

Background

The Problem:

Imagine a secretariat and project manager, they must send out a meeting summary and action items in a timely manner for follow up and decisions to be made. During the meeting, it is difficult to actively participate in the meeting while simultaneously creating extensive notes.

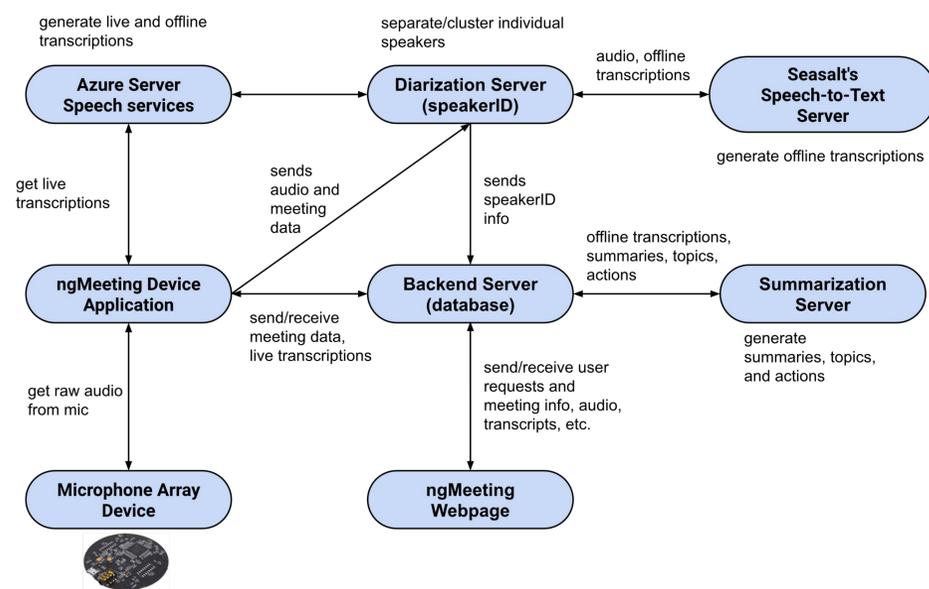
To allow the secretariat/project manager the ability to both keep records of the meeting, while also contributing to the meeting, a solution must be able to accurately (>80% accuracy) tag multiple speakers, and deliver close to real time transcriptions and summaries.

Our Solution:

ngMeeting is a meeting transcription system that records in-person and telephone meeting conversations to make spoken communications easily searchable and actionable. It uses a microphone array to record speakers' voice, identify different speakers, and generate meeting transcriptions of multiple speakers in real-time or asynchronously, with meeting summarization and to-do "actions" list created shortly after the meeting.

A meeting page will display the name, date, and start time of the meeting. The host users can add participants, edit meeting information and verify transcripts, and download meeting content. Participants can view meeting content, add actions, listen to meeting audio, and watch live transcriptions.

System Architecture



Microphone Array for Far-Field Speech Recognition



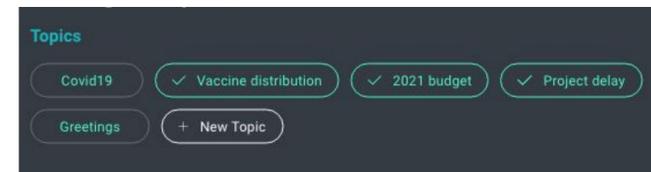
A microphone array is used to collect audio, the array allows for spatial information to be collected. The direction of arrival of the signals assist in noise reduction and speaker identification. Beamforming takes the direction of arrival so that sounds from different directions can be suppressed (noise) or enhanced (speakers).

Transcriptions and Speaker Diarization

- Speakers are separated via Diarization after the meeting has concluded
- Transcriptions are separated by speaker
- The original generated Azure Speech transcription is found on the left side, and the user corrected/verified transcripts are on the right
- Audio clips for each utterance can be played using the audio player next to the speaker ID.



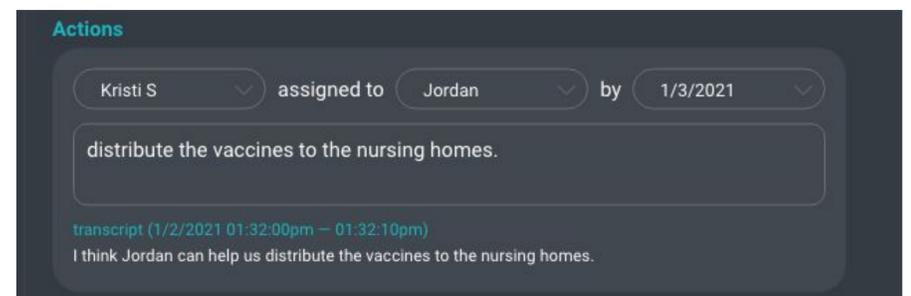
Topic Extraction



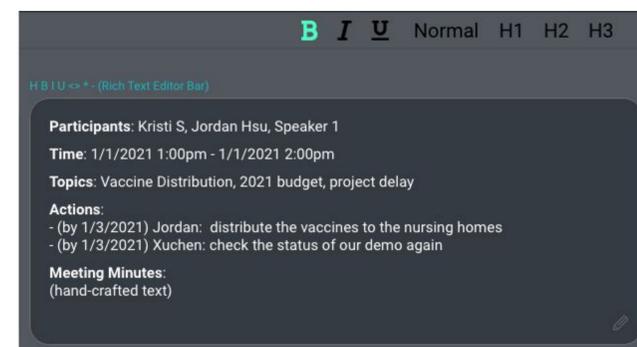
Topics are generated via keyword extraction using keyBERT [1] and topic abstraction using fastTEXT [2] and Wikipedia data.

Actions Extractions

Actions are extracted from the transcripts by locating 'actionable' segments, then it takes the previous 2 segments for context and produces an action summary based on those three segments.



Summarization Generation



A machine learning summarization model developed from the BART [3] and SAMSum [4] dataset for abstractive summarization.

Future Work, References, and Acknowledgments

- Create user authentication
- Remotely start meeting from website
- Improve accuracy of ML models
- Improve UI/UX and conduct user tests
- Support more languages, especially less supported languages.

- [1] [KeyBERT](#)
- [2] [fastTEXT](#)
- [3] [BART](#)
- [4] [SAMSum](#)

Special Thanks to the Seasalt.ai Team and our TA Haobo Zhang