

Subjective Effect of Synthesis Conditions in 3D Sound Field Reproduction System Using a Few Transducers and Wave Field Synthesis

Toshiyuki Kimura¹², Munenori Naoe¹²,

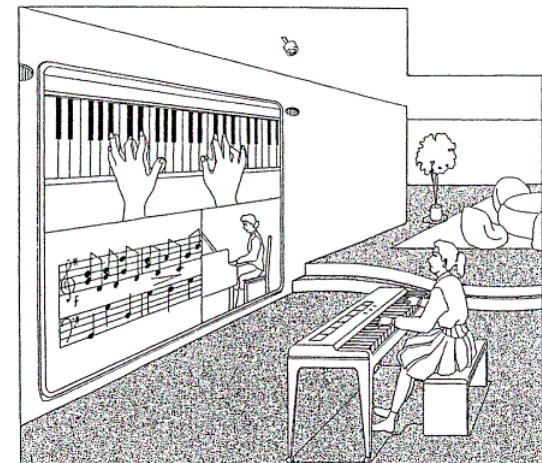
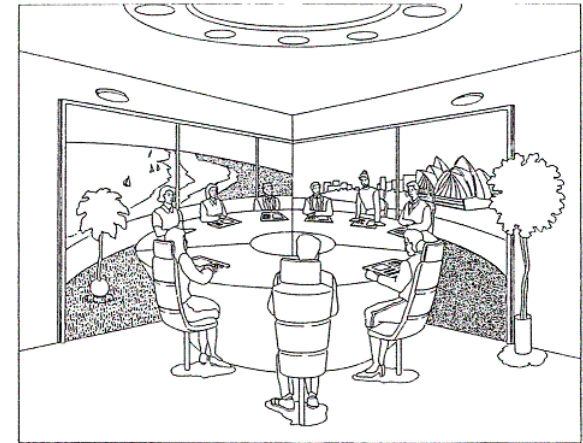
Yoko Yamakata¹ and Michiaki Katsumoto¹

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

²Graduate School of Engineering, Tokyo University of
Agriculture and Technology, Japan

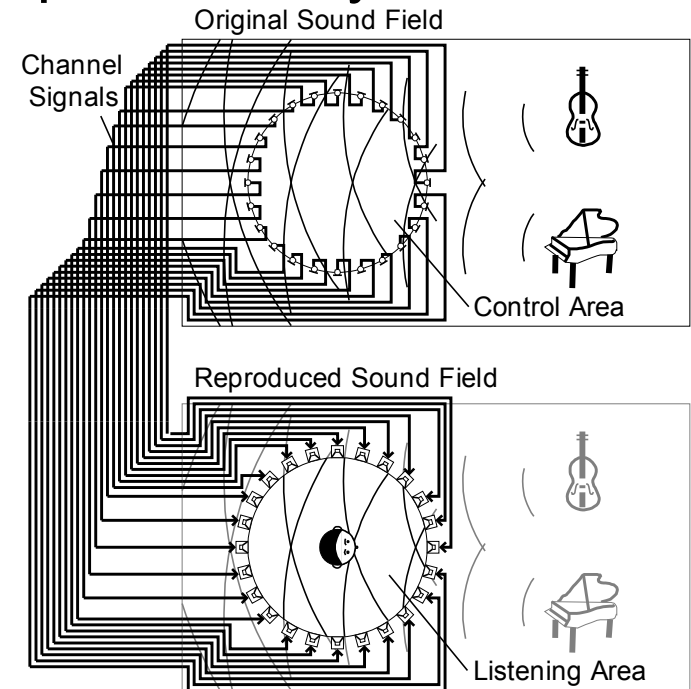
3D Sound Field Reproduction System

- 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 System

- Original sound field
 - Sound is captured by the microphone array
- Reproduced sound field
 - Captured sound is replayed by the loudspeaker array
 - Wave fronts are reproduced based on Kirchhoff-Helmholtz integral equation
- Feature
 - Multiple listeners can move or turn their heads while listening to a sound



