National Institute of Information and Communications Technology

Performance Evaluation of 3D Sound Field Reproduction System Using a Few Loudspeakers and Wave Field Synthesis

Munenori Naoe[12], Toshiyuki Kimura[12],

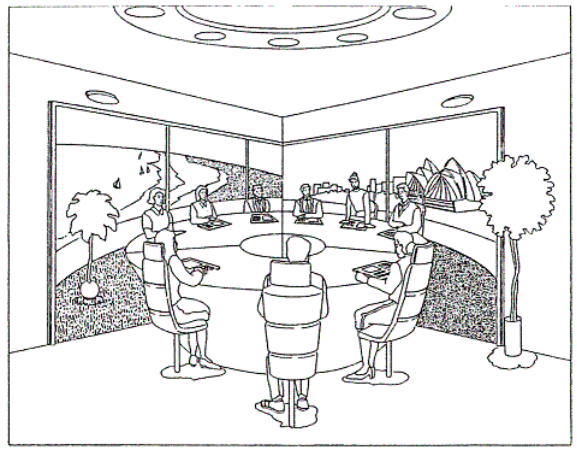
Yoko Yamakata[1] and Michiaki Katsumoto[1]

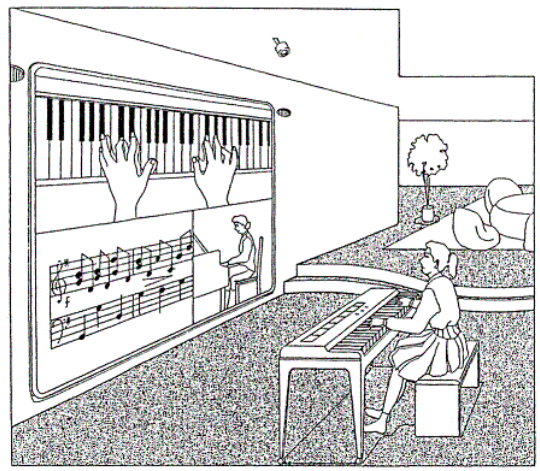
[1]Universal Media Research Center, National Institute of Information and Communications Technology

[2]Graduate School of Engineering, Tokyo University of Agriculture and Technology

3D Sound Field Reproduction System

- More realistic sensation than conventional systems
  - TV-phone, 5.1ch audio
- Tele-conference
  - People in different places feel **as if they have a meeting in the same room**
- Tele-ensemble
  - People in different places feel **as if they play a music in the same concert hall**

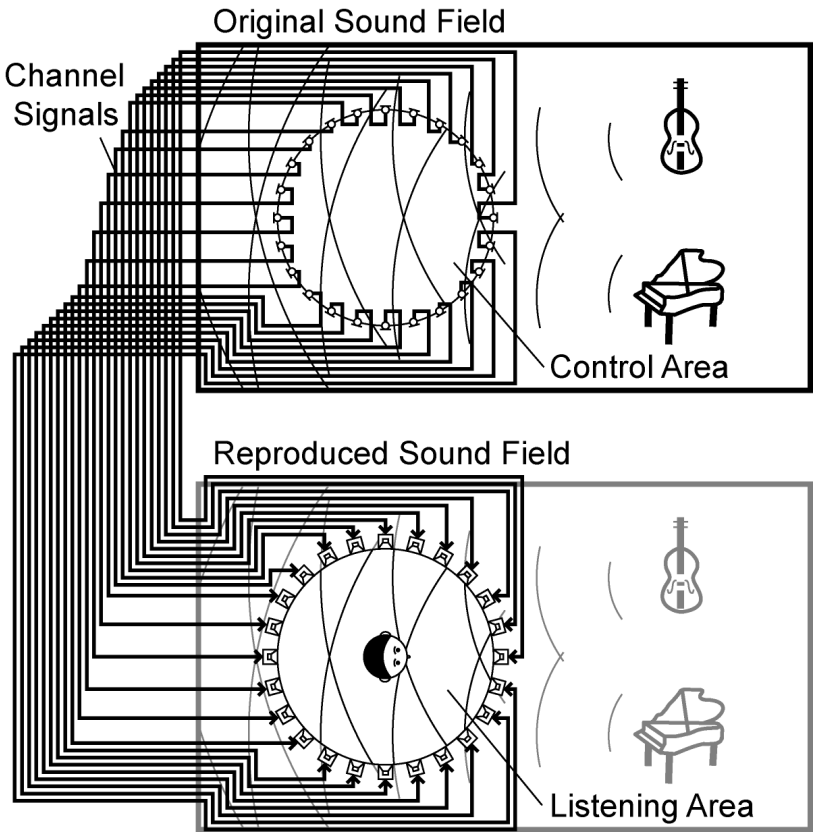




Wave field synthesis is used

System Using Wave Field Synthesis

- Original sound field
  - Sound is recorded by the microphone array
- Reproduced sound field
  - Recorded sound is played by the loudspeaker array
  - Wave fronts are synthesized based on Huygens' principle (Kirchhoff-Helmholtz integral equation)
- Feature
  - Listeners can turn their heads while listening to a sound



Conventional 3D Sound System

- Ise *et. al.* at ICA2007
  - Based on boundary surface control principle
    - Based on Kirchhoff-Helmholtz integral equation
  - Use 70 microphones and 70 loudspeakers
  - Because loudspeakers are visible in the listener's field of vision, it is very difficult to construct an audio-visual system

The number of microphones and loudspeakers should be reduced in order to construct an audio-visual system

Aim of Study

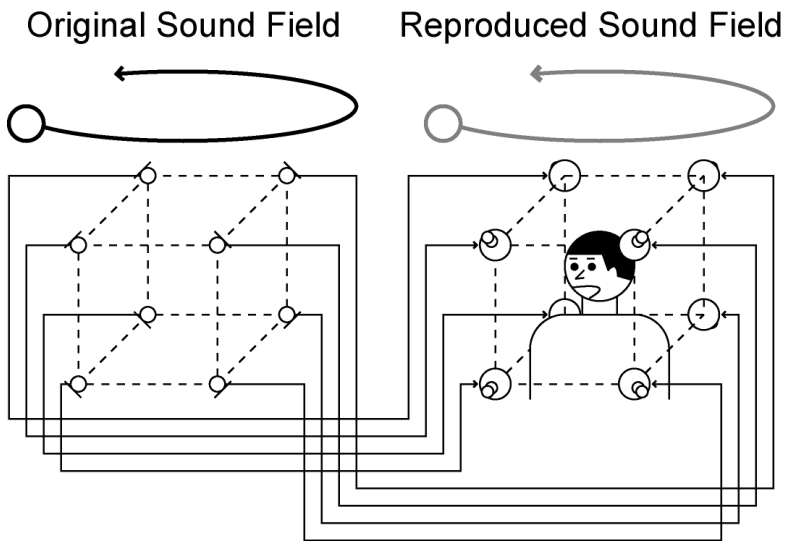
- New 3D sound field reproduction system is proposed
  - 8 microphones and 8 loudspeakers
  - Wave field synthesis technique

    ↓

  - Prevent the loudspeakers from appearing in the listener's field of vision
- The auditory capability of the proposed system is evaluated
  - Localization test

Diagram of Proposed System

- Sound is recorded by 8 microphones
- Recorded sound is played by 8 loudspeakers
- 3D sound field is reproduced
  - When the sound source is moving above the microphone array, the listener feels that the sound image is moving above their head

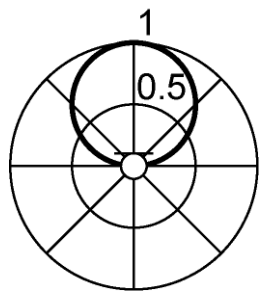


Synthesis of Multi-channel Signals

- Signals $x_i(n)$ are synthesized on a computer
  - i=1...8

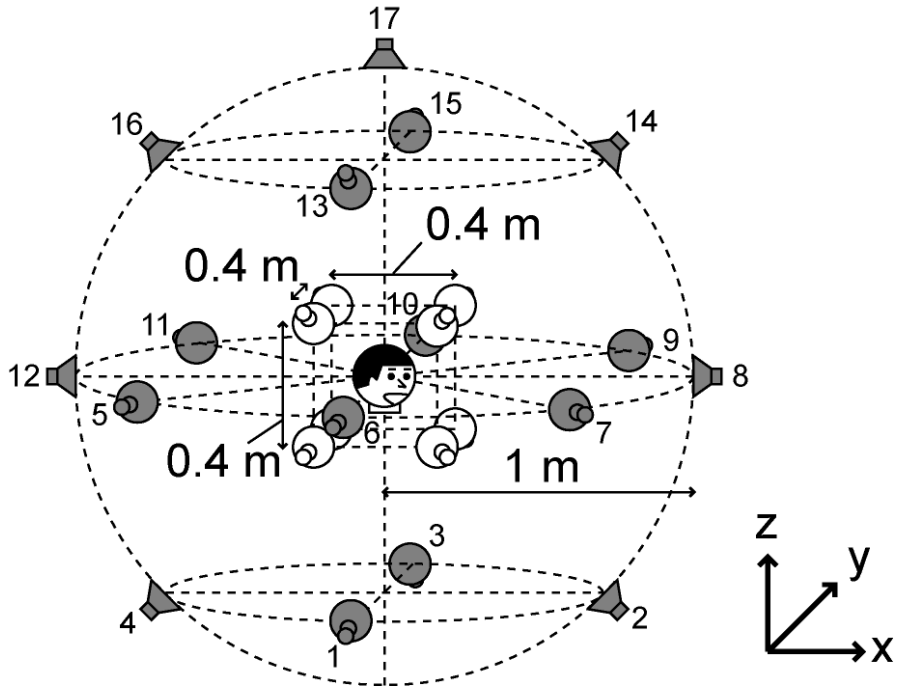
$$x_i(n) = \frac{D_i}{d_i} s\left\{ n - \mathrm{round}\left(\frac{d_i F_S}{c} \right) \right\}$$

