

In Person Conversation Analysis

Shruti Badrish, Shantanu Govalkar, Iris Litiu



Agenda

Objective

Prior Work

Data Collection

Speech Diarization

Metrics

Accuracy & Error

Privacy

Next Steps

Demo



Objective

The end goal of this project is to:

- Provide readable conversation feedback to users
- Record audio live or use pre-recorded audio
- Help users visualize a conversation
- Gives graphical and text metrics that are beneficial

Prior Work in Space

- Zoom conversation analysis
 - Had individual microphones for each speaker
 - Zoom automatically highlights speakers, so it's easier to detect interruptions by finding overlap in speech
- Speaker diarization
 - Models have already been developed which segment a recording into its component speakers

Audio Input





Step 1 - Data Collection

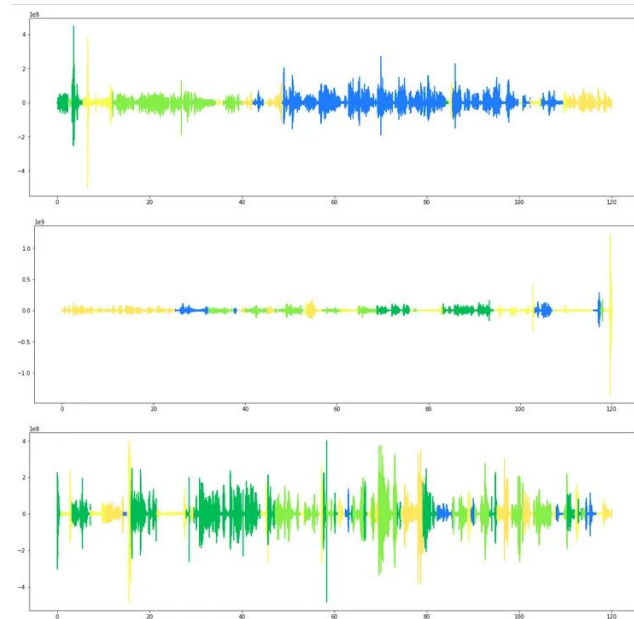
- Initial Data Collection
 - Recorded most use cases off the top of our head
 - Helped give preliminary testing data for our metrics
- Synthetic Data
 - Use pre-recorded snippets, randomly select 1 second segments, make them overlap for a random amount of time
 - Easily generate large amounts of data

Step 2 - Speech Diarization

Tested multiple models:

- diart (<https://github.com/juanmc2005/diart/tree/main>)
- pyAudioAnalysis (<https://github.com/tviannak/pyAudioAnalysis>)
- SpeechBrain (<https://github.com/speechbrain/speechbrain>)
- Pyannote (<https://github.com/pyannote/pyannote-audio>)
- Google Cloud
(<https://cloud.google.com/speech-to-text/docs/multiple-voices>)

We ultimately decided on **Google Cloud** because of its ability to transcribe the audio as well as its csv file output.



<https://medium.com/@ishantwadhwa4/plot-speaker-diarization-with-pyplot-7c8f6e677da8>



Metrics

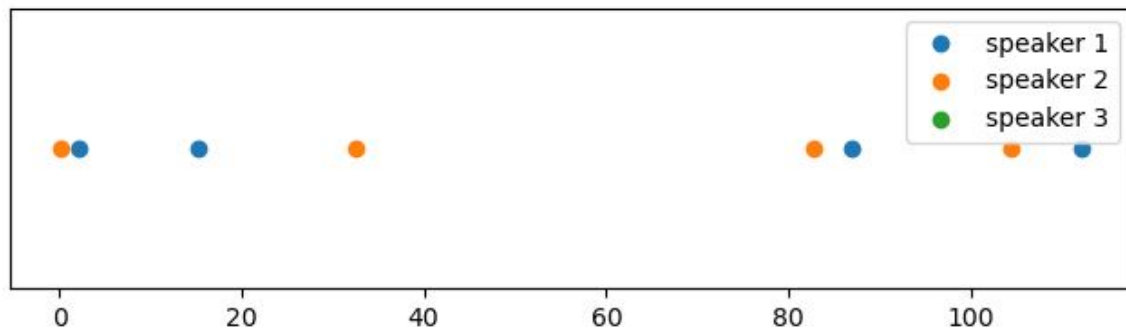
Figures:

- Speaker Visualization
- Interruptions - Required Model Training
- Breaks in Conversation
- Overall Sentiment
- Sentiment by Speaker
- Change in Volume
- Filler Words