Conventional Systems

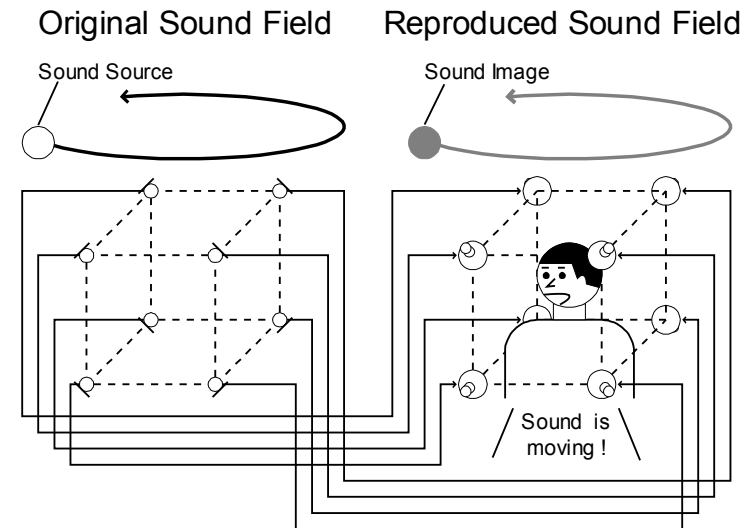
- Boundary Surface Control (BoSC) system
 - BoSC principle
 - Based on Kirchhoff-Helmholtz integral equation and uses inverse filters
 - 70 transducers are used
 - Loudspeakers are visible in the listener's field of vision
 - It is very difficult to construct an audio-visual system



- The number of transducers should be reduced in order to construct an A-V system

Diagram of Our Proposed System

- Original sound field
 - Sound is captured by 8 microphones
- Reproduced sound field
 - Captured sound is replayed by 8 loudspeakers
 - 3D sound field is reproduced
- Feature
 - Listener can turn their heads while listening to a sound
 - Loudspeakers are not visible in the listener's field of vision



Aim of Study

- We have evaluated the localized performance of the proposed system by the localization test
 - 12 directions' performance was good
 - Remaining 5 directions' performance was not good
 - Synthesis conditions were **fixed** in a past test
 - Size of cubic arrays
 - Directivity of microphones
- ↓
- Effect of synthesis conditions is evaluated by the localization test

Multichannel Signals

- Signals $x_i(n)$ are synthesized on a computer

- $i = 1 \dots 8$
$$x_i(n) = \frac{D_i}{d_i} s \left\{ n - \text{round} \left(\frac{d_i F_s}{c} \right) \right\}$$

- $s(n)$: Sound source signal

- $F_s (=48 \text{ kHz})$: Sampling frequency

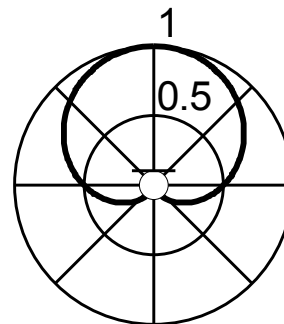
- $c (=340 \text{ m/s})$: Sound velocity