  - $s(n)$: Sound source signal
  - F_s(=48kHz): Sampling frequency
  - c(=340m/s): Sound velocity
  - d_i: Distance between the sound source and the ith microphone
  - D_i: Directivity of the ith microphone
    - Shotgun Directivity



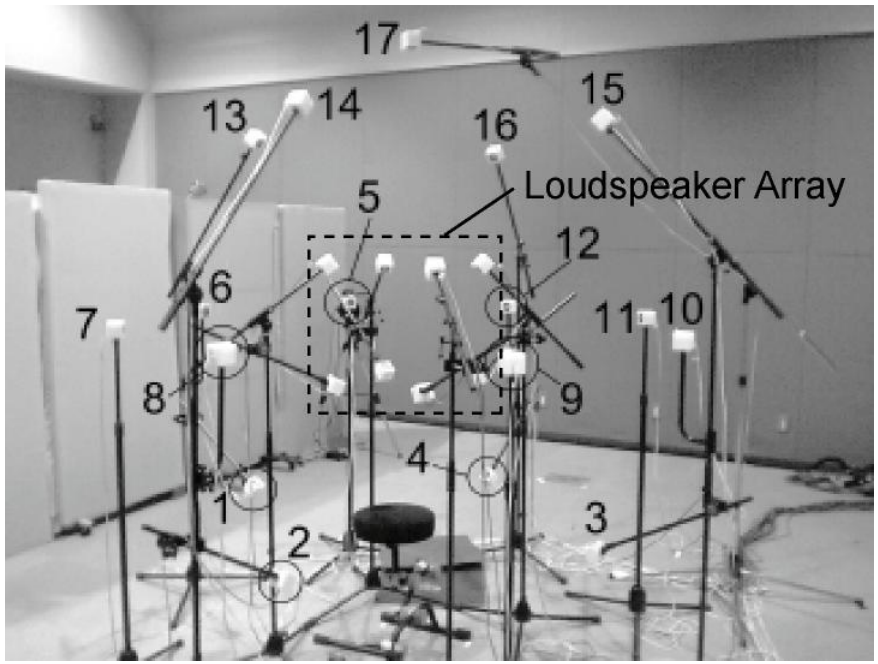
Experimental Environment

- Listening position
  - Center of a sphere
- 25 loudspeakers
  - 8 loudspeakers (white)
    - For loudspeaker array
    - At the vertex of a cube having sides measuring 0.4 m
  - 17 loudspeakers (grey)
    - For control condition
    - On a sphere with a radius of 1 m



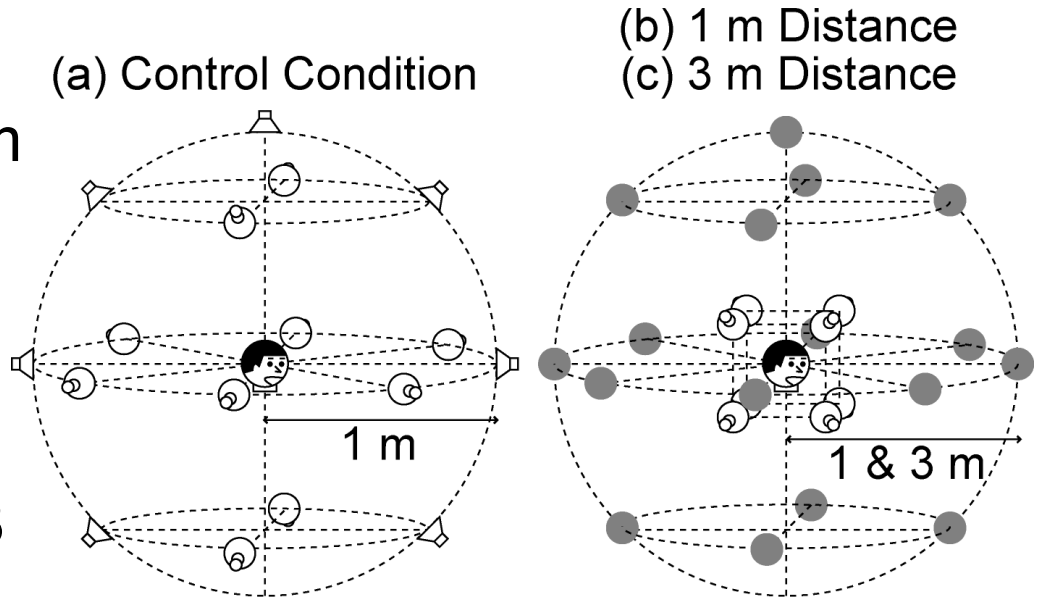
Setup of Loudspeakers

- Experimental room conditions
  - Reverberation time: 180 ms
  - Background noise level: 23 dB(A)
  - Sound pressure level: 60 dB(A) (listening position)



