ESTIMATION OF SOUND SOURCE POSITIONS USING A SURROUNDING MICROPHONE ARRAY

T. Okamoto, R. Nishimura, Y. Iwaya

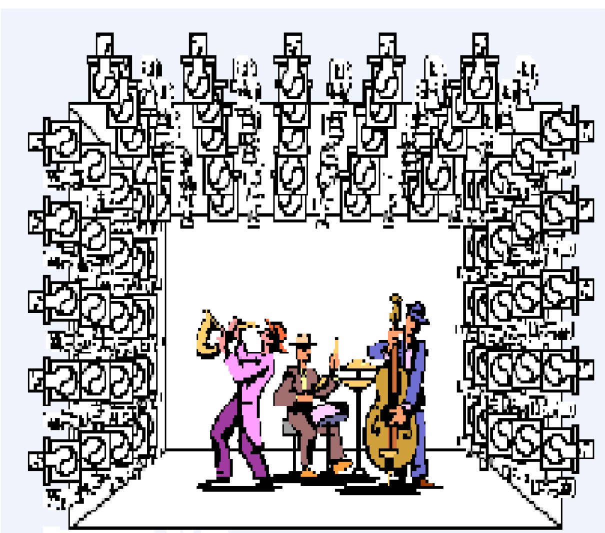
Research Institute of Electrical Communication,
Graduate School of Information Sciences,
Tohoku University

Outline

- Introduction
- Surrounding Microphone Array
- Estimation of Sound Source Positions
  - MUltiple Signal Classification (MUSIC) method (R. O. Schmidt *et el.* 1986)
  - Spatial Smoothing Technique (T. Shan *et al.* 1985)
  - Developing RAP-MUSIC
- Performance Evaluation
- Conclusion

Introduction

- FIR^3 (FIeld Recording, Recognition and Reproduction) system to record a sound field for later reproduction



Reconstructing the sound information in another place at another time

FIR^3 Prototype

Recording Room

Surrounding microphone array (157 ch)

B&K Type 4591 (157 ch)

Sound information

Reproduction Room

Surrounding speaker array (157 ch)

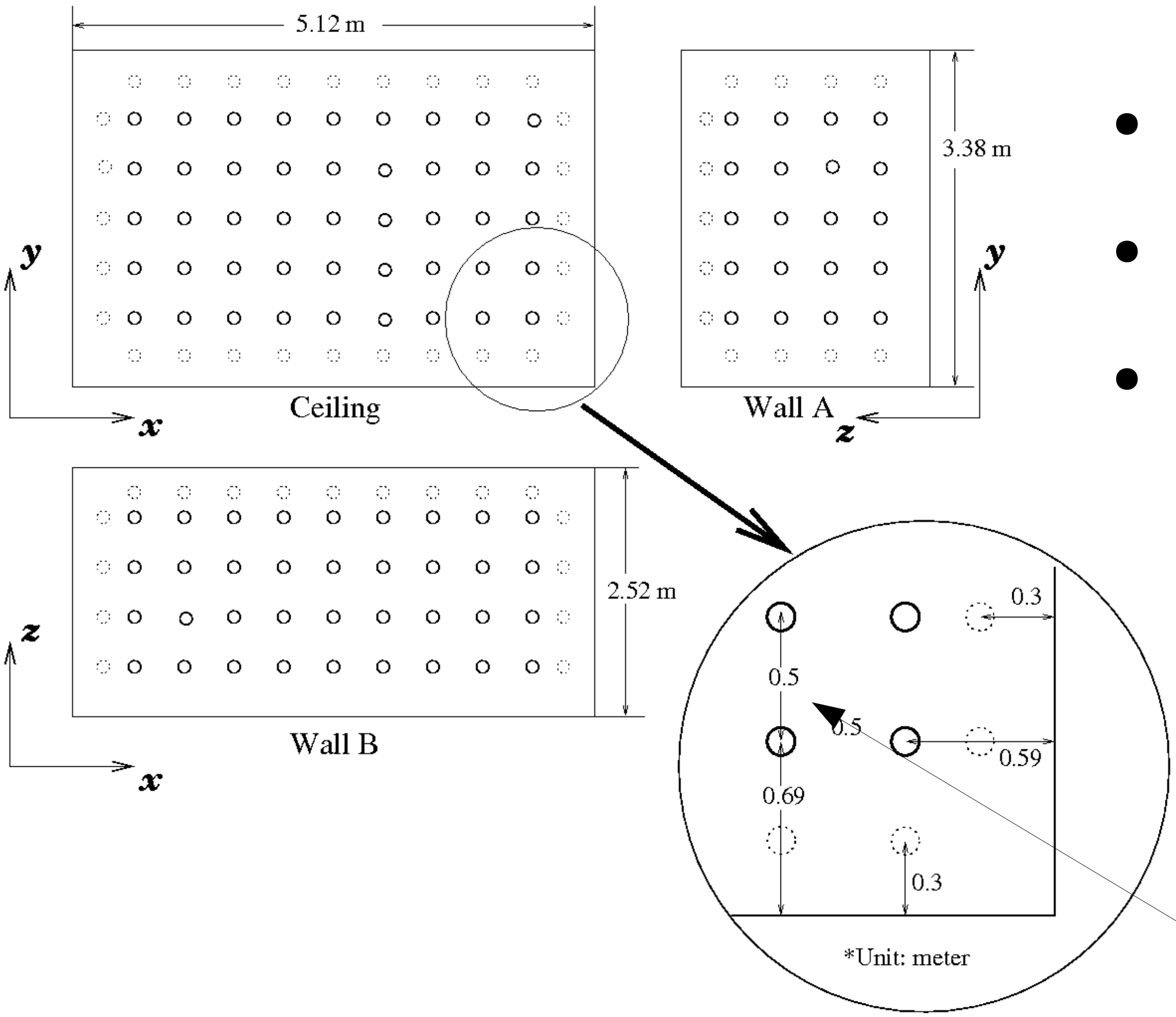
FOSTEX FE38E (157 ch)

