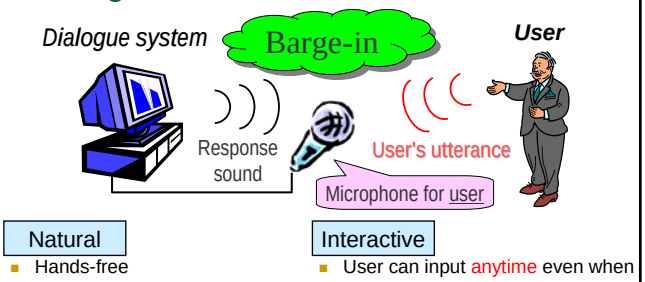
Evaluation of Spoken Dialogue Interface Based on Sound Field Control and Source Separation

Speech and Acoustics Laboratory
D2 Shigeki Miyabe

Overview

- Background
- Conventional methods
  - Acoustic echo canceller
  - MOMNI method
- Proposed method
- Speech recognition experiment
- Conclusion & Future Works

Background



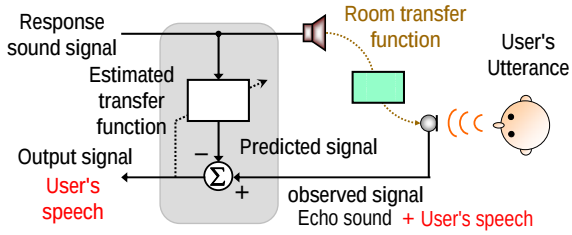
Natural
- Hands-free
- Wearing no special equipments

Interactive
- User can input anytime even when system outputs response sound

⇨ **Degradation of speech recognition performance**

Goal: **Realize the interface for barge-in free spoken dialogue system**

Conventional Method: Acoustic Echo Canceller

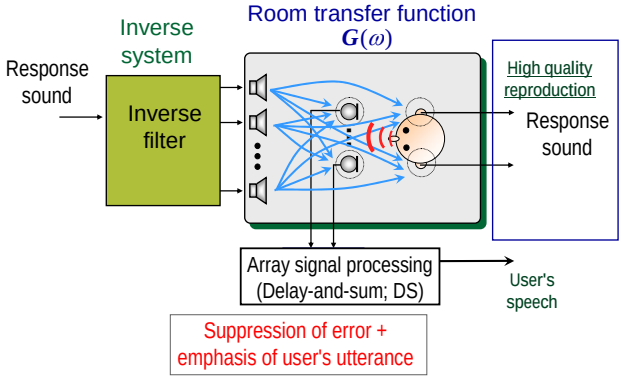


Problem

Failure estimation in barge-in situation causes lowing of performance

⇨ Double-talk detection is necessary — Difficult in noisy environment

Our previous approach: Multiple-Output and Multiple-No-Input (MOMNI) Method



Suppression of error + emphasis of user's utterance

Problems and approach

- Features of MOMNI method
  - Advantage
    - Robust control without adaptation and double-talk detection is unnecessary
    - High-quality sound reproduction
  - Disadvantage
    - Requires too many loudspeakers for stable control with many microphones

Approach

Introduce more efficient array signal processing with fewer microphones

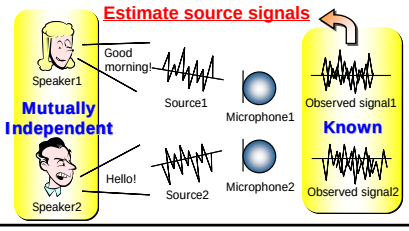
- Features of various array signal processing

Processing	Double-talk detection	Performance
Delay-and-Sum (DS)	○ Unnecessary	× Requires many microphones
Adaptive beamformer	× Necessary	○ Good
Blind source separation	○ Requires no detection	○ Good

So we try to introduce blind source separation (BSS)

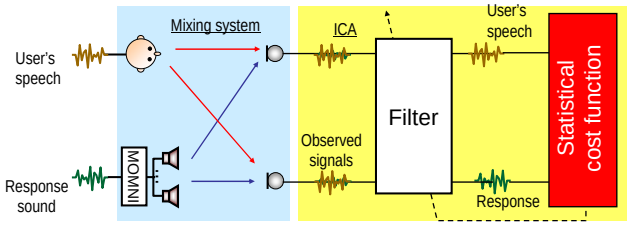
Blind source separation (BSS)

- BSS：A technique to estimate source signals
  - Only with observed signals
  - Without any priori information
- One of the most promising BSS strategy is that based on Independnt component analysis (ICA)
  - Assuming statistical independence between sources