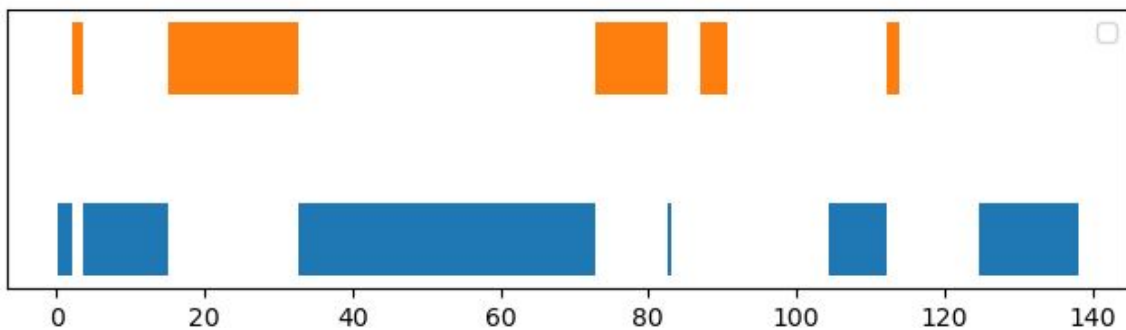
Text:

- Speaker Speeds
- Overall Sentiment Score
- Speech Length
- Awkwardness Metrics
- Turns per Speaker

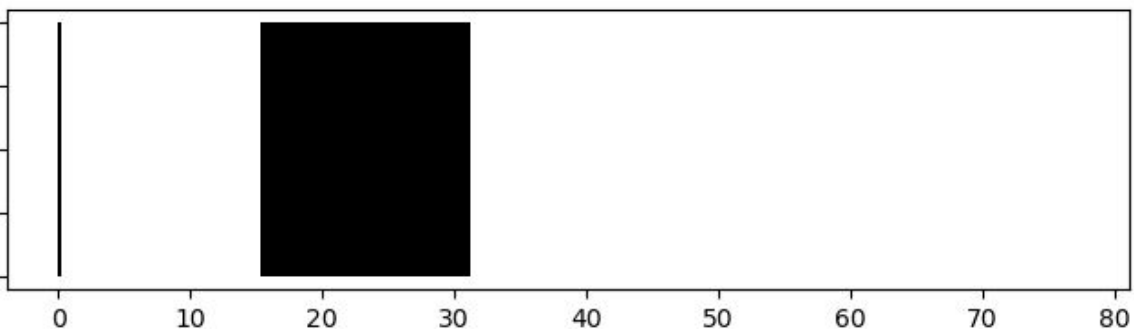
Interruptions:



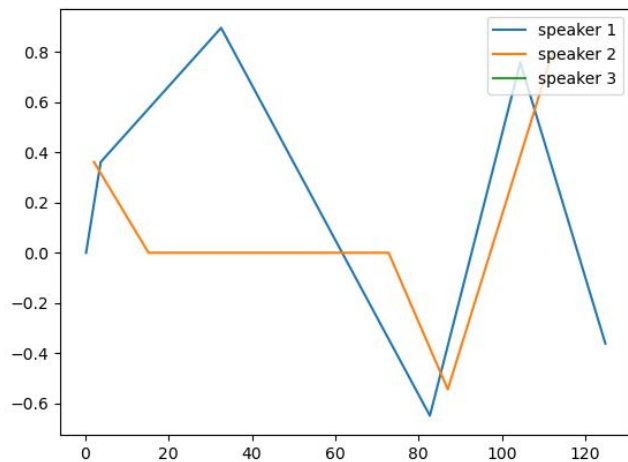
Speaker Visualization:



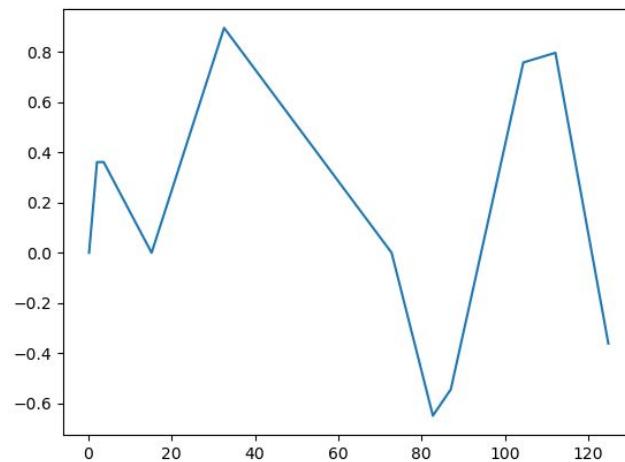
Awkwardness:



