National Institute of Information and Communications Technology

Listening Test for Three-dimensional Audio System Based on Multiple Vertical Panning

Toshiyuki Kimura and Hiroshi Ando

Universal Communication Research Institute,

National Institute of Information and Communications Technology (NICT), Japan

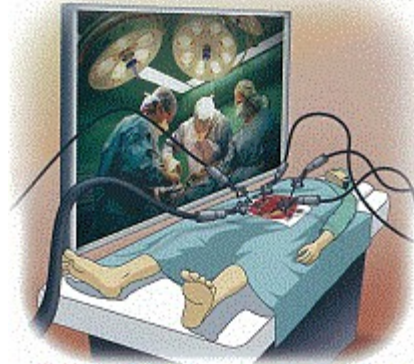# Ultra-Realistic Communications Technique

- Realistic 3D video and audio appear in a 3D space by these techniques

- More realistic form of communication
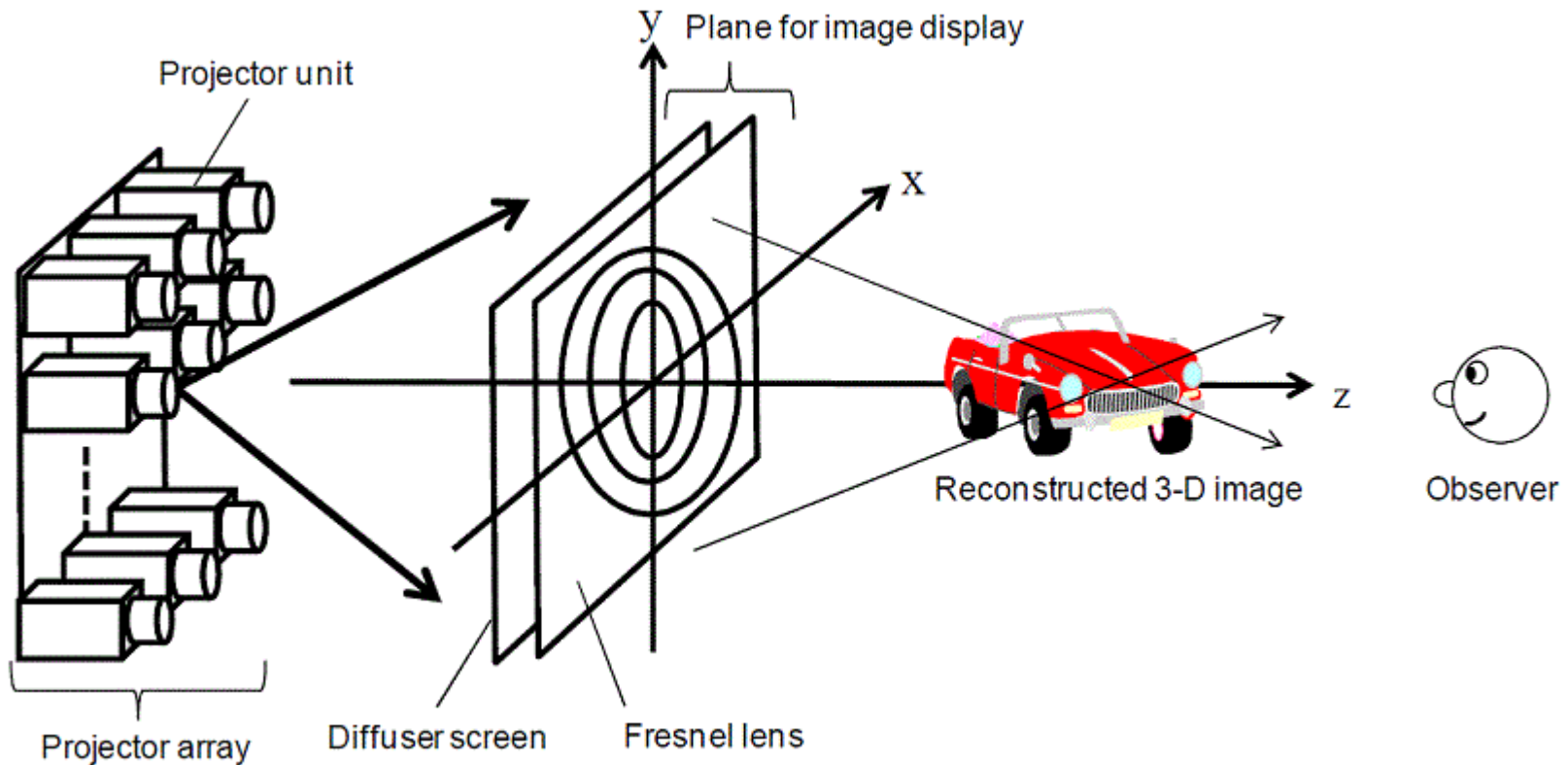
3D television   3D telesurgery   3D teleconference

