

15

Voice command and speech communication in car

- World's best voice capture and recognition technologies -

Abstract

Our technology can **support a speech command and hand-free communication** even in noisy environment such as road noise without any stresses. Clear speech can be picked up from the noise-mixed sound in order to realize speech command with high accuracy. A lot of computational complexy and memory was required to keep speech quality and reduce only noise so far. This problem can be solved by **using our acoustical knowhow**, moreover, low latency was able to be also achieved. In addition, **a sign of howling was able to be detected rapidly** by combining multiple microphone array. Our goal is **to improve an in-car acoustical environment** by reducing noises which are road noise, engine noise, and any sound from other cars. We will also try to establish an event detection technology in order to **help a driving assistant or an early maintenance** by detecting emergency car or anomalous in sound.

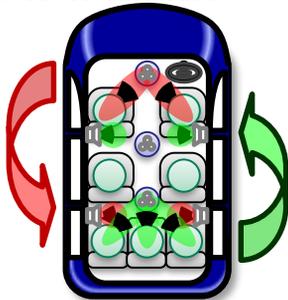
Differences from conventional

IM : Intelligent Microphone
 ASTER : Anti-distortion Suppression of noise with mask-based TransER function estimation

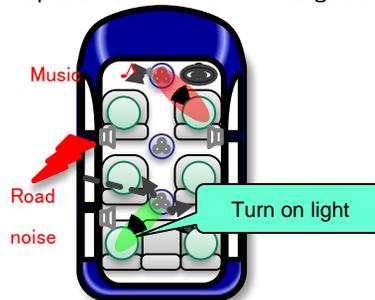
	Conventional NTT technology 1 (IM*)	Conventional NTT technology 2 (ASTER*)	This technology (IM-ASTER)
Abstract	Extracting target sound from noise-mixed sound by combining a linear and non-linear process.	Reducing noise rom noise mixed speech signal while minimizing speech distortion.	Achieving low computational complexity and small amount of memory, and having advantages of IM and ASTER.
Image of process  (Input signal)	 Speech is degraded by too much noise reduction	 Noise is removed while minimizing speech distortion	 Noise is removed while minimizing speech distortion
Minimize distortion	△	○	○
High level noise	○	△	○
Computational complexity / memory	○	×	○

Demonstration

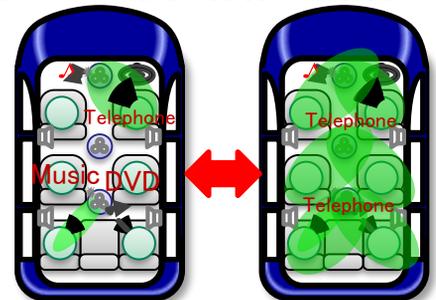
In car communication



Speech command from target seat



Hand-free communication



References

- [1] Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, Y. Haneda, "Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 6, pp. 1240-1250, 2013.
- [2] T. Yoshioka, N. Ito, M. Delcroix, A. Ogawa, K. Kinoshita, M. Fujimoto, C. Yu, W. H. Fabian, M. Espi, T. Higuchi, S. Araki, T. Nakatani, "The NTT CHIME-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices," in *Proc. IEEE ASRU*, pp. 436-443, Dec. 2015.

Contact

Noboru Harada Email: cs-liaison-ml at hco.ntt.co.jp
 Media intelligence laboratory



Innovative R&D by NTT
 Open House 2019