

KEYNOTE

The road towards inclusive speech technology and why phonetic knowledge and sciences are essential

Odette Scharenborg

TU Delft

Automatic speech recognition (ASR) is increasingly used, e.g., in emergency response centers, domestic voice assistants, and search engines. Because of the paramount relevance spoken language plays in our lives, it is critical that ASR systems are able to deal with the variability in the way people speak (e.g., due to speaker differences, demographics, different speaking styles, and differently abled users). ASR systems promise to deliver objective interpretation of human speech. Practice and recent evidence however suggest that the state-of-the-art ASRs struggle with the large variation in speech due to e.g., gender, age, speech impairment, race, and accents. The overarching goal in our research is to develop inclusive speech technology, i.e., speech technology that works for everyone irrespective of their voice, language, or the way they speak. In this talk, I will present systematic experiments aimed at quantifying, identifying the origin of, and mitigating bias in state-of-the-art ASRs on speech from different “diverse”, typically low-resource, groups of speakers, and I will argue why phonetic knowledge and the phonetic sciences are essential in developing inclusive speech technology.