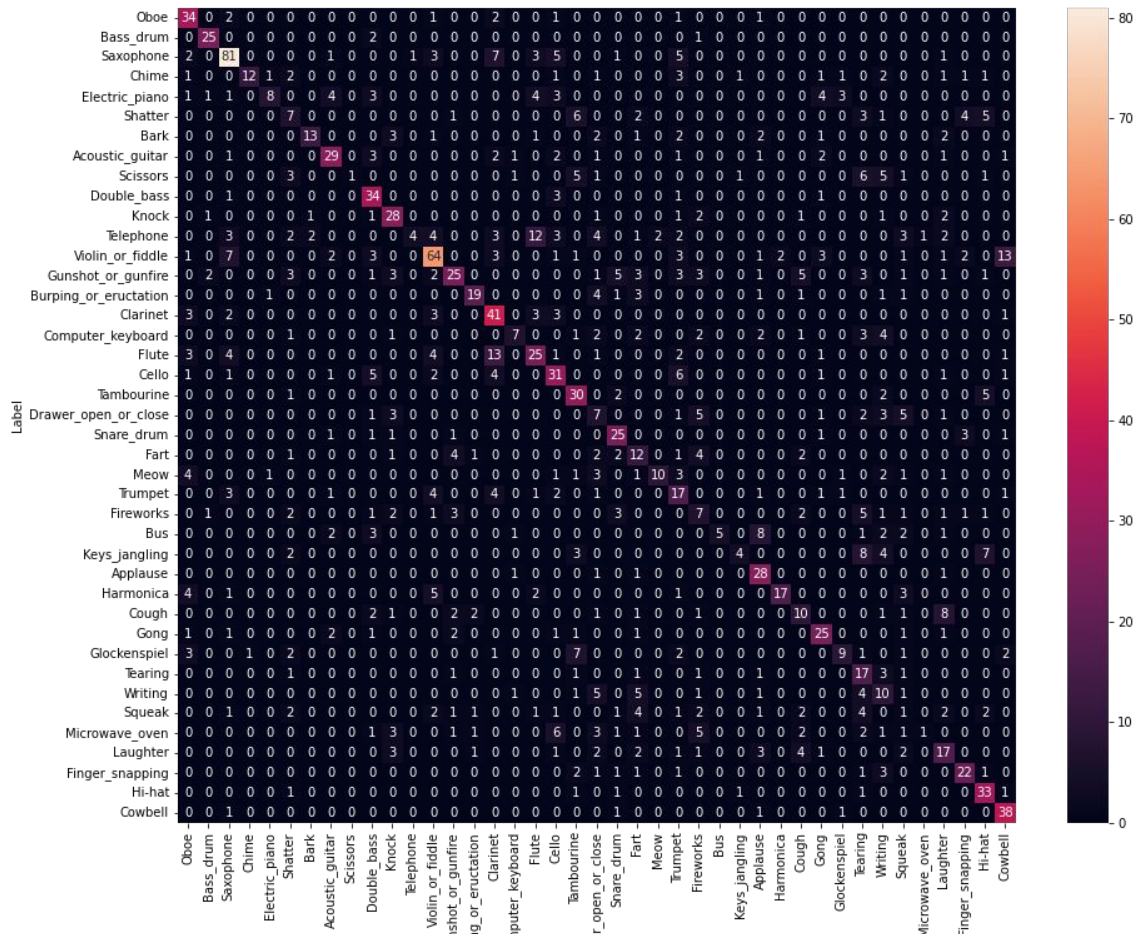
Across Label

Within Label

Results

Detectable Stimuli

41 Inputs

Accuracy

50%

Opportunities, and... Future Growth Potential (~90% acc.)

Barks, meows

Pet care nudge

Chime, knock

Guest or delivery interaction

Laughter, finger-snapping

Party DJ

Shatter, gunshot

Home security

Drawer, microwave oven

Audible recipe

instruction

Next Steps

Maximize audio information

Adaptive input size

Isolate target sounds

Signal filtering

Optimization

Hyperparameter tuning

Length of Audio Files