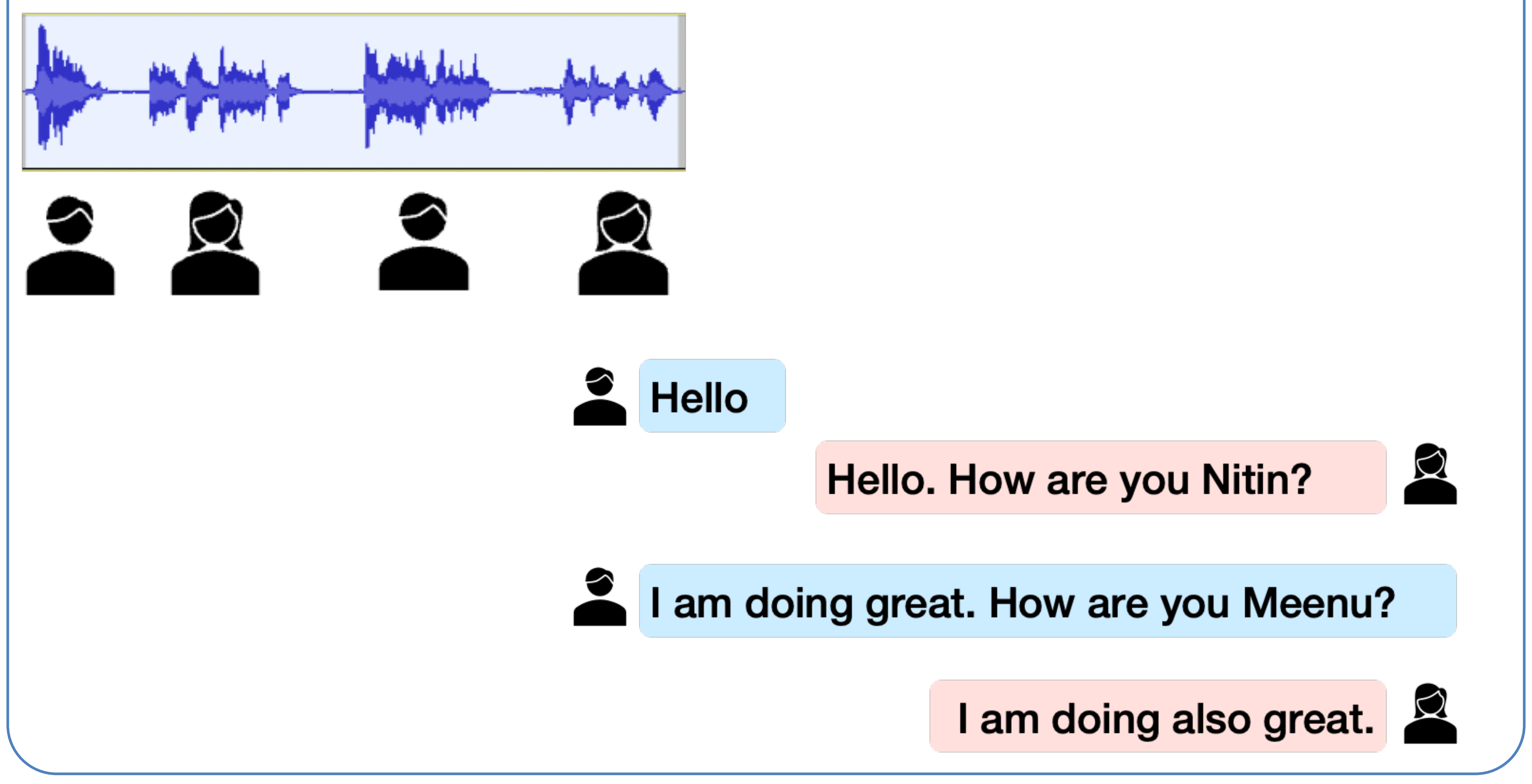
Who spoke when in a conversation?

