15 Joint signal separation, dereverberation and noise reduction

Extracting voices out of noise & reverberation

Abstract

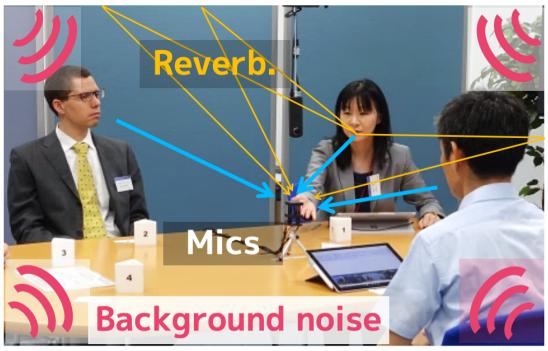
To enable such audio devices as smart speakers to accurately recognize human voices in real-world environments, we must reduce the noise and reverberation from the signals observed by microphones and extract each individual's voice. State-of-the-art (SOTA) technology addresses this problem by sequentially applying the following three techniques: (a) dereverberation, (b) source separation, and (c) denoising. However, SOTA is ineffective in noisy reverberant conditions because all three techniques, (a), (b), and (c), are optimized individually without considering the overall performance. In this exhibit, we introduce a new technology that jointly optimizes (a), (b), and (c) to maximize the quality of the output audio. Our new technology significantly improves the speech recognition performance compared to the SOTA method. It will contribute to a more convenient world where people and computers can interact smoothly in our daily environments, including train stations, streets, and shopping malls.

Speech Extraction from Microphone Recordings

Computers struggle to accurately recognize audio signals recorded by distant microphones

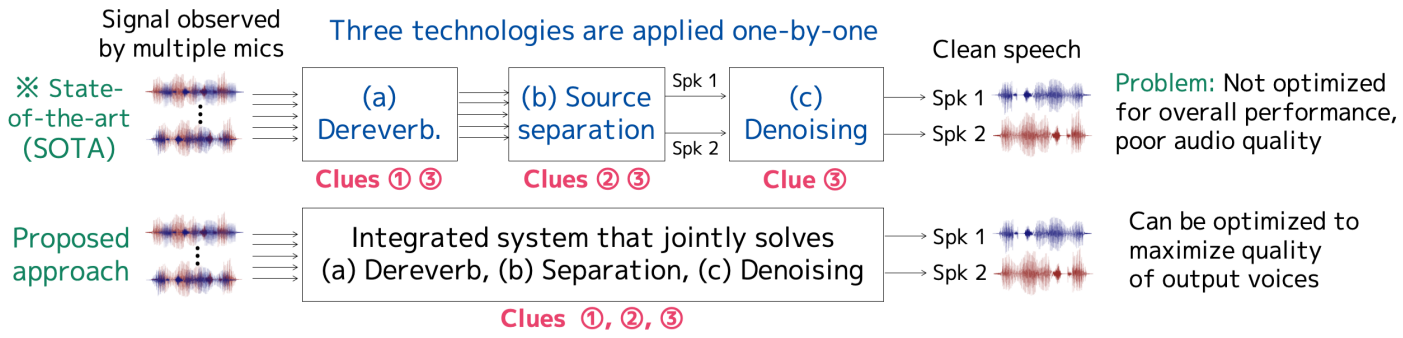
Causes of difficulty	Required technologies
High level of reverberation	(a) Dereverberation
Multi people talking simultaneously	(b) Source separation
High level of background noise	(c) Denoising

⇨ We accurately extract each individual's voice using signal processing technologies



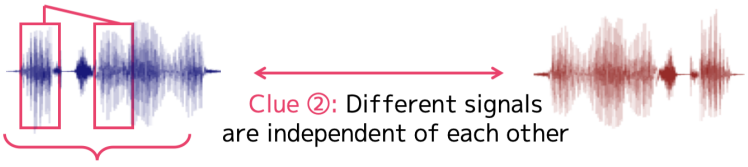
Problem of SOTA technique & our result

※ NTT's CHiME3 technology remains a widely used SOTA technology



Proposed framework unifies (a), (b), and (c)

Clue ① Different time intervals are not similar (independent)



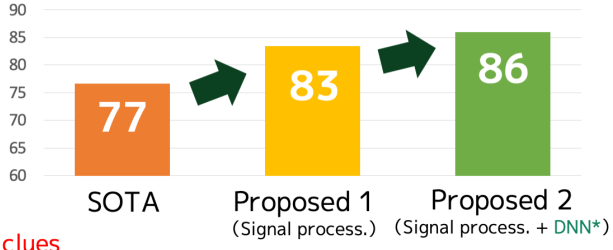
Clue ②: Different signals are independent of each other

Clue ③: Voice intensity is time-varying (nonstationary)

⇨ Integrate (a), (b), and (c) by simultaneously considering all three clues

Speech recognition performance [%]

Speakers	2
Mics	8



*When estimating part of a probabilistic model using DNN

References

[1] T. Nakatani, C. Böddeker, K. Kinoshita, R. Ikeshita, M. Delcroix, R. Haeb-Umbach, "Jointly optimal denoising, dereverberation, and source separation," in *Proc. IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2267-2282, 2020.
[2] R. Ikeshita, T. Nakatani, S. Araki, "Block coordinate descent algorithms for auxiliary-function-based independent vector extraction," in *Proc. IEEE Trans. Signal Process.*, 2021, to appear.
[3] R. Ikeshita, T. Nakatani, "Independent vector extraction for fast joint blind source separation and dereverberation," in *Proc. IEEE Signal Process. Lett.*, 2021, to appear.

Contact

Rintaro Ikeshita / Signal Processing Research Group, Media Information Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp