Slide 1

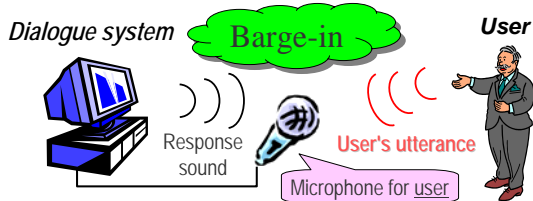# Barge-in Free Spoken Dialogue Interface Using Sound Field Control and Semi-Blind Source Separation

Speech and Acoustics Laboratory
D1 Shigeki Miyabe

Slide 2

Overview

- Background
- Conventional method
- Proposed method
- Speech recognition experiment
- Conclusion & Future Works

Slide 3

Background

Dialogue system — Barge-in — *User*

Response sound

User's utterance

Microphone for user

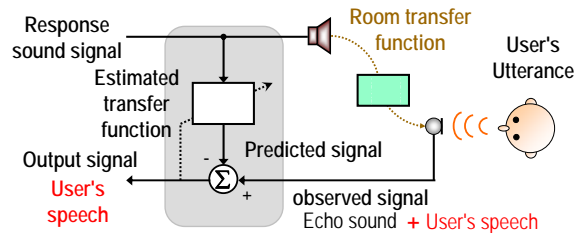| Natural	Interactive
- Hands-free - Wearing no special equipments	- User can input anytime even when system outputs response sound

➡ **Degradation of speech recognition performance**

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

Slide 4

Conventional Method: Acoustic Echo Canceller

Response sound signal

Room transfer function

User's Utterance

Estimated transfer function

Output signal
User's speech

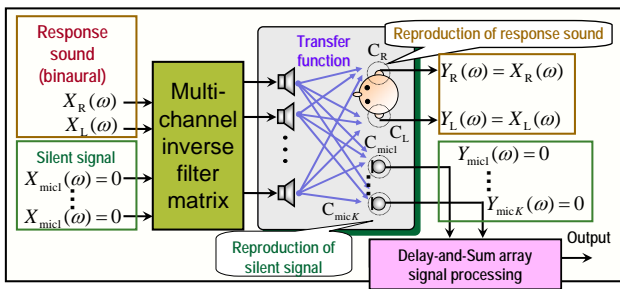Predicted signal

observed signal
Echo sound + User's speech

Problem

Failure estimation in barge-in situation causes lowing of performance

➡ double-talk detection is necessary | Accurate detection is difficult

Slide 5

Proposed Method: Multiple-Output and Multiple-No-Input (MOMNI) Method (Hinamoto et al., 2003)

- Combination of sound field reproduction and microphone array
  Sound field reproduction … eliminates response sound before observation
  Delay-and-sum array … emphasizes user's speech and improves stability

Response sound (binaural)
$X_R(\omega)$
$X_L(\omega)$

Multi-channel inverse filter matrix

Silent signal
$X_{mic1}(\omega) = 0$
$X_{mic1}(\omega) = 0$

Transfer function

C_R
C_L
C_{mic1}
C_{micK}

Reproduction of response sound
$Y_R(\omega) = X_R(\omega)$
$Y_L(\omega) = X_L(\omega)$
$Y_{mic1}(\omega) = 0$
$Y_{micK}(\omega) = 0$

Reproduction of silent signal

Delay-and-Sum array signal processing

Output

Slide 6

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	Unnecessary	Requires many microphones
Adaptive beamformer	Necessary	Good
Blind source separation	Requires no detection	Good

So we try to introduce blind source separation (BSS)!

Simple idea: MOMNI + BSS

- Separate sources only with observed signals
  - Number of sound sources are two
  - Independent component analysis (ICA) can be performed with two observed signals
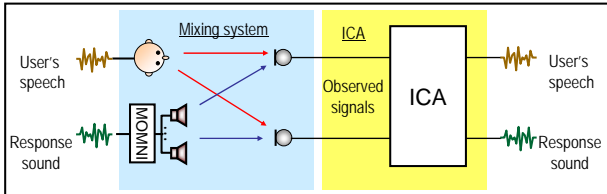
Proposed method: MOMNI + Semi-blind source separation

- Problems of MOMNI + BSS
  - Post processing is difficult because of complex directivity
- One of sound sources is already available!
  - Input one observed signal and response sound source
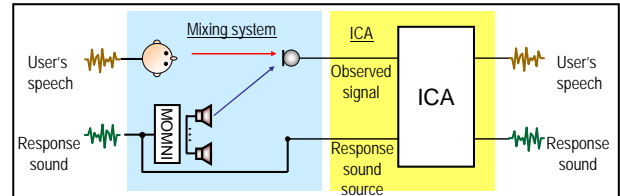
Advantages of semi-blindness

- Reduction of control points
  - Improves stability of sound field control
- Making problem easier by giving a part of answer
  - Good initial value can be made
 - No post-processing is required

Simulation

- Content of experiment
  - Speech recognition experiment
- Plan of experiment
  - Simulation of spoken dialogue system using impulse response measured in real-environment
  - Comparison of robustness of control
- Comparing with
  - Acoustic echo canceller
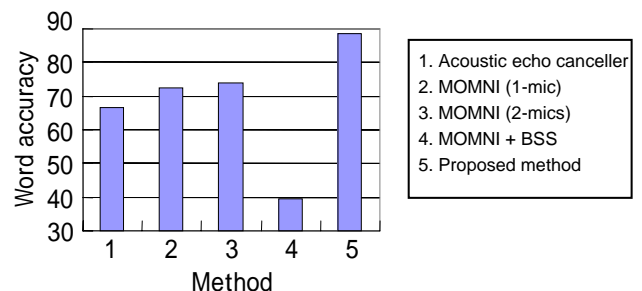 - MOMNI method
  - MOMNI+BSS

Condition of Measuring Impulse Responses

Evaluate robustness against fluctuation of transfer function
Measure impulse responses caused fluctuation.

- Placing interference to cause fluctuation
  - a mannequin at 12 positions
  - imitating 12 kind of fluctuations
- Reverberation time
  - 160 ms
- Sampling frequency
  - 16 kHz

Results

Legend:
1. Acoustic echo canceller
2. MOMNI (1-mic)
3. MOMNI (2-mics)
4. MOMNI + BSS
5. Proposed method
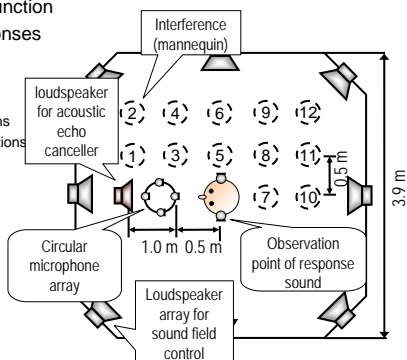
- MOMNI + BSS degrades the performance (because of difficulty in post-processing
- Proposed method performs best

2

Conclusions

- We proposed semi-blind source separation and used it in spoken dialogue interface with sound field reproduction
- Proposed method shows higher performance in speech recognition experiment

Future works

- Elimination of environmental noises by increasing the number of microphones