Meeting Recognition and Tracking

Alex Waibel

Interactive Systems Labs.
Carnegie Mellon University and Karlsruhe University

Abstract

In recent years, we have conducted an active research program aimed at the capture, transcription, tracking, description, review, access, recall and summarization of human-human interaction in meetings. The task extends previous research toward truly natural, unprepared and unconstrained human interaction. The problem is inherently multimodal and involves capturing all available signals. Based on these signals, it involves robust processing, fusion and understanding of the full breadth of human communicative hints and signals, without prior preparation, segmentation or artificial restrictions in recording style.

The processing problems include large vocabulary conversational speech recognition, microphone independence, cross talk, sound source localization, recognition of emotion (from speech and facial expression), identification of participants from speech and face), tracking of topics, summarization from speech, visual eye-gaze, pose and focus of attention tracking. In this talk I will describe the problem, our current research efforts, the databases and evaluation methodologies currently used, and directions for future research and resource requirements.