- d_i : Distance between the sound source and microphones

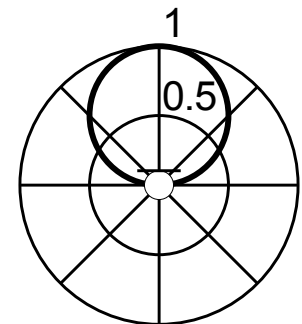
- D_i : Microphone directivity

- Unidirectional
- Shotgun

Unidirectional

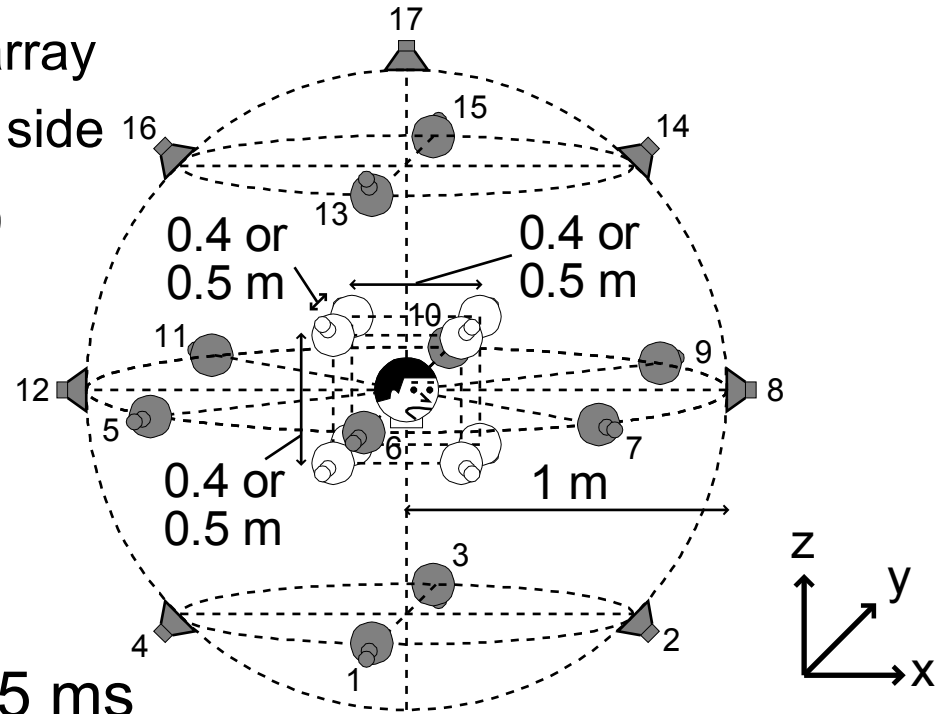


Shotgun



Experimental Environment

- 25 loudspeakers
 - 8 loudspeakers (white)
 - For cubic loudspeaker array
 - 0.4 m or 0.5 m on a side
 - 17 loudspeakers (gray)
 - For control condition
- Listening position
 - Center of a sphere
- Room conditions
 - Reverberation time: 115 ms
 - Background noise level: 20 dB(A)
 - Sound pressure level: 60 dB(A) (listening position)



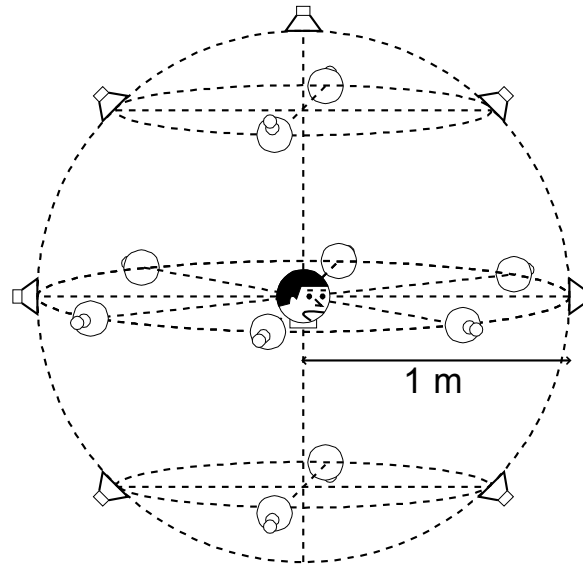
Synthesis Conditions

- Synthesis condition (iii)
 - Condition of the previous localization test
- The localized performance was compared among the four synthesis conditions

	Microphone directivity	Array size
(i)	Unidirectional	0.4 m
(ii)	Unidirectional	0.5 m
(iii)	Shotgun	0.4 m
(iv)	Shotgun	0.5 m

Experimental Condition

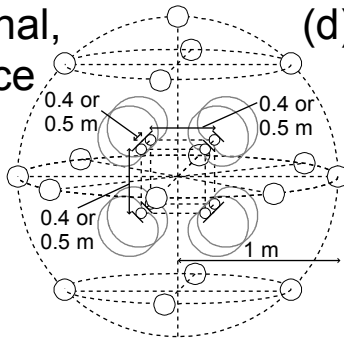
- Control condition (a)
 - Sound source signal is directly replayed from one loudspeaker selected from 17 loudspeakers
 - Listeners feel that there are one sound source in the position of white loudspeakers