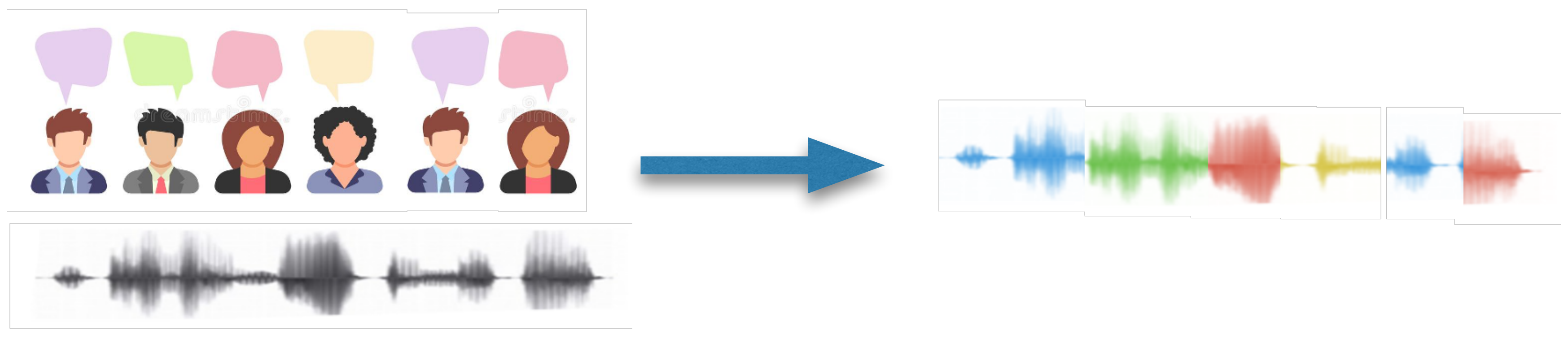
Prachi Singh, Sriram Ganapathy
LEAP Lab, Electrical Engineering,
Indian Institute of Science, Bangalore

Introduction

Conversational audio contains multiple speakers engaged in a conversation. Transcribing audio into text using speaker information generates much meaningful text.


Hello
Hello. How are you Nitin?
I am doing great. How are you Meenu?
I am doing also great.

Who spoke when?



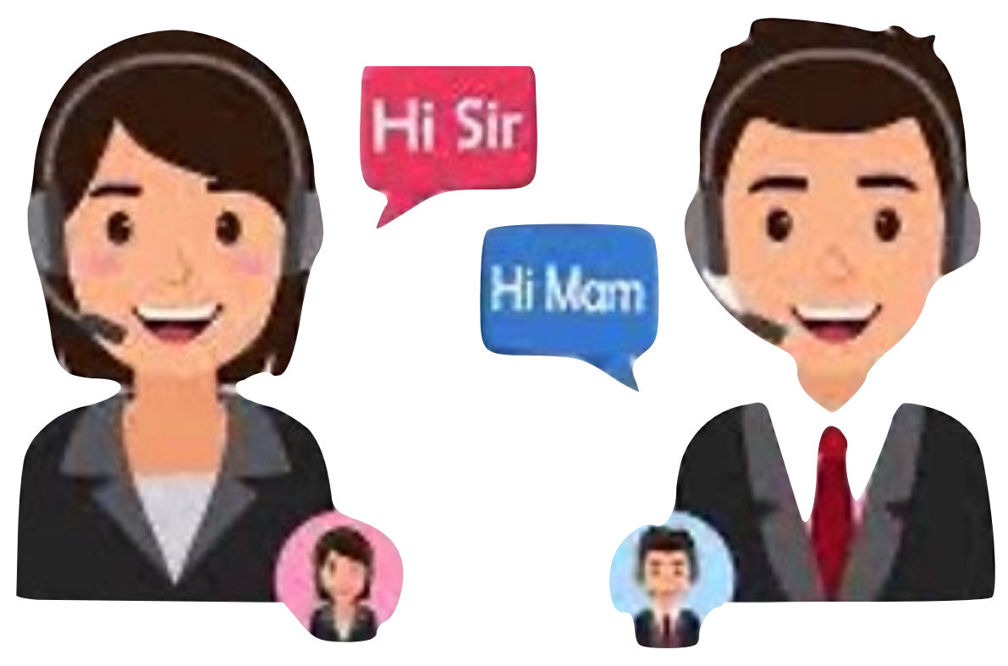
- **Speaker Diarization** is the task of finding "**who spoke when**?" in a multi-speaker conversational audio.
- It involves partitioning an input audio stream into segments based on speaker sources.

