A screenshot of a computer

AI-generated content may be incorrect.[Speaker Recognition Based on Deep Learning: An Overview](https://arxiv.org/pdf/2012.00931)

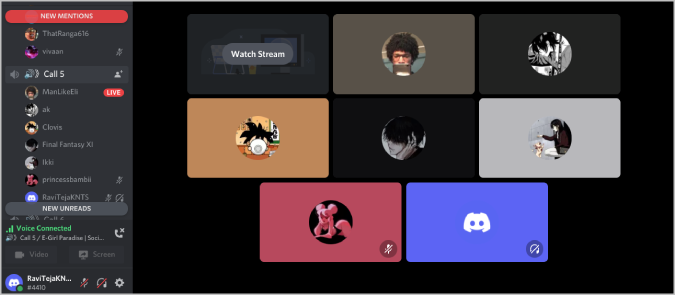
*Problem*

“In a busy room full of people I know, I cannot understand who is speaking and what they are saying”

1. Who is speaking?
2. What are they saying?

*Solution*

An app on a phone or tablet which identifies speakers and what they are saying. (Imagine a Discord call with speech-to-text).



Users can add multiple ‘profiles’ including a profile picture and a training set of their voice.

The app will be able to identify each of these voices in a noisy environment and transcribe what each of them are saying.

*Back-End/ML Process*

1. Speaker diarization of *n* speakers
2. For speaker *i*🡪*n*:

* Identify the speaker, or return “unsure” (avoid hallucination)
* Identify what they are saying (hallucination is possible here)

*Potential issues*

* Is there a max no. identifiable speakers? (eg. 3 at once is okay. 6 at once? 8?)
* For now only English would be possible
* How will speaker diarization be affected in a noisy environment?
* Will multiple speakers at once be able to be transcribed?
* Scope creep already happening and will get worse (see below)

*Reducing Scope – alternative project ideas*

**Segmented Speech-to-Text with Simultaneous Speakers:** “Can multiple people be transcribed, speaking at the same time?”

**Speaker Diarisation with Many Speakers:** “How many speakers can be diarised at once?”

**Speaker Diarisation with Similar Speakers:** “Is diarisation still possible with very similar speakers? What changes need to be made to a model to capture smaller differences in an input?”

**Single Speaker Diarisation and Transcription:** “Can a speaker be diarised and transcribed simultaneously?”

*Further Related Ideas*

**Speech-Related Sound Augmentation:** “Can the noise of an environment be cancelled, leaving only human noise?” (Noise Cancelling Hearing Aids)