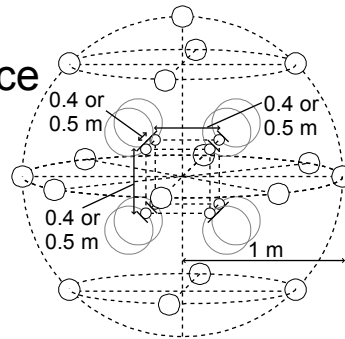
Experimental Condition(Cont')

- Four conditions (b)-(e)
 - 8ch signals are replayed from 8 loudspeakers
 - Listeners feel as if there are one sound image in the position of gray circles

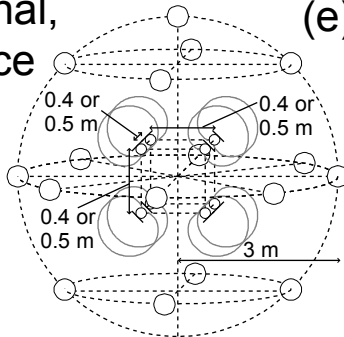
(b) Unidirectional,
1 m Distance



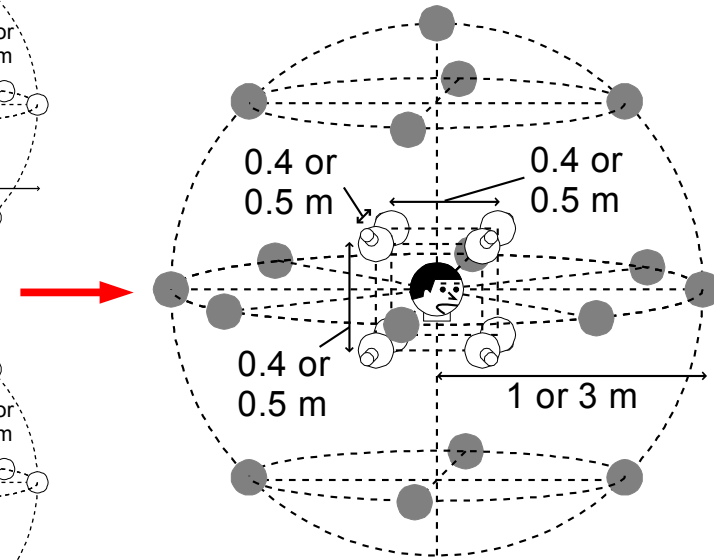
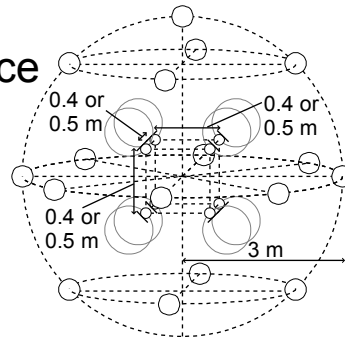
(d) Shotgun,
1 m Distance