3D teleshopping   3D remote education

Large Grasses-free 3D Video Display System

- ## This system provides parallax videos according to horizontal positions
  - ### Several people can observe parallax videos according to horizontal viewing positions


Projector unit

y Plane for image display

x

z

Reconstructed 3-D image

Observer

Projector array

Diffuser screen

Fresnel lens

Aim of Study

- 3D audio system matched to the developed 3D video display system

$$\downarrow$$

- Novel 3D audio system is proposed
  - Based on Multiple Vertical Panning (MVP) method

Basic Configuration of Proposed System

- 2 loudspeakers are placed at the top and bottom of the position of the 3D object

- Sound is played by the "vertical panning"
  - Listeners feel that a sound image is played between 2 loudspeakers

- Appropriate sound level differences
  - Multiple listeners can feel that a sound image is played at the position of the 3D object

Sound Image Position
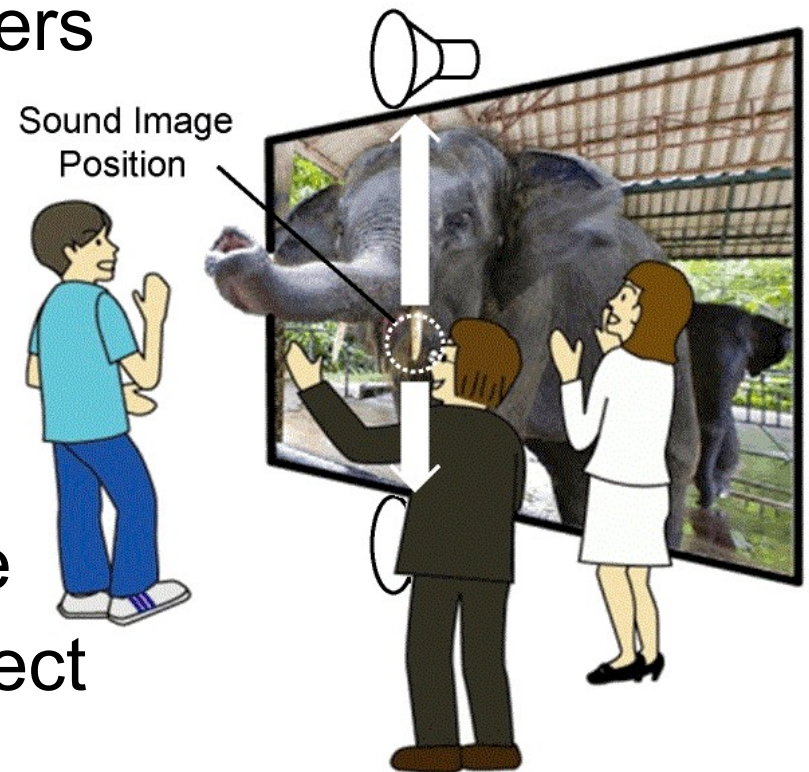
Basic Configuration of Proposed System

- Multiple loudspeaker pairs are placed at the top and bottom of the screen
  - Sound image positions are also expanded to the left-right direction
- Multiple listeners can simultaneously feel multiple sound images at the position of 3D objects
  - regardless of listening position