Reverberant time of both rooms : 0.15 s

Today's our presentation,

Estimation of Sound Source Positions using a Surrounding Microphone Array in a Reverberant Sound Field

Surrounding Microphone Array



Arrangement of microphones

- Ceiling 45 ch
- Wall A 20 ch × 2
- Wall B 36 ch × 2

157 ch

Spacing between microphones 50 cm

Estimation of Sound Source Positions

- MUSIC method (R. O. Schmidt *et al.* 1986)

Input Signal

$$\boldsymbol{x}(\omega, t) = [x_1(\omega, t), x_2(\omega, t), \cdots, x_n(\omega, t)]^T$$

Spatial Correlation Matrix

$$\boldsymbol{R} = \mathrm{E}[\boldsymbol{x}(\omega, t)\boldsymbol{x}^H(\omega, t)]$$

Extraction →

Base Vector of the orthogonal complements from $\boldsymbol{R}$

$$\boldsymbol{V}$$

Transfer Function between microphone i and search point (x, y, z)

$$a_i(\omega, x, y, z) = \frac{1}{r_i} e^{j\omega \frac{r_i}{c}}$$

→ Steering Vector

$$\boldsymbol{a} = [a_1, a_2, \cdots, a_n]^T$$

MUSIC method

$$P(x, y, z) = \frac{1}{\| \boldsymbol{V}^H \boldsymbol{a}(x, y, z) \|^2} \quad (1)$$