(c) Unidirectional,
3 m Distance

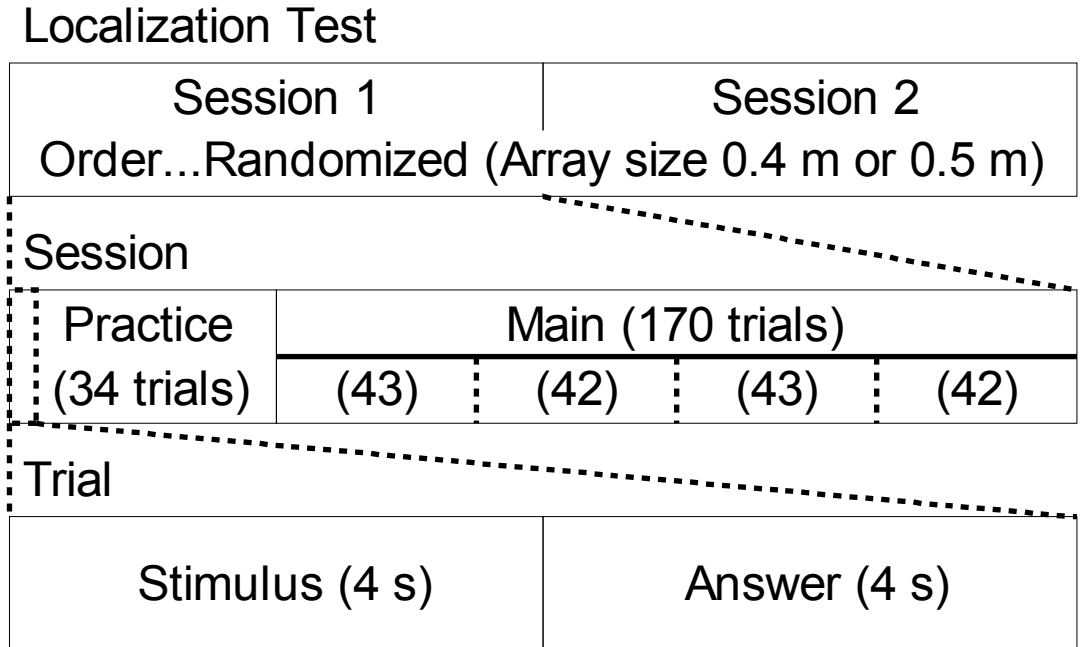


(e) Shotgun,
3 m Distance



Experimental Flowchart

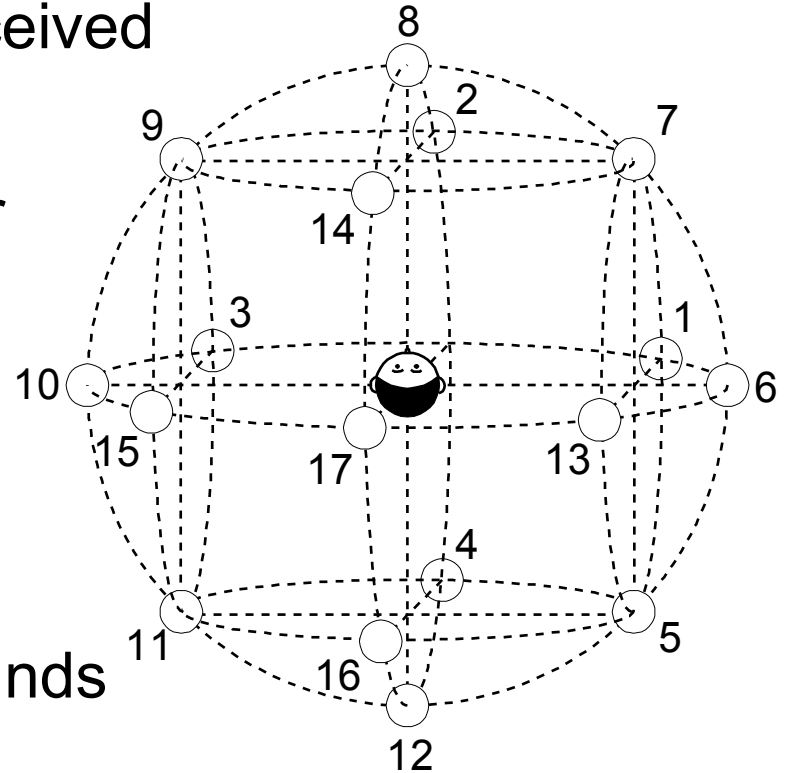
- Listeners
 - 7 males
 - 3 females
- Source
 - White noise



	Element	Note
Practice(34)	= 2 conditions x 17 directions	Experimental conditions (a) & (b)
Main(170)	= 5 conditions x 17 directions x 2 repetitions	Experimental conditions (a)-(e)