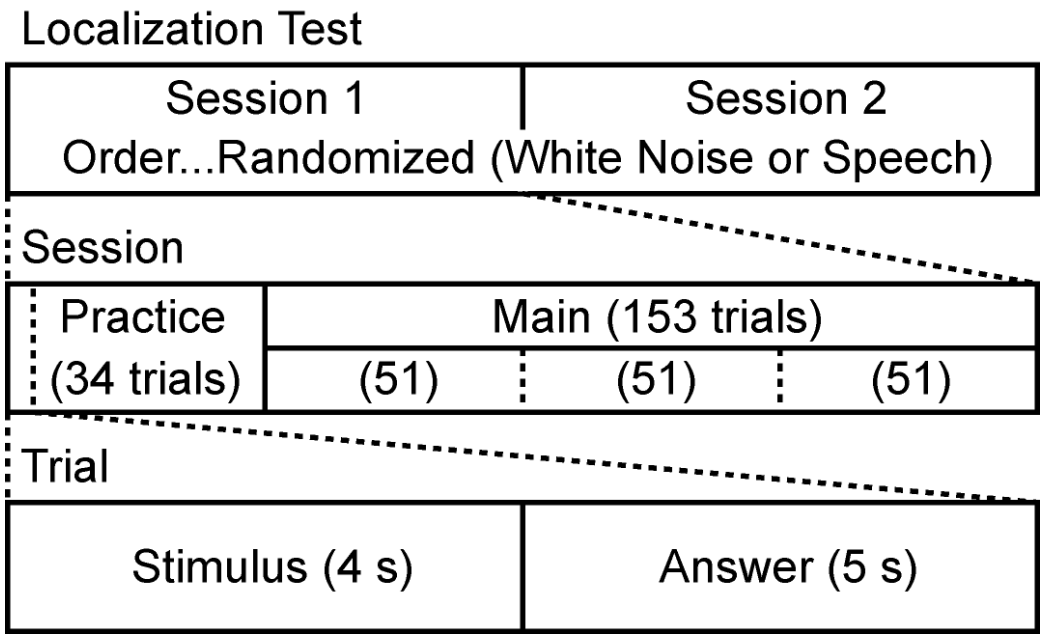
Experimental Condition

- Control condition (a)
  - One sound source was presented from one loudspeaker
- Conditions (b)&(c)
  - 8 channel signals were played from 8 loudspeakers
  - One synthetic sound image was presented
- Listeners reported the direction of sound sources and sound images



Experimental Flowchart

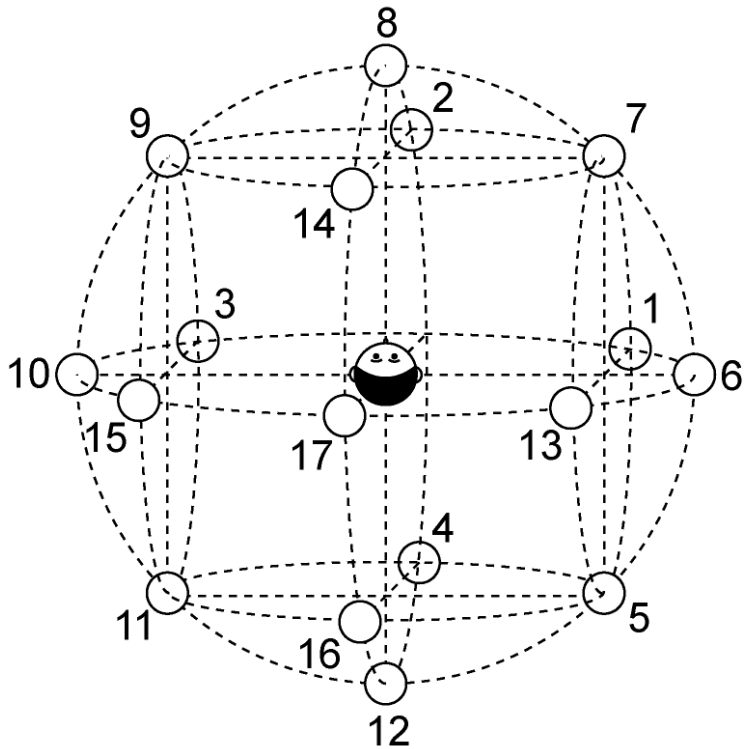
- The number of listeners
  - 6 males and 1 female