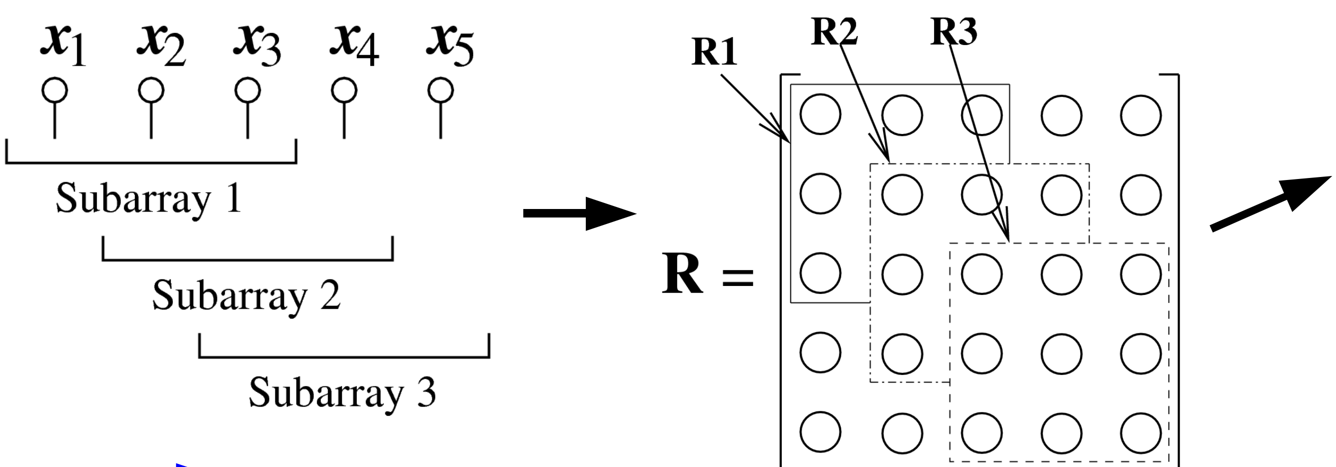
P has a peak because $\boldsymbol{a}$ is orthogonal to $\boldsymbol{V}$ if $\boldsymbol{a}$ fits the source positions

Estimation of Sound Source Positions in Reverberant Sound Field

- MUSIC method (R. O. Schmidt *et al.* 1986)

➡ Negative estimation in reverberant sound field !!!

- Spatial Smoothing Technique (T. Shan *et al.* 1985)
  - Dividing the array into N subarray and calculating subarray's spatial correlation matrices



$$\bar{\boldsymbol{R}} = \sum_{n=1}^{N} \frac{1}{N} \boldsymbol{R}_n \quad (2)$$

MUSIC

➡ Reducing negative effect of reflected sounds

RAP-MUSIC (Rearrangement And Pre-smoothing for MUSIC)

- RAP-MUSIC (Re-arrangement And Pre-smoothing for MUSIC)
  - Making virtual channels from 4 channels at their median point
  - Converting 4 signals at the real channels into 4 signals at the virtual channels by both of amplitude and phase correlation

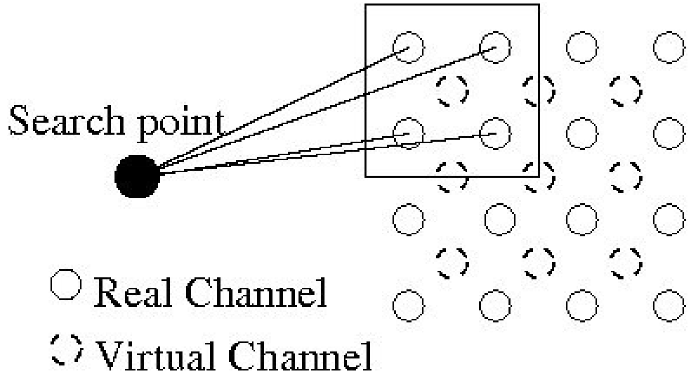
$$x_i'(t) = \frac{r_i}{r_i'} x_i(t) e^{-j\omega\left(\frac{r_i - r_i'}{c}\right)} \quad (3)$$

  - Averaging the signals at virtual channels after synchronizing the phase

    ||

    (Delay-and-Sum beamforming at virtual channels)

  - Applying MUSIC method to averaged signals



Keeping the phase of each signal at the virtual channel

➡ Reducing negative effect of reflected sounds more effectively !!!

Delay-and-Sum > Averaging spatial correlation matrices

Performance Evaluation

Experiments of Estimation of Sound Source Positions

- Experimental Condition
  - Source 1 : male voice (a/i/kya/ku) 63.7 dB Lceq
  - Source 2 : female voice (i/chi/yu/u) 50.4 dB Lceq
  - Sampling frequency : 48000 Hz
  - Time length of input signal : 1 s
  - Hamming window (512 points)
  - Frequency band : 500 ~ 8000 Hz
- Evaluation value of the peak of P

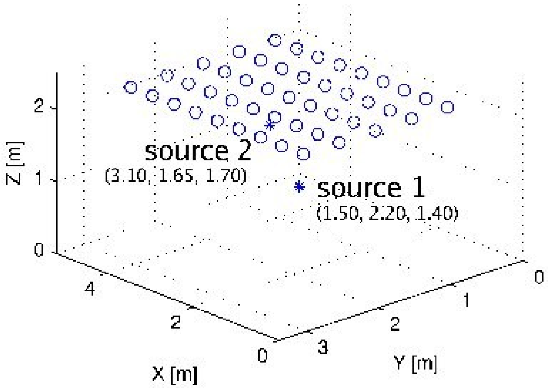
$$k = \frac{P \text{ at each search position}}{P \text{ at the source position}} \qquad (4)$$

Experimental Condition

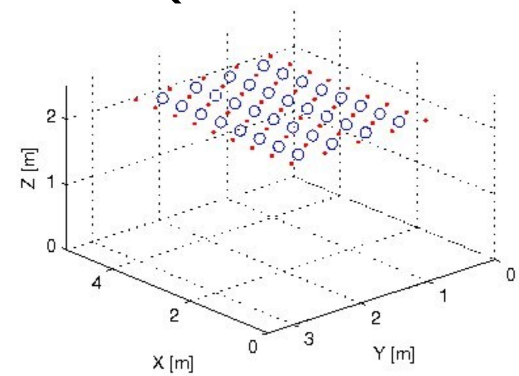
Using two equilateral 32-hedron point source loud-speakers
(7 cm in diameter)

Arrangements of Microphones

- 2-dimensional arrangement (Pattern 1)

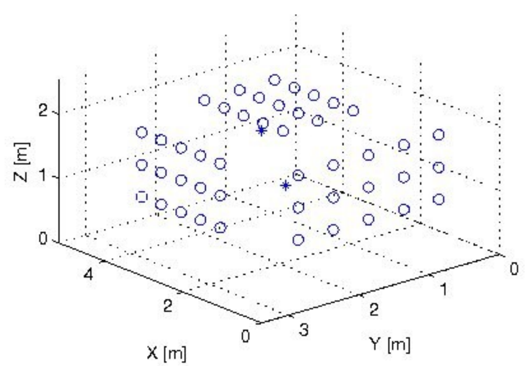


Music (45 ch)

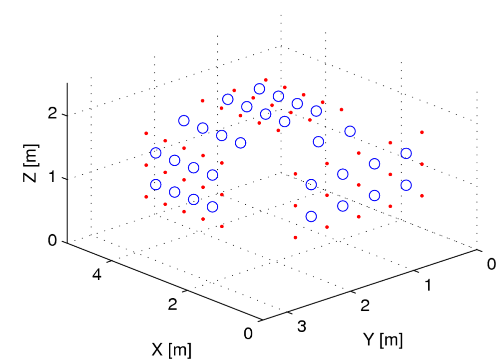


Spatial Smoothing Technique / RAP-MUSIC(32 ch)

- 3-dimensional arrangement (Pattern 2)



Music (45 ch)

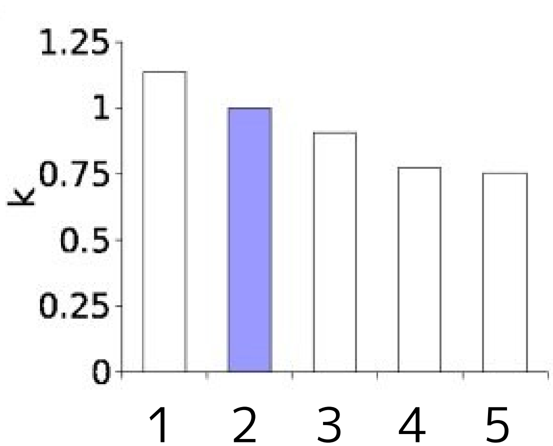


Spatial Smoothing Technique / RAP-MUSIC(30 ch)