Experimental Procedure

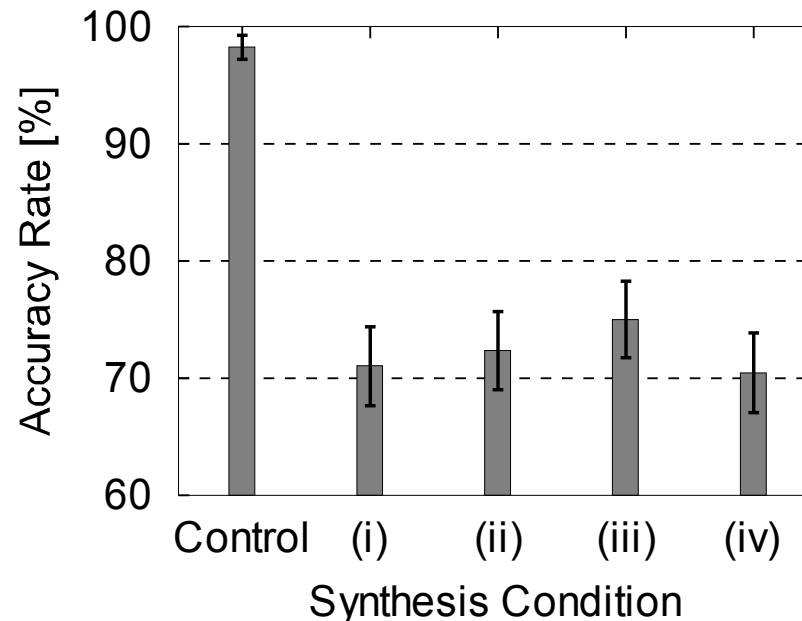
- Instruction
 - Listeners report the perceived direction of sound
 - Listeners list the number of the direction in an answer sheet
 - Listeners are allowed to turn their heads freely while listening to the sounds



Results in Each Synthesis Condition

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

- **Synthesis condition (iii) is best**
- Accuracy rate of all synthesis conditions
 - Lower than that of the control condition



χ^2 -test Result in Each Direction

- Significant difference of 1% level in (iii)
 - Yes: 6 directions (No. 4, 6, 8, 10, 12, and 17)
 - No: other 11 directions
- **Performance of (iii) is best in 11 directions**

	Control	(i)	(ii)	(iii)	(iv)
1	100%	70%	65%	95%	88%
2	98%	73%	88%	85%	88%
3	100%	75%	55%	88%	65%
4	98%	75%	63%	70%	70%
5	100%	98%	100%	100%	98%
6	100%	50%	33%	40%	30%
7	100%	100%	100%	100%	100%
8	100%	40%	43%	33%	28%

	Control	(i)	(ii)	(iii)	(iv)
9	100%	100%	98%	98%	100%
10	100%	35%	28%	38%	18%
11	100%	95%	98%	95%	100%
12	100%	45%	35%	53%	40%
13	100%	80%	88%	90%	83%
14	98%	93%	90%	95%	95%
15	95%	78%	88%	85%	90%
16	88%	50%	55%	65%	78%
17	95%	53%	65%	48%	73%

Answer Rates of Six Directions

- Six directions
 - Behind and downward direction (No. 4)
 - Left direction (No. 6)
 - Front direction (No. 8)
 - Right direction (No. 10)
 - Behind direction (No. 12)
 - Top direction (No. 17)