Sound Image Position 1

Sound Image Position 2
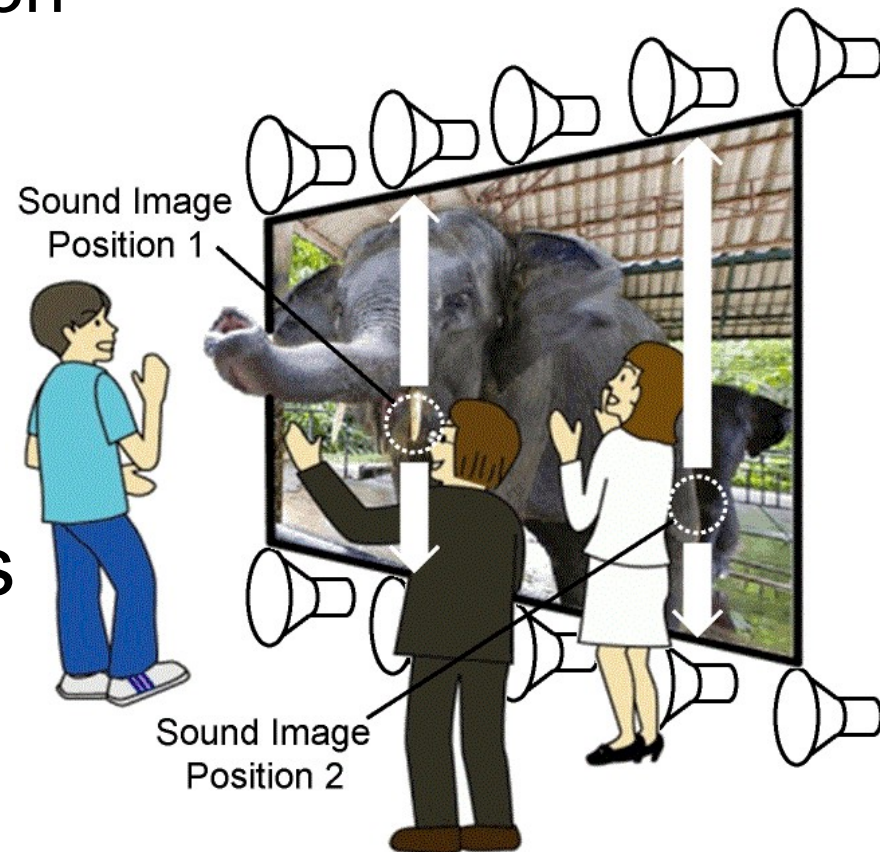
Experimental Environment

- ## ATR variable reverberation room
  - Reverberation time: 140 ms, 1030 ms
  - Background noise level
    - 14 dBA (140 ms), 22 dBA (1030 ms)

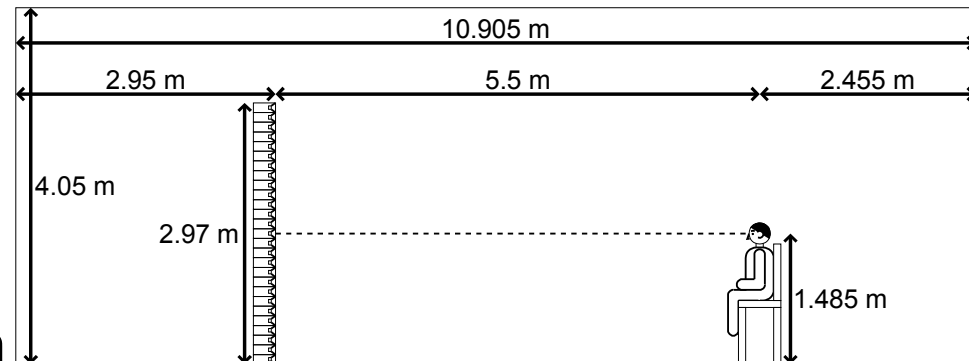140ms   1030ms

Experimental Environment

- 27 loudspeakers are placed in the vertical line
  - Height of array
    - 2.97 m (=11 cm×27)
  - Listening position
    - 5.5 m distance from the array
  - Sound level
    - About 70 dBA at listening position