Applications

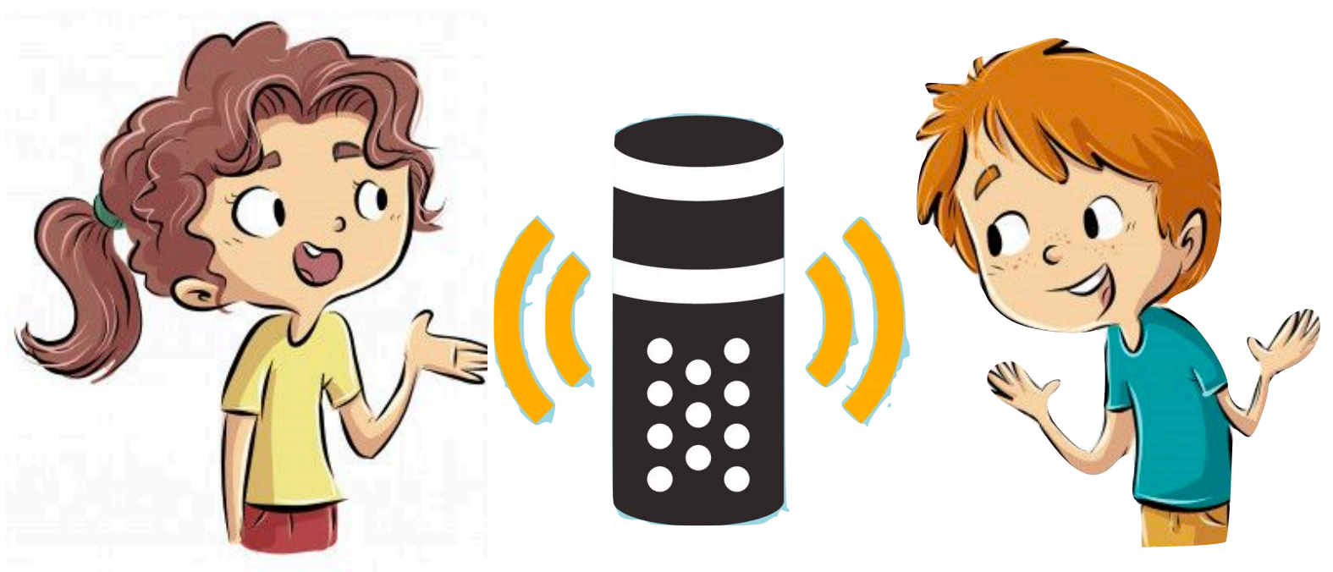
Converting text to speech

Transcription

Behavioural analysis of agents and customer to understand customer satisfaction



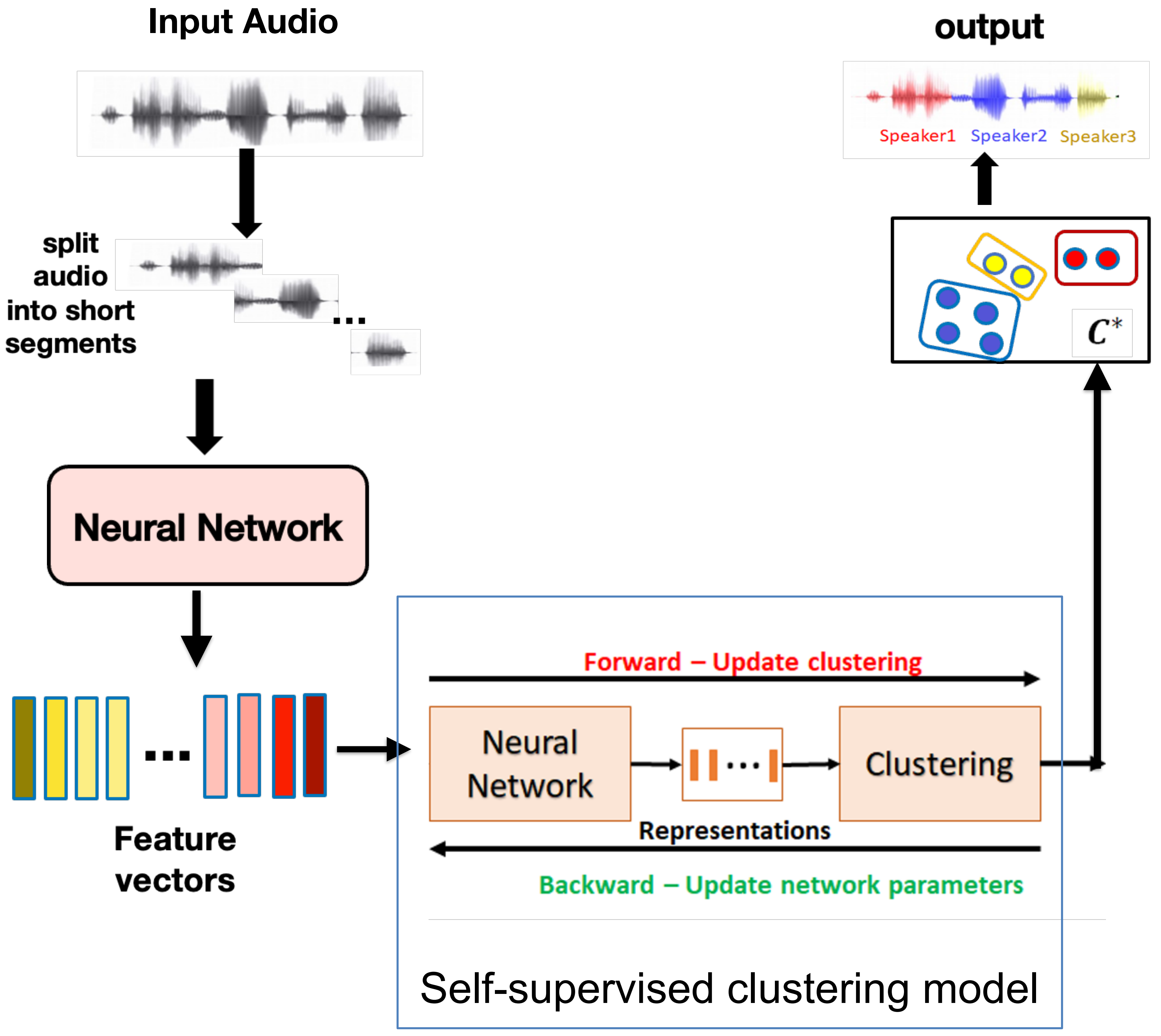
Call center



Smart assistant

Improving human machine interaction

Approach



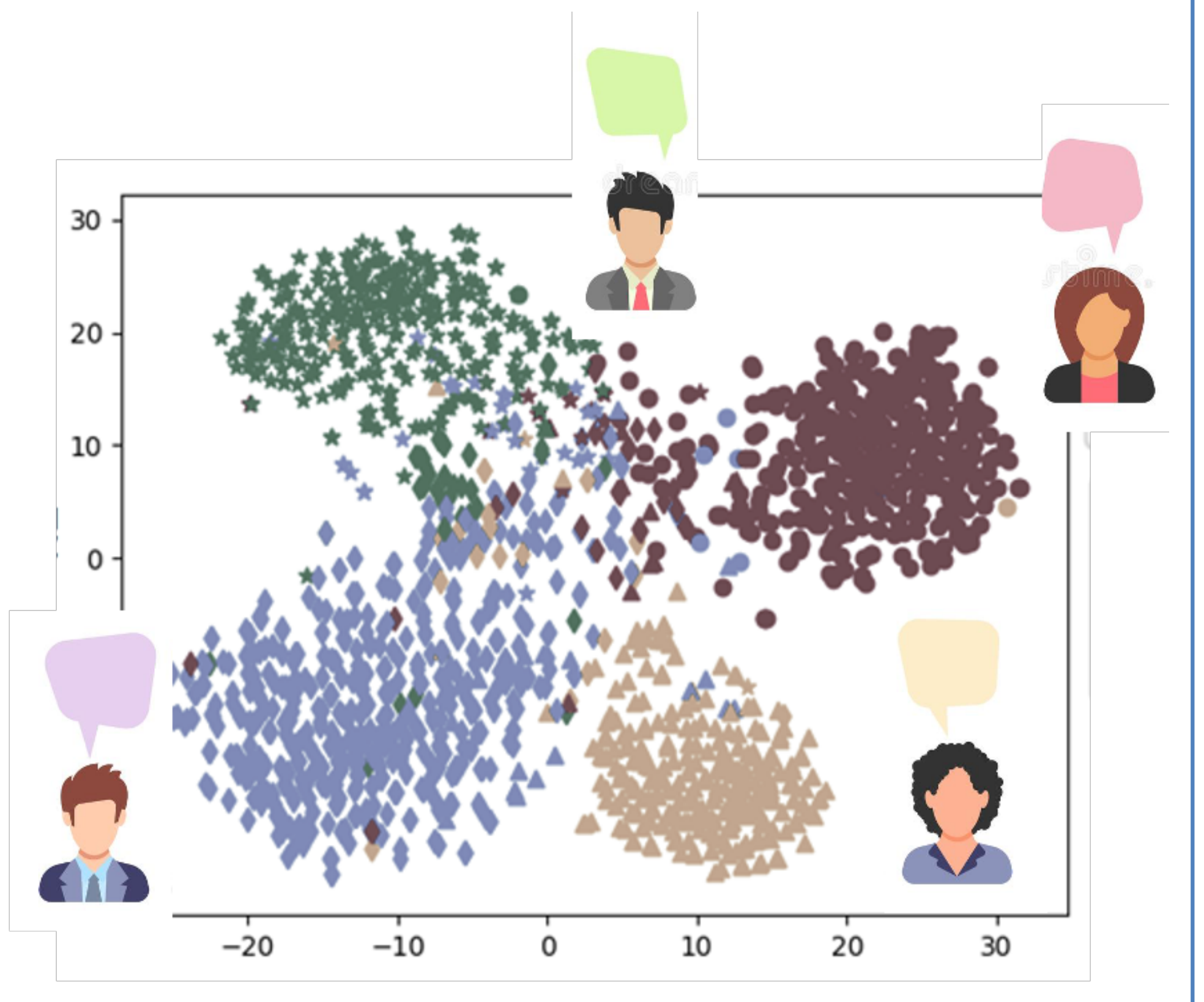
Self-supervised clustering model

- This is a multi-step approach which includes segmentation, feature generation and self-supervised clustering (SSC).
- SSC model improves clustering by learning new representations/features iteratively.

Results & Conclusion

Datasets:
AMI dataset**:** Meeting dataset containing 3-5 speakers, 20-60mins audio
DIHARD dataset: Multi-domain dataset with 1-10 speakers domains ranging from meeting to web videos, 1-10mins audio

Evaluation metric:
Diarization Error Rate (DER) =
Speaker Confusion Error + False Alarm + Miss Rate



2-d plot of learned feature vectors. Each colour indicates a cluster/speaker

- Proposed model helps to separate features in speaker space.
- Reduces the DER by upto 60% compared to baseline models

References

- Snyder et. al., X-vectors: Robust DNN Embeddings for Speaker Recognition, ICASSP, 2018
- Prachi Singh et. al., "Self-supervised representation learning with path integral clustering for speaker diarization", IEEE TASLP 2021
- Prachi Singh et. al. ,"Self-Supervised Metric Learning with Graph Clustering for Speaker Diarization", IEEE ASRU 2021