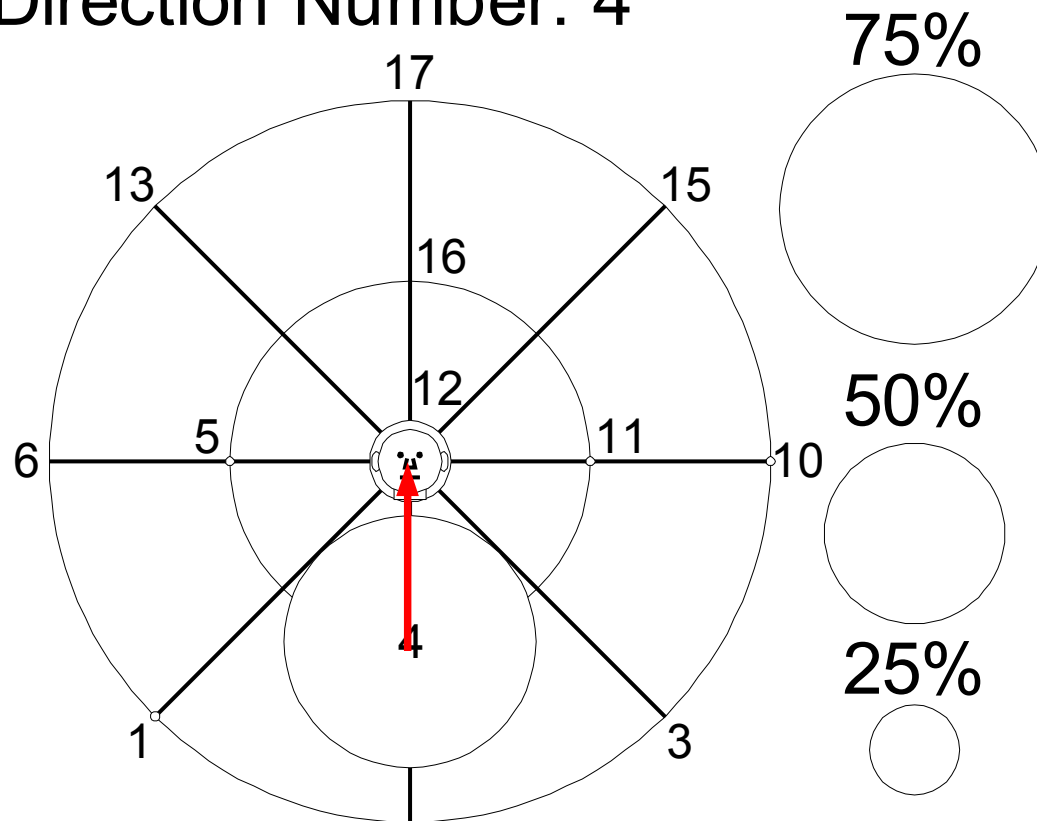
- Definition of answer rates

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

Answer Rates in Downward Direction

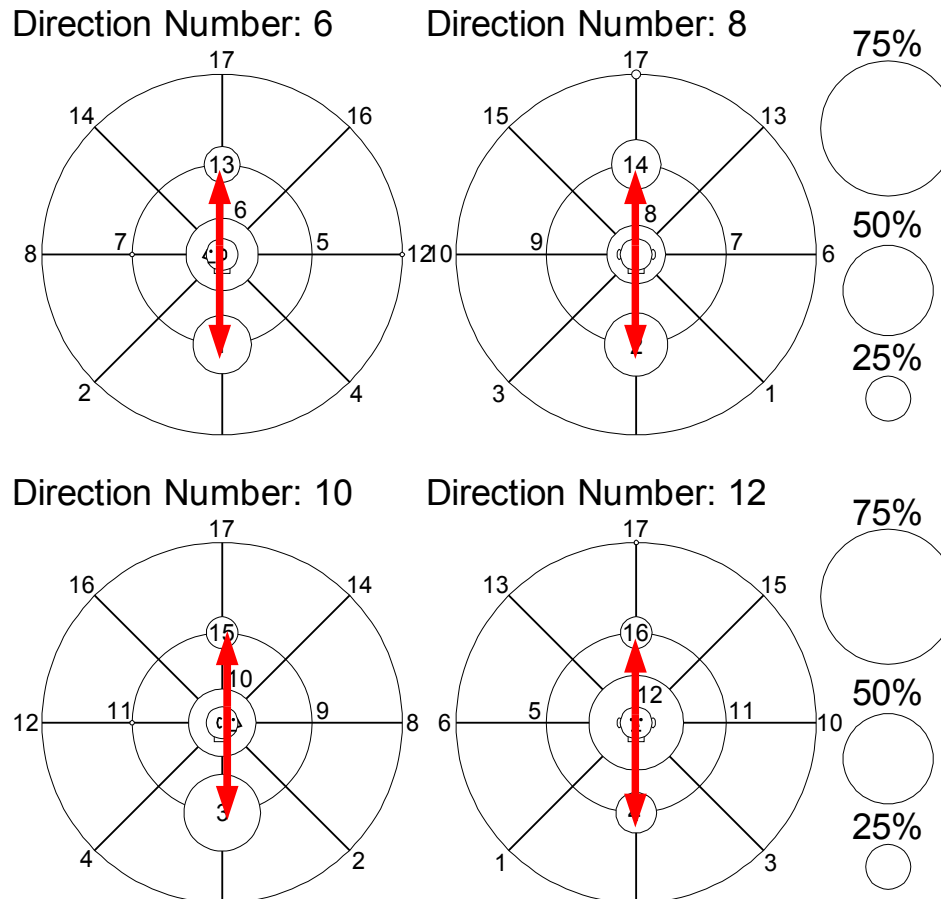
- Sound images are biased toward the upper direction