Plane View

10.905 m

2.86 m

2.95 m   5.5 m   2.455 m

6.67 m

3.81 m
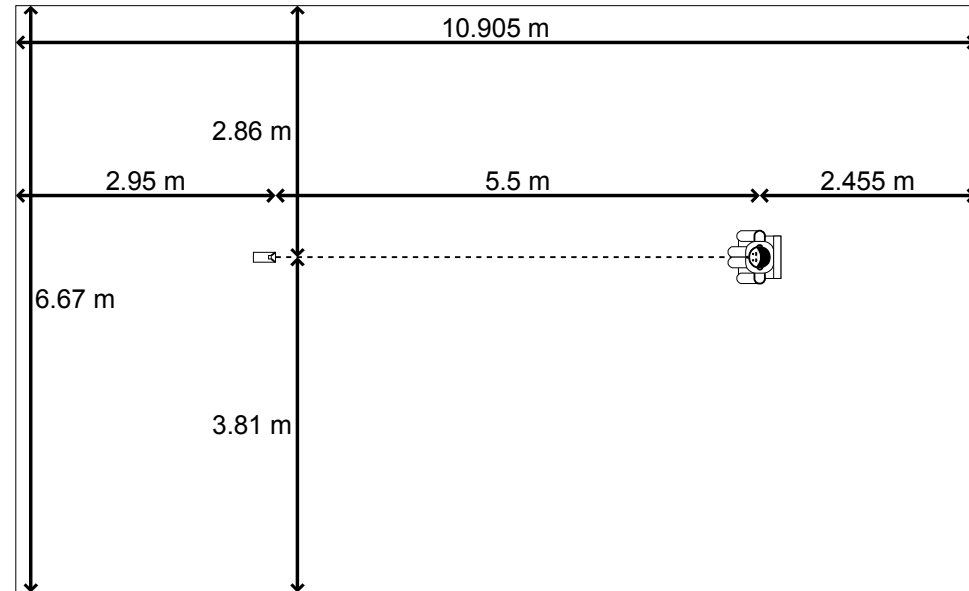
Cross-sectional View

10.905 m

2.95 m   5.5 m   2.455 m

4.05 m

2.97 m

1.485 m

Experimental Condition

- ## Panning condition
  - Level different sounds are played from 2 loudspeakers
  - Number: 31 (-15dB~15dB, 1dB interval)

  (a) Panning Condition

- ## Control condition
  - Sound is played from 1 loudspeaker selected from 27 loudspeakers
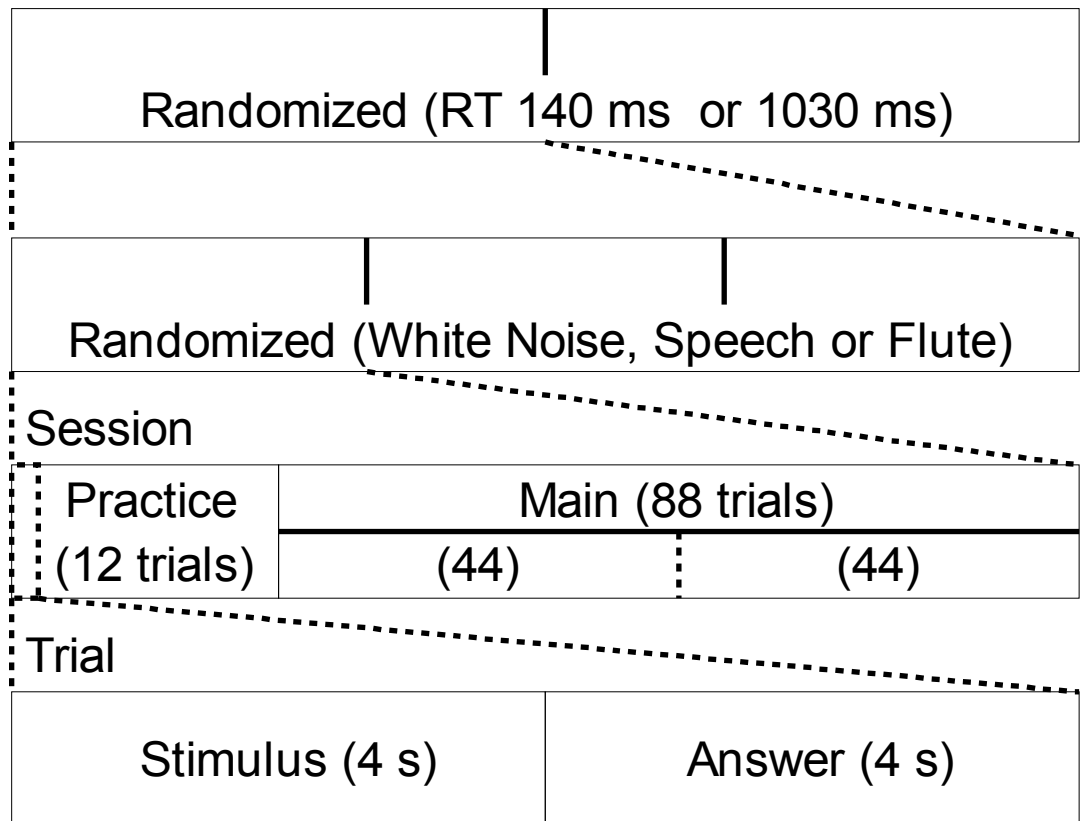 - Number: 13

