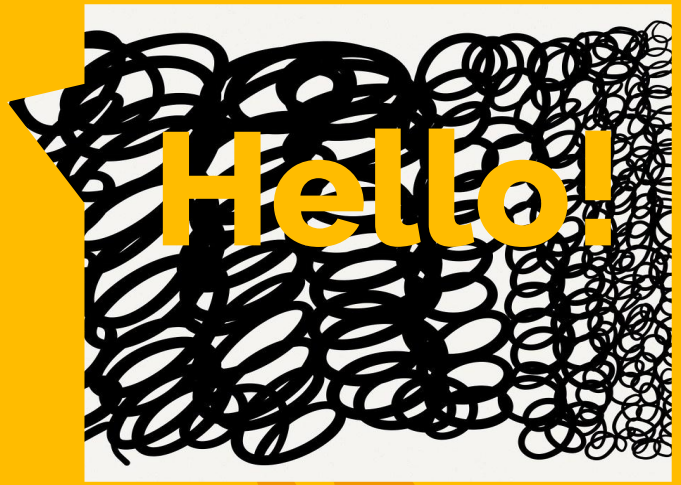
Multi-Speaker Identification Breakdown

1. Our Products Mission is ...!

For software developers who want to be able to recognize the key speaker on the fly. The Multi-speaker identifier is an open source product that integrates signal processing and machine learning to be able to amplify the user's voice over other noise and people. Unlike AI assistants and APIs, our product will do its processing in real time and it not limited to the person whose voice was sampled.



2. Existing Works

SincNet

Uses machine learning to figure out the start and stop frequencies of the filter (in the time domain, which is why it is the sinc waveform) of each segment of the voice signal [1]

Google API

“Speech-to-Text can recognize multiple speakers in the same audio clip.” [2]

Siri

The “Hey SIri” function “ “uses a Deep Neural Network (DNN) to convert the acoustic pattern of your voice at each instant into a probability distribution over speech sounds.” [3]

MVP

- Real-time voice recognition
- Must recognize at least one voice from the current situation
- Must be able to filter out environment's noise

MVP User Stories! -- Possible SW Developer uses

I am a WFH 9-5 employee,

And I want to be able to speak with my manager without voices seeping through as background noise

I am physically disabled,

And I want to be able to use my phone by voice without worrying to compromise effectiveness due to my environment

Technology

Evaluate:	ML	DSP
Reasoning	**Allows for functionality in a changing environment**	**Allows for necessary functions (e.g., filters and domain transformations)**

Set-Up

- Set up Python coding environment
- Set up the machine learning model's base
- Training data format decision and collection

Thanks!

Any questions?

You can find me at EC 601 Section 1A Lecture

Credits

- [1] https://github.com/mravanelli/SincNet
- [2] https://cloud.google.com/speech-to-text/docs/multiple-voices
- [3] https://machinelearning.apple.com/research/hey-siri