Direction Number: 4



Answer Rates in Horizontal Direction

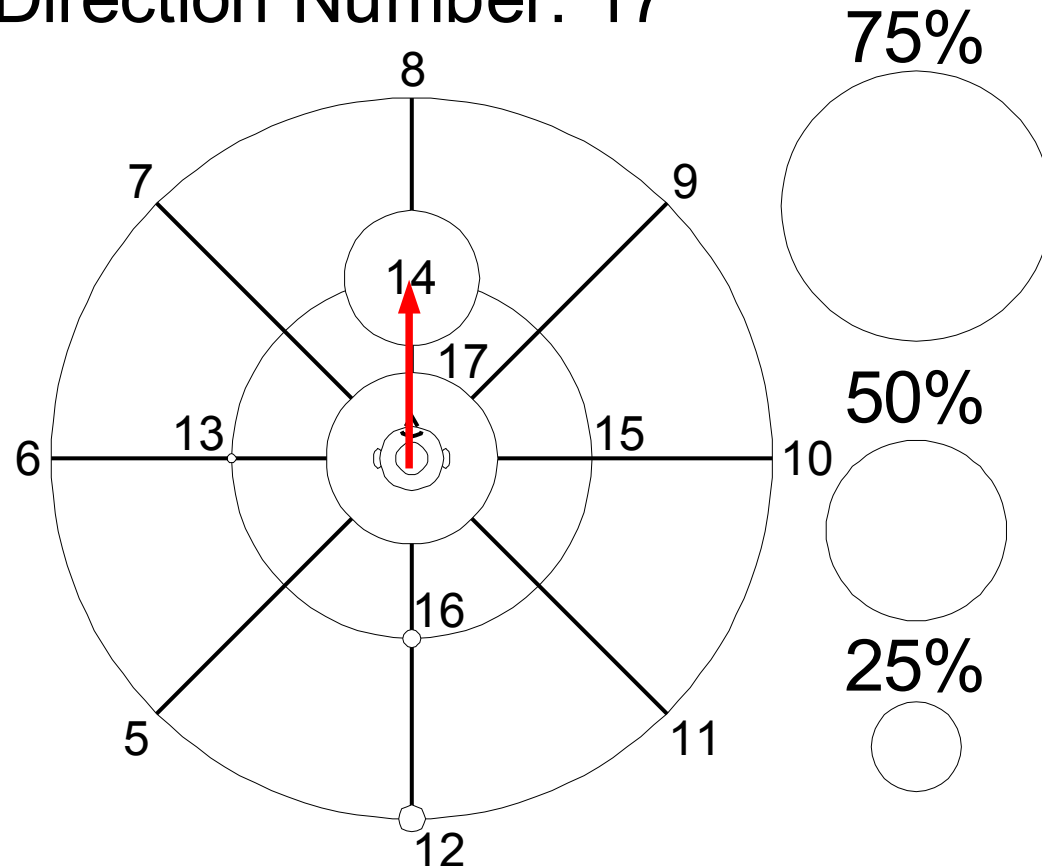
- Sound images are blurred toward the vertical direction



Answer Rates in Top Direction

- Sound images are biased toward the forward direction

Direction Number: 17



Conclusions

- In 3D sound field reproduction system using 8 transducers and wave field synthesis, the effect of synthesis conditions was evaluated by the localization test
- Cubic array size: 0.4 m, shotgun microphones
 - The localized performance was best in 11 directions of 17 directions used in the test
 - The bias and blur of sound images occurred in remaining 6 directions
- Future works
 - Improvement of performance in remaining 6 directions