

NEWS RELEASE

September 29, 2016

**Development of New Acoustic Signal Processing Technology that
Simultaneously Recognizes the Voices of Multiple Speakers
Making it possible to convert conversations into text using speech recognition**

Researchers from the laboratory of Associate Professor Nobutaka Ono of the Principles of Informatics Research Division, National Institute of Informatics, the Research Organization of Information and Systems, an Inter-University Research Institute Corporation (NII; Director General, Masaru Kitsuregawa; Chiyoda Ward, Tokyo) and researchers including Professor Shoji Makino from the Life Science Center of Tsukuba Advanced Research Alliance (TARA), University of Tsukuba (President, Kyosuke Nagata; Tsukuba City, Ibaraki Prefecture) have developed a new acoustic signal processing technology that makes it possible to separate individual voices from the overlapping voices of multiple speakers recorded using multiple devices, and thereby, to simultaneously recognize multiple voices. This result was announced by Specially Appointed Assistant Professor Keiko Ochi from Ono Laboratory/Principles of Informatics Research Division, NII, at Interspeech, a leading international conference in the field of speech communication, held this month in San Fransisco, USA. The technology allows speech recognition in situations such as meetings where multiple speakers are talking at the same time, without using any special equipment.

Speech recognition has improved greatly in recent years. However, in multi-speaker environments such as conversations or meetings, recognition deteriorates due to the voices of different speakers overlapping, and this is a major problem. With the aim of solving this problem, research is being carried out on "sound source separation" technology that separates the voices of each speaker in situations where the voices of multiple speakers are mixed together.

When sound is recorded using multiple recording devices, not only will the recording start time of the signals recorded by each device be different but the sampling frequency (frequency when converting sound pressure signals to discrete signal time series) will differ slightly for each device. Conventional sound source separation technology cannot deal with these kinds of signals, and therefore, a special piece of equipment called a microphone array is needed to make a simultaneous recording using multiple microphones.

However, the research group from NII and the University of Tsukuba has developed a new signal processing technology that synchronizes multiple signals after they have been recorded asynchronously using multiple devices.

By combining this technology with high-speed blind source separation^(*1) technology developed previously in the Ono Laboratory to separate conversations in which the voices of multiple speakers are intermingled into individual voices, and then employing speech recognition, the researchers have succeeded in greatly improving speech recognition performance in a multi-speaker environment.

A possible example of application of this technology is an automatic minutes-taking system for meetings, as described below:

- ① Recording: Meeting participants each bring their own smartphone and use it to record the conversation during the meeting.
- ② Synchronization: The speech recorded by each participant's smartphone is not synchronized, and so the signal processing technology developed through this joint research is used to automatically synchronize the speech at an accuracy within 1 millisecond.
- ③ Separation: The speech recorded by each smartphone will include voices other than that of the smartphone's owner, so blind source separation technology is used to separate the voices of each participant.
- ④ Recognition: The separated voices are converted into text using speech recognition technology.

Research is also underway at NII to implement this technology as a Web-based GUI system.

These studies are supported by a JSPS Grant-in-Aid for Scientific Research (JP16H01735).

END

[Contact information for inquiries from the media]

National Institute of Informatics

Communications

TEL:03-4212-2164 FAX:03-4212-2150

E-mail : media@nii.ac.jp

University of Tsukuba

Public Relations Office

TEL: 029-853-2039 FAX : 029-853-2014

E-mail : kohositu@un.tsukuba.ac.jp

(*1) Blind source separation: Blind source separation is a method of estimating individual sound sources using only recorded mixed speech, without using position information of speakers or smartphones.