

Speaker Embedding and Recent Advances in Speaker Recognition Evaluation

Kong Aik Lee¹

概要 : Automatic speaker recognition is the task of identifying or verifying an individual's identity from their voice samples using machine learning algorithms, without any human intervention. Voice is a combination of physical and behavioral biometrics characteristics. The physical features of an individual's voice are based on the shape and size of the vocal tract, mouth, nasal cavities, and lips that involve in producing speech sound. The behavioral aspects, which include the use of a particular accent, intonation style, pronunciation pattern, choice of vocabulary and so on, are associated more with the words or lexical content of the spoken utterances. Speaker recognition has seen significant advancements over the past few decades, giving rise to the successful introduction of commercial products. With the advent of Big Data and the resurrection of data-hungry modeling techniques such as artificial neural networks, more recently the research focus has shifted from a more controlled scenario towards larger and more realistic speaker in the wild scenarios. The latest cycle of NIST evaluations (SRE'18), which in addition to traditional conversational telephony speech (CTS) involves voice over IP (VOIP) data as well as audio extracted from online videos. This talk aim to present recent technical advances in speaker recognition and NIST SRE'18 from NEC perspective.

経歴

Kong Aik Lee is a Senior Principal Researcher, Biometrics Research Laboratories, NEC Corp., Japan. He received his B.E. (First Class Honours) in Electrical Engineering from University Technology Malaysia (UTM), Malaysia in 1999 and Ph.D. from Nanyang Technological University (NTU), Singapore in 2005. From 2006 to 2018, He was a research scientist at the Human Language Technology Department, Institute for Infocomm Research (I2R), A*STAR, Singapore. His current research interests include speaker recognition and characterization, multilingual recognition and identification, speech analysis and processing, machine learning and digital signal processing. He also serves as an Editorial Board Member for Elsevier Computer Speech and Language, and Associate Editor for IEEE/ACM Transactions on Audio, Speech and Language Processing. He is a senior member of IEEE.

¹ NEC Corp, Central Research Laboratories, Japan