(b) Control Condition

Experimental Design

- ## Sound source
  - White noise, speech and flute
    - Duration: 4 seconds
- ## Subject
  - 12 persons
    - 6 males
    - 6 females
  - Age
    - 21-32
  - Audibility
    - Normal in daily life

Listening Test

Randomized (RT 140 ms or 1030 ms)

Randomized (White Noise, Speech or Flute)

Session

Practice (12 trials)	Main (88 trials)	
	(44)	(44)

Trial

Stimulus (4 s)	Answer (4 s)

Details of Trials

- Number of sessions
  - 6 =sound source (3)×reverberation time (2)
  - Presentation order: randomized

- Practice trials
  - 12=(7(panning)+5(control))×1(repetition)
  - Presentation order: randomized

- Main trials
  - 88=(31(panning)+13(control))×2(repetition)
    - Rest period: every set of 44 trials
  - Presentation order: randomized

Experimental Procedure

- Report the perceived height of sound images by listing the 27 index of heights in an answer sheet

  - Index: presented in the right side of loudspeakers
  - If subjects felt multiple sound images, they could list multiple indexes
  - Allowed to move their heads and upper bodies freely while listening to the sounds
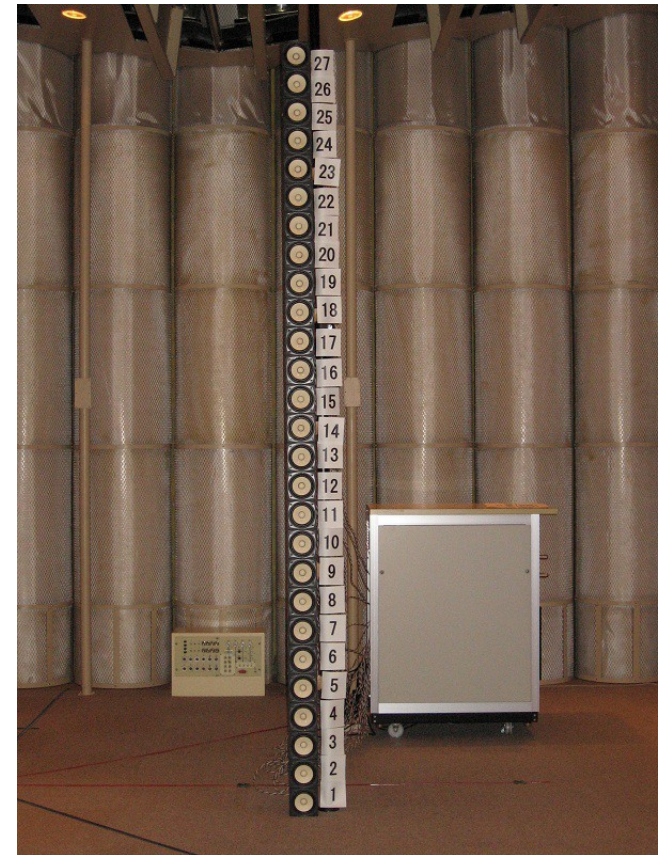
Analysis of Experimental Result

- Reduce the answers of multiple indexes
- Calculate the perceived height of sound images

$$H_{per}[\mathrm{m}] = (I_{ans} - 14) \times 0.11$$
