AVA: ONLINE CAPTIONING & SPEAKER DIARIZATION

Ava Team
Ava US & France, `https://www.ava.me`
Alexey Ozerov (contact): `alexey@ava.me`

Our mission in Ava is to break communication barriers between deaf and hard of hearing (DHH) users and the rest of the world. Our product, Ava app, is an online captioning system of conversations which is already used by thousands of DHH users to navigate in their everyday life. During a conversation, Ava app allows several users to connect via different devices (e.g. smartphones, laptops) to the same session (in person or remotely). The app leverages automatic speech recognition (ASR) to instantly and accurately transcribe speech of all the participants of the conversation.

However, accurately transcribing the speech of all speakers is not enough to help DHH users navigate in their surrounding world. In a multi-talker environment, it is also crucial to indicate "who is saying what?". This can be achieved via online speaker diarization (Park et al., 2022). Inspired by recent research work (Coria et al., 2021), we have developed a low-latency neural-based speaker diarization approach and integrated it into Ava app.

In this demo, we will demonstrate Ava app with a particular focus on the online speaker diarization functionality. Figure 1 shows an example of Ava app's transcripts of an in-person conversation. Each speaker is represented by its own color. It is also possible for a user to enter names of particular speakers. At Ava AI Lab in Paris, we are currently working on various problems to improve our acoustic front-end such as echo cancellation, speech separation and noise reduction.

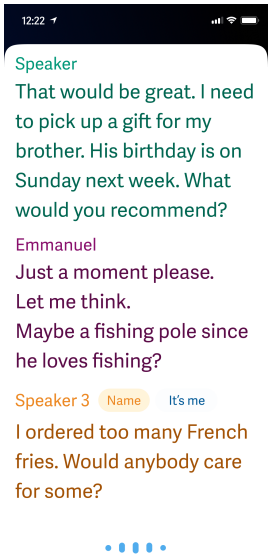

Figure 1: An example screenshot of Ava app. Each speaker is represented by its color, and users may tag speakers.

REFERENCES

Juan Manuel Coria, Hervé Bredin, Sahar Ghannay, and Sophie Rosset. Overlap-aware low-latency online speaker diarization based on end-to-end local segmentation. In *IEEE ASRUW*, 2021.

Tae Jin Park, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J Han, Shinji Watanabe, and Shrikanth Narayanan. A review of speaker diarization: Recent advances with deep learning. *Computer Speech & Language*, 72:101317, 2022.