Yan-min Qian, Chao Weng, Xuan-kai Chang, Shuai Wang, Dong Yu, 2018. Past review, current progress, and challenges ahead on the cocktail party problem. *Frontiers of Information Technology & Electronic Engineering*, 19(1): 40-63. https://doi.org/10.1631/FITEE.1700814

Past review, current progress, and challenges ahead on the cocktail party problem

Key words: Cocktail party problem; Computational auditory scene analysis; Non-negative matrix factorization; Permutation invariant training; Multi-talker speech processing

Corresponding author: Yan-min QIAN
E-mail: yanminqian@tencent.com
ORCID: http://orcid.org/0000-0002-0314-3790

Motivation

- The cocktail party problem, i.e., tracing and recognizing the speech of a specific speaker when multiple speakers talk simultaneously, is one of the critical problems yet to be solved to enable the wide application of automatic speech recognition (ASR) systems.
- Although the processing mechanisms seem clear and related tasks are easy for humans, researchers have found it surprisingly difficult to give machines the same ability
- This paper aims to provide a comprehensive survey of the popular and effective solutions to the cocktail party problem developed in the past two decades, and the remaining difficulties and challenges ahead.

Method

1. Conventional single-channel techniques such as computational auditory scene analysis (CASA), non-negative matrix factorization (NMF) and generative models.
2. Conventional multi-channel techniques such as beamforming and multi-channel blind source separation.
3. The newly developed deep learning-based techniques, such as deep clustering (DPCL), deep attractor network (DANet), and permutation invariant training (PIT).
4. The new techniques developed to improve ASR accuracy and speaker identification in the cocktail party environment.

Conclusions

- We have described past efforts in attacking the cocktail party problem. We can observe that the majority of the efforts focus on the speech separation task, and some of the work targets the speaker tracing and speech recognition.
- Effectively exploiting information in the microphone array, the acoustic training set, and the language itself using a more powerful model.
- A better optimization objective and techniques will be the approach to solving the cocktail party problem.