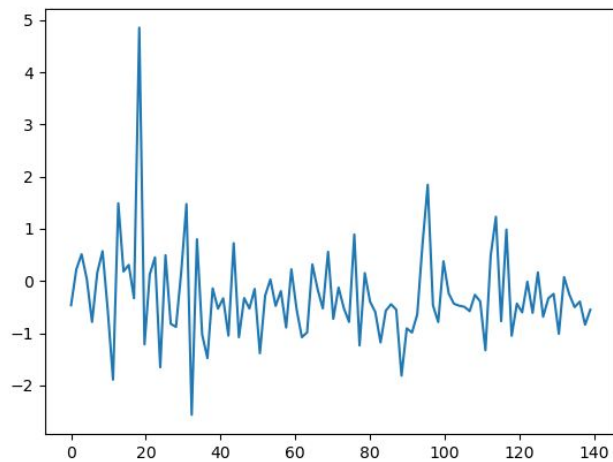
Sentiment
by Speaker



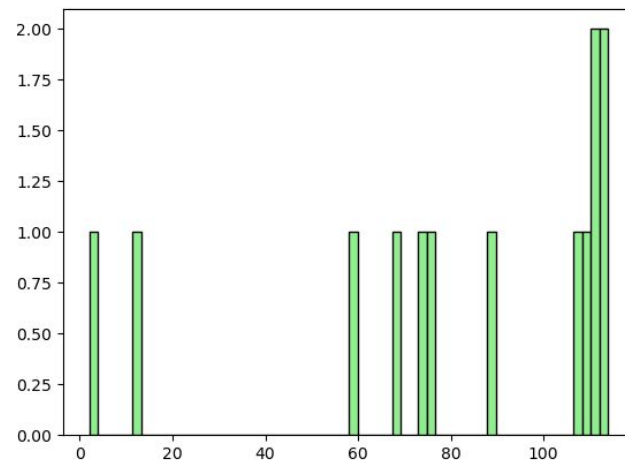
Overall
Sentiment



Total
Volume



Filler
Words





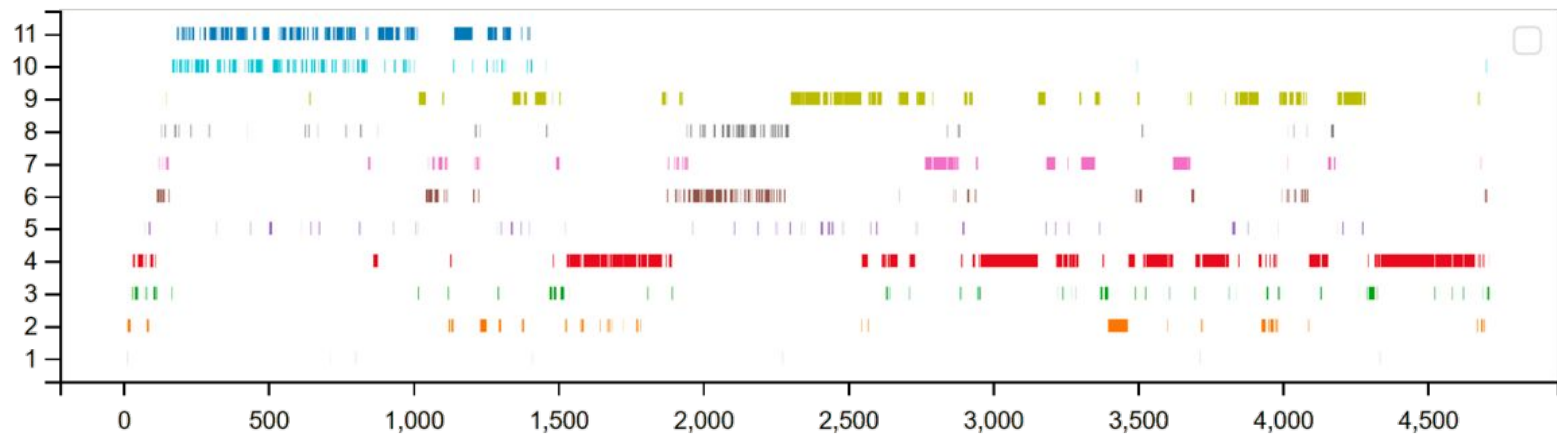
AI Feedback

We used ChatGPT to generate feedback based on details about the conversation. Users input information about recordings prior to requesting feedback. In this example, I wrote: *Two people are giving a presentation to a group about a website which analyzes conversations.*