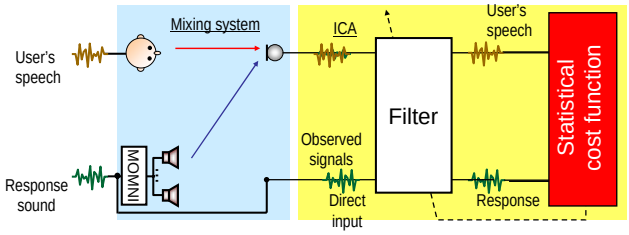
Straightforward idea: MOMNI + BSS

- BSS can separate sources only with observed signals
  Number of sound sources are two
  ⇔ Independent component analysis (ICA) can be performed with two observed signals



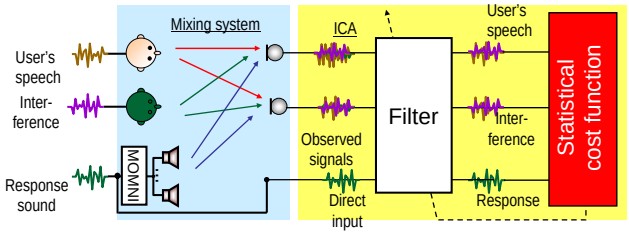
Proposed method: MOMNI + Semi-blind source separation

- One of sound sources is already available
  - Input one observed signal and response sound source
  - One of output (response sound) is already optimal in beginning of optimization



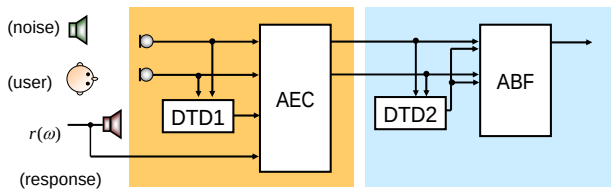
Proposed method: MOMNI + Semi-blind source separation

- In existence of additional interfering noise
  - By using one more microphone, user's speech can be separated from the noise just as in ordinary BSS.
  - Noise reduction without double-talk detection



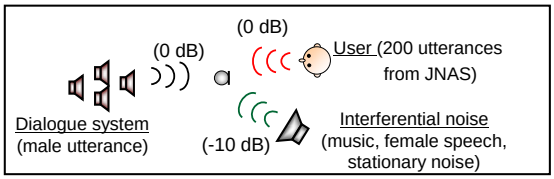
Competitive conventional method: AEC + Adaptive Beamformer (ABF)

- Compare the performance of unsupervised proposed method with conventional method with double-talk detection (DTD).
- To simulate ideal behavior of DTD, we gave true single-talk durations to adaptive filters.



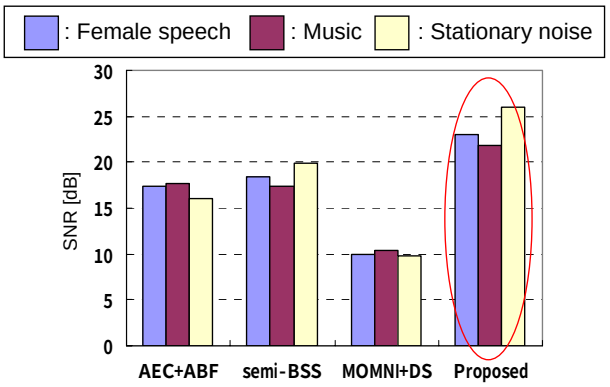
Simulation

- Evaluation in double-talk and noisy environment

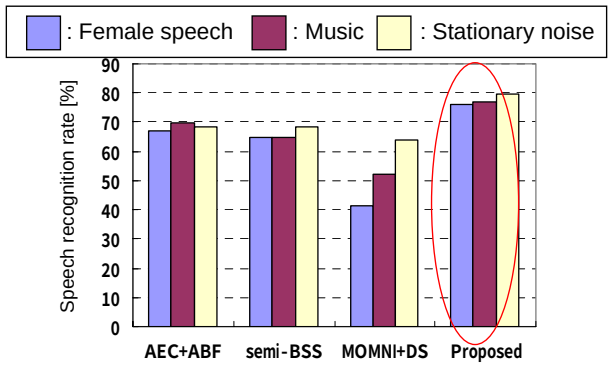


- Compared with
  - AEC+ABF (using ideal DTD, 1 loudspeaker, 2 microphones)
  - Semi-BSS (1 loudspeaker, 2 microphones
  - MOMNI + DS (8 loudspeakers, 3 microphones)
  - Proposed method:
    MOMNI + semi-BSS (8 loudspeakers, 2 microphones)

Processed SNR



Speech Recognition Performance



Conclusions

- We proposed semi-blind source separation and used it in spoken dialogue interface with sound field reproduction
- Semi-blind source separation can eliminate both response sound and interfering noise
- Proposed method shows higher performance in speech recognition experiment

Future works

- Expansion of the area where sound source is reproduced