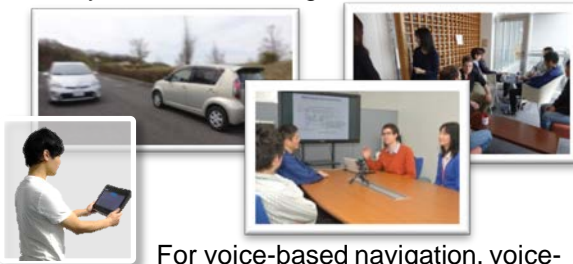


Abstract

While **automatic speech recognition (ASR) technology** has been recently greatly improved and increasingly come into our daily lives, **its use scene is still limited**. For example, it is still difficult for ASR systems to perform reliably in noisy environments such as street, cafés and exhibition halls, especially when they are used in the situation where microphone(s) and users' mouth are not close enough, and thus the recorded speech signal contains **noise and reverberation**. Through this poster, we introduce some of **our recently developed key fundamental technologies** such as distortion-less speech enhancement and deep learning-based ASR technologies to address such issues, with which we won the CHiME-3 Challenge, an international program for evaluating the performance of speech recognizers in noisy outdoor public areas. In the future, these technologies will serve as a key to help us enhance the quality of ASR in smartphones and development of communication robots.

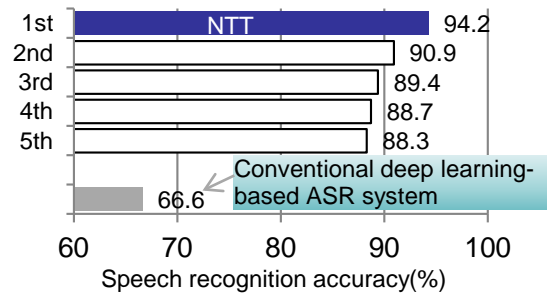
Expanding use scene of ASR technology

To noisy environments, e.g. street and cafés

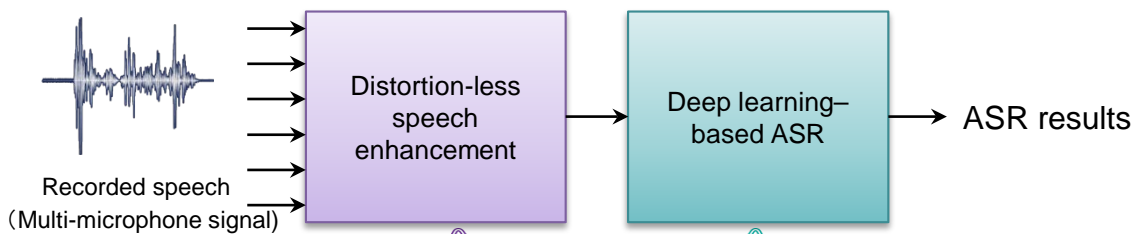


For voice-based navigation, voice-based controls of electric appliances and etc

Achieved the world's best performance



Our technologies



- Capable of **canceling ambient noise and reverberation**, which has detrimental effects on speech recognition performance, without distorting target voices.
- Suitable for the subsequent deep learning-based ASR system

- **Robust representation of speech signals** by using deep convolutional networks based on the network-in-network (NiN) approach
- Capable of **accounting for long-term dependency in word sequences** by using recurrent neural networks

【Reference】

- [1] T. Yoshioka, N. Ito, M. Delcroix, A. Ogawa, K. Kinoshita, M. Fujimoto, C. Yu, W. J. Fabian, M. Espi, T. Higuchi, S. Araki, and T. Nakatani, "The NTT CHiME-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices," in *Proc. ASRU2015*, 2015.
- [2] S. Araki, M. Okada, T. Higuchi, A. Ogawa and T. Nakatani, "Spatial correlation model based observation vector clustering and MVDR beamforming for meeting recognition," in *Proc. ICASSP2016*, 2016.

【Contact】

Shoko Araki & Keisuke Kinoshita Signal Processing Research Group, Media Information Laboratory
E-mail : {araki.shoko, kinoshita.k}@lab.ntt.co.jp