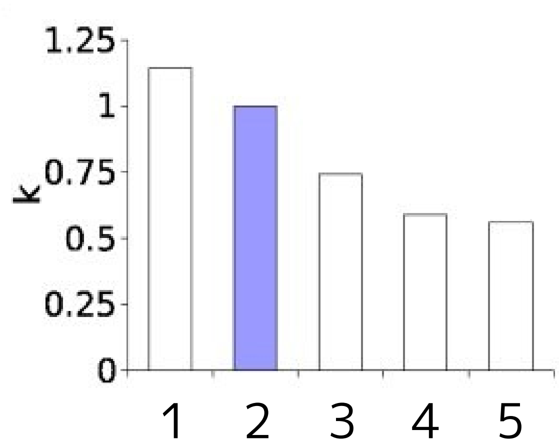
Experimental Results

- Comparison of the performances of three methods for Source 1 (Pattern 1)

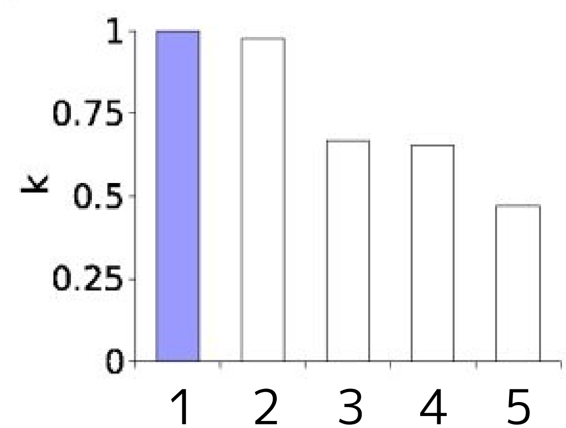
MUSIC



Spatial Smoothing



RAP-MUSIC



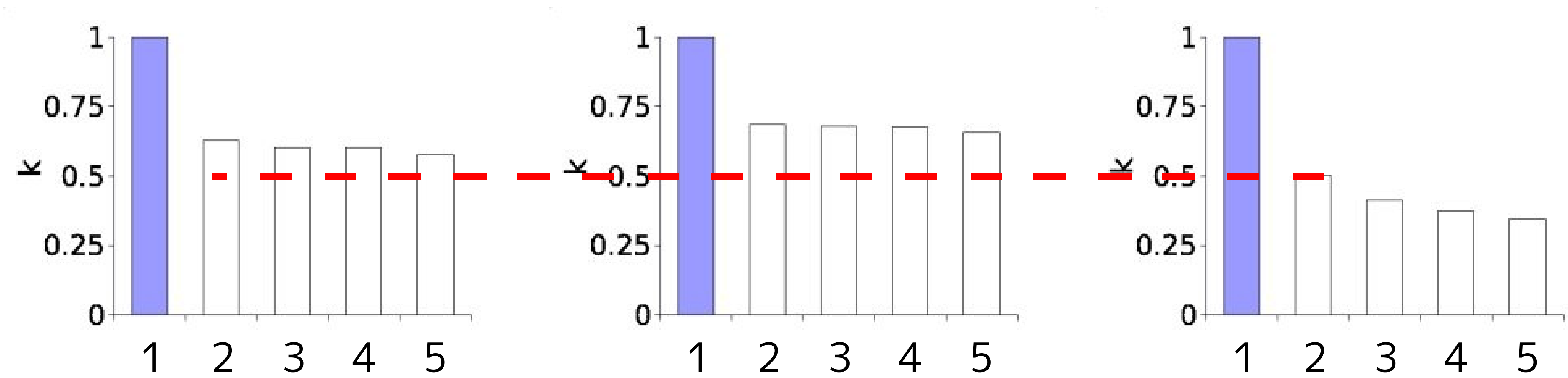
↓

Only RAP-MUSIC succeeded in estimating correct position

Experimental Results

- Comparison of the performances of three methods for Source 1 (Pattern 3)

MUSIC　　Spatial Smoothing　　RAP-MUSIC



↓

RAP-MUSIC is superior to MUSIC and Spatial Smoothing Technique in a Reverberant Environment

Conclusion

- Developing an effective estimation method of sound source positions using the surrounding microphone array for FIR^3
  - Proposing RAP-MUSIC to reduce the influence of reverberant waves