Sentiment Feedback	Speech Feedback	Contribution Feedback
<p>Speaker [1, 2]: To more positively contribute to a conversation within a two-person lecture presentation, it's essential to be an active listener. Pay attention to the content presented, jot down any questions or comments that come to mind, and wait for the appropriate time to voice them. Always respect the presenters by not interrupting them mid-speech. Show engagement by nodding, maintaining eye contact, and showing appropriate reactions. When asking questions or giving feedback, be clear, concise, and respectful. Make sure your contributions are relevant to the topic at hand. Encourage others in the group to express their thoughts and ideas.</p> <p>Remember, your goal is to enhance the learning experience for everyone, not to dominate the conversation. Lastly, always appreciate the efforts of the presenters, it encourages a positive environment.</p>	<p>['Speaker 1']: Good speaking speed ['Speaker 1']: Good speaking speed ['Speaker 2']: To talk slower in a lecture, first, practice your presentation beforehand to get a sense of timing. Use a metronome or timer to pace yourself. During the lecture, remember to breathe and pause between points for emphasis. Articulate each word clearly and avoid rushing through your material. Maintain a relaxed posture and make eye contact with your audience to gauge their understanding and adjust your speed accordingly. It's okay to take brief pauses to gather your thoughts. This not only helps you slow down but also allows your audience to absorb the information.</p>	<p>['Speaker 1']: Well balanced contribution</p>

Robustness

This website works for large recordings, including long meetings.





Interruptions

Version 1: Chat GPT

- Asking Chat GPT if there was a likely interruption (over 70%)
- If true, mark as an interruption

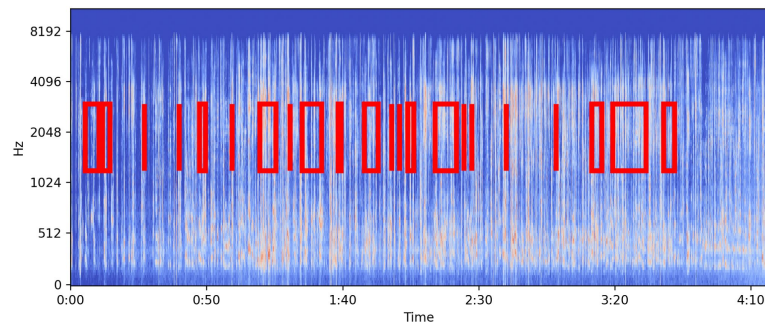
Version 2:

- Random Forest model: further train an existing model for interruptions
- Still R&D

Interruption Detection With ML

Implementation of a lightweight Random Forest classifier model to detect interrupts in user speech.

- Trained on YAMNet library data and our own synthetically generated interruption recordings.
- We extracted MFCC feature vectors from each recording, and labeled the recordings at each timestamp with a 1 (corresponding to an interruption) or a 0 (no interruption), the latter of which constituted our target dataset.



Display of speaker overlap within mel spectrogram



Accuracy & Error

When provided with a clear recording, as well as an accurate number of speakers, the Google Cloud diarization model is able to effectively transcribe what is being said, with some error.

Example:

- Original Dialogue: **This is a test Google Cloud recording**
- Recorded Dialogue: **this is a test to Google Cloud recording**

Ethical Concerns

- Privacy issues arise when recording conversation
- Concerns how website is used
 - Can companies use it the second employees walks in?
How does it impact employees?
 - Do companies tell employees they use it?
 - If they tell, employees act differently
 - If they don't, privacy violation (?)
- Possible solutions:
 - Company policies on AI feedback
 - Ask for permission to record and use recording data before every meeting





Conclusions & Future Work

Future Work:

- Improve accuracy of the model - Train Google Cloud diarization pipeline further
- Improve graphs (esp. for longer recordings)
- Web app
 - Flask auth
 - Store user data in a SQL server (in progress)
- Address privacy concerns
- Start user studies
 - Find more useful metrics
- Write a paper



Acknowledgements

- Our mentors, Devesh Sarda and Joe Breda (for scaffolding our design and coding processes)
- Internship coordinator, Jerry Cao
- UbiComp Lab and the University of Washington
- The developers of the numerous libraries we used in the process of creating this web app

Thank you for listening!