	Element	Note
Practice (34)	= 2 conditions × 17 directions	Conditions (a) & (b)
Main (153)	= 3 conditions × 17 directions × 3 repetitions	Conditions (a)–(c)

Experimental Procedure

- Instruction
  - Listeners report the perceived direction of sound
  - Listeners list the number of the direction in an answer sheet
  - Listeners are allowed to turn their heads freely while listening to the sounds



Relation between the perceived directions and direction numbers

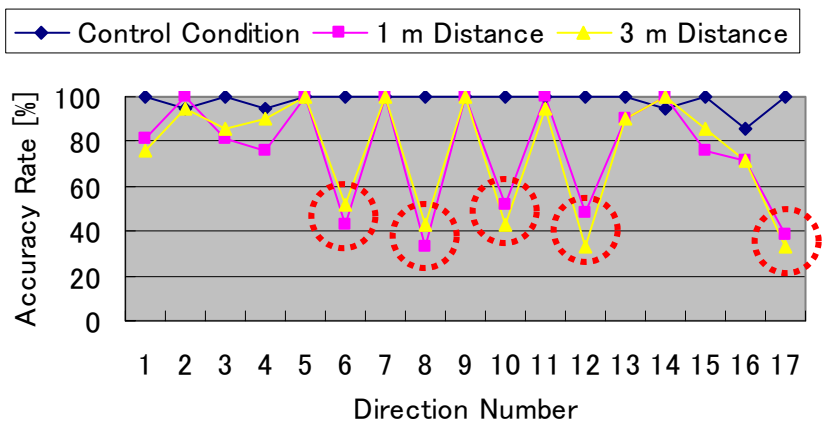
Results

- Accuracy rate
  - Five directions...Lower than control condition
  - Others...Almost the same as control condition

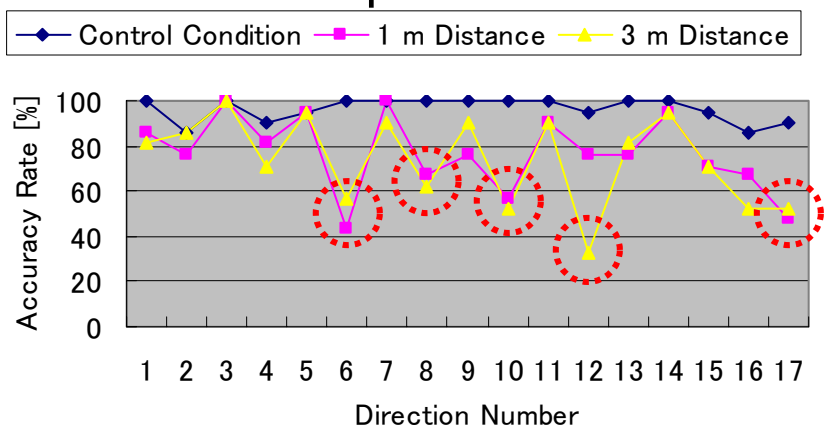
$$\text{Accuracy rate}\,[\%] = \frac{\text{The number of correct answers}}{\text{The number of presentations}}$$

The auditory capability of the proposed system is enough in 12 directions

White noise



Speech



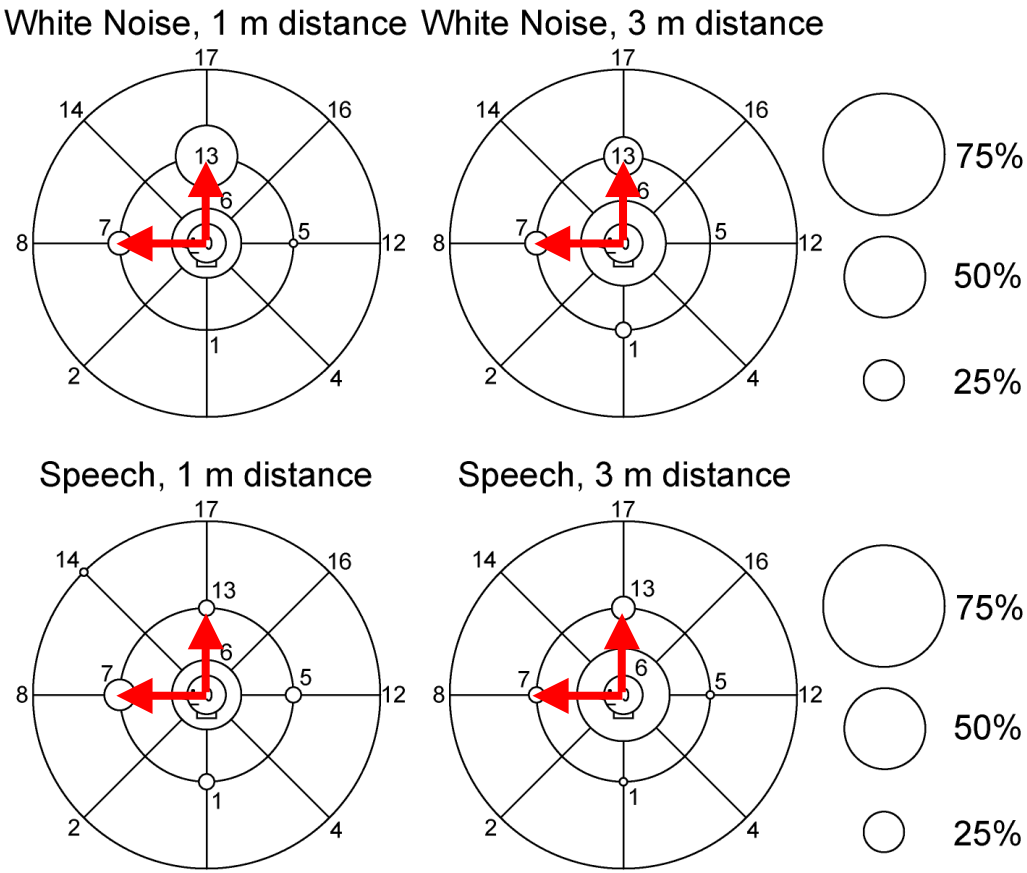
Answer Rates of Five Directions

- Five directions
  - Left direction (No. 6)
  - Front direction (No. 8)
  - Right direction (No. 10)
  - Behind direction (No. 12)
  - Upper direction (No. 17)
- Definition of answer rates

$$\text{Answer rate}\,[\%] = \frac{\text{The number of answers}}{\text{The number of presentations}}$$

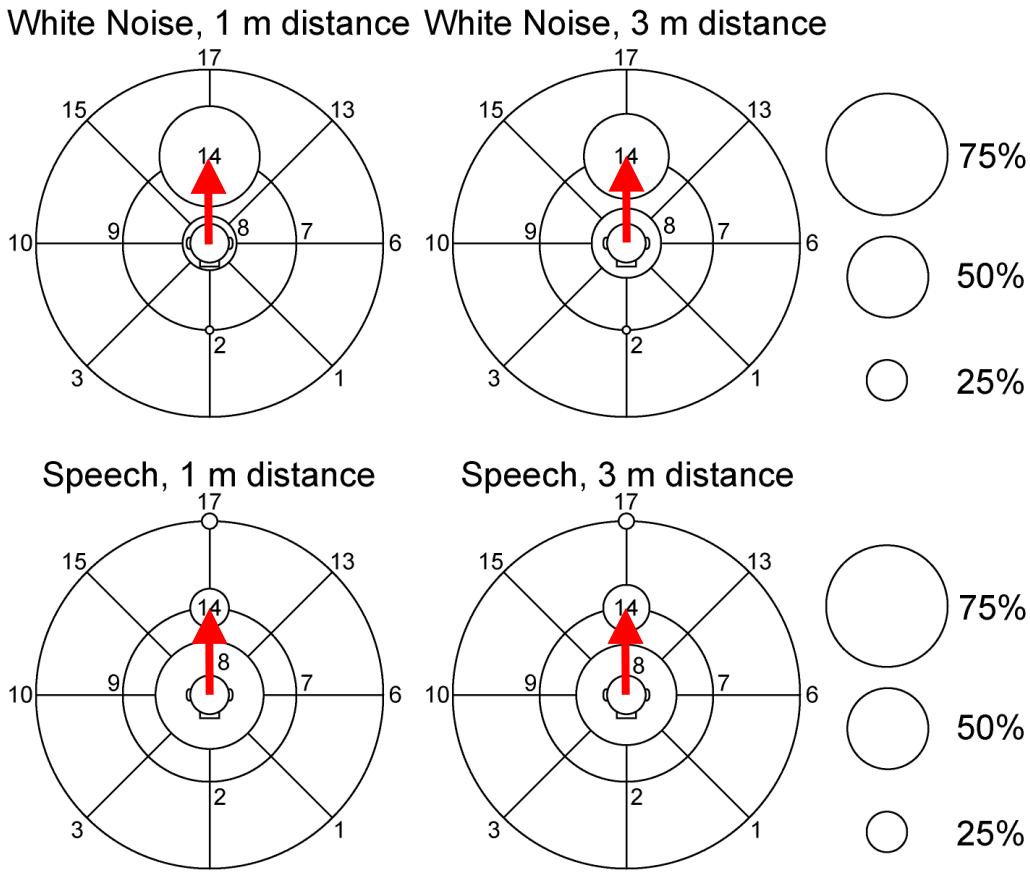
Answer Rates in Right Direction

- Sound images are biased
  - Forward & upper directions



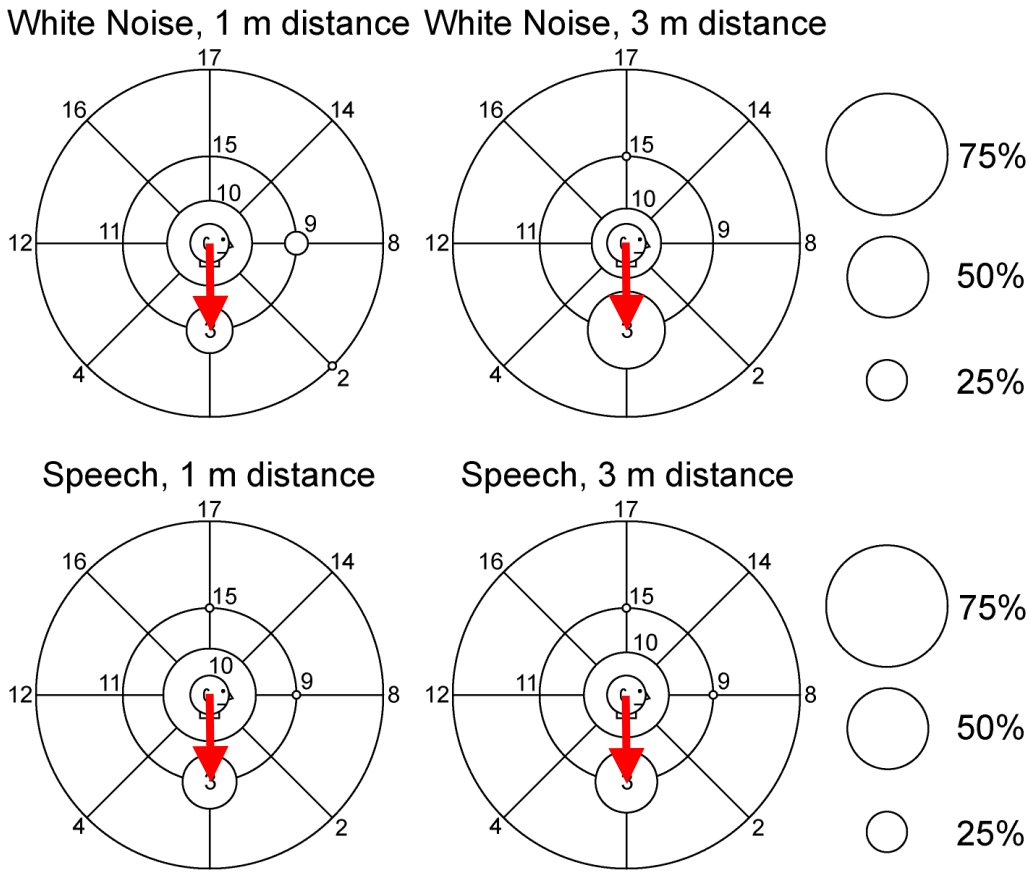
Answer Rates in Front Direction

- Sound images are biased
  - Upper direction



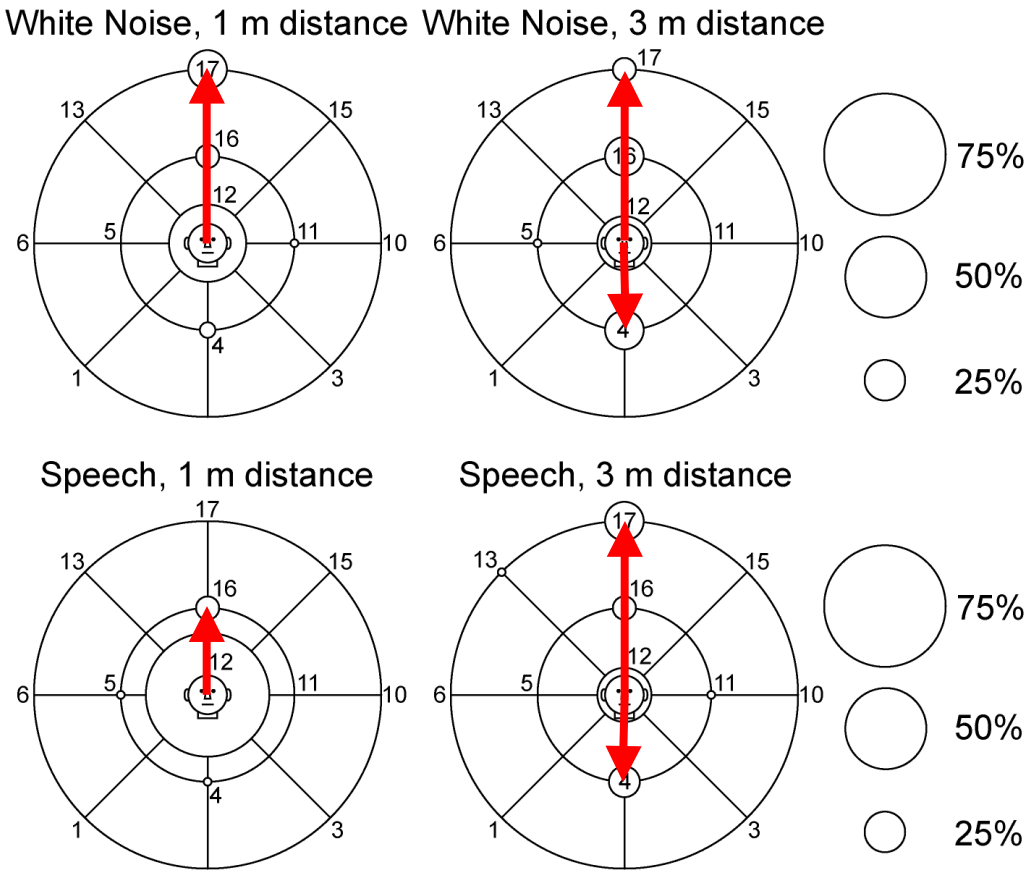
Answer Rates in Left Direction

- Sound images are biased
  - Downward direction



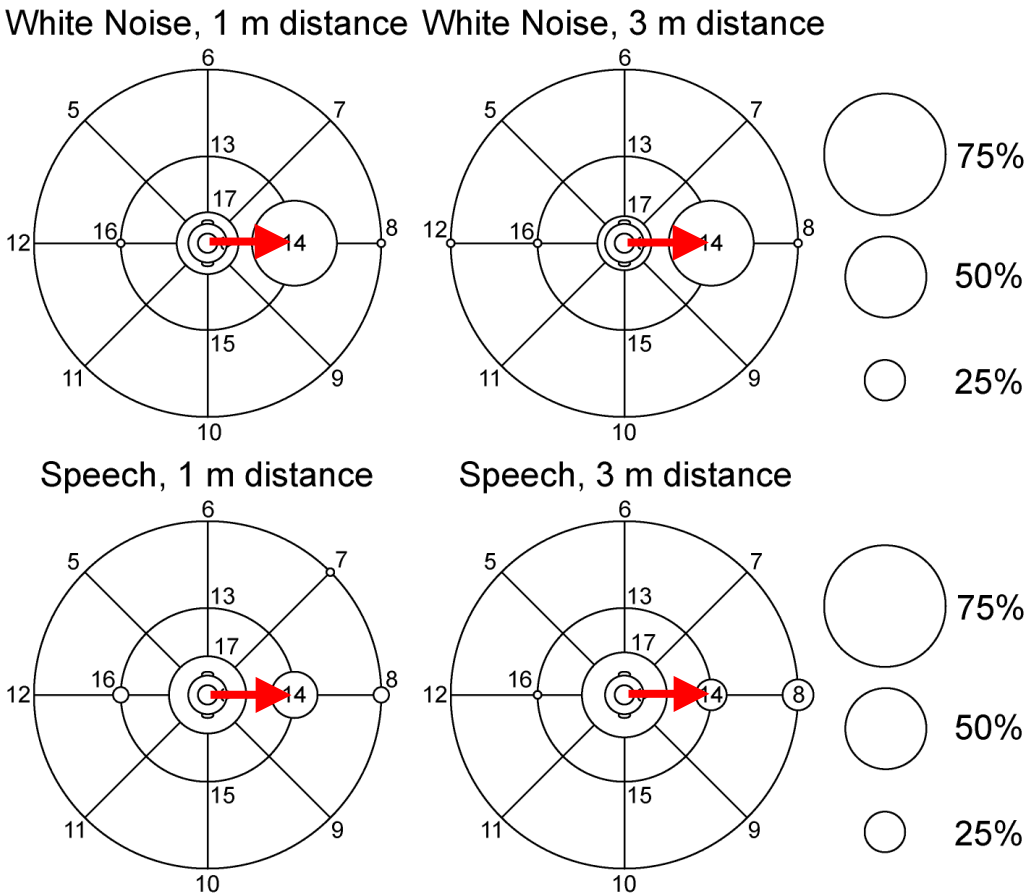
Answer Rates in Behind Direction

- Sound images are blurred
  - Vertical direction



Answer Rates in Upper Direction

- Sound images are biased
  - Forward direction



Conclusions

- New 3D sound field reproduction system was proposed
  - 8 microphones and 8 loudspeakers
  - Wave field synthesis technique
- The auditory capability of proposed system was evaluated by the localization test
  - Performance of proposed system was enough in 12 directions of 17 directions used in the test
  - The bias and blur of sound images occurred in remaining 5 directions
- Future works
  - Improvement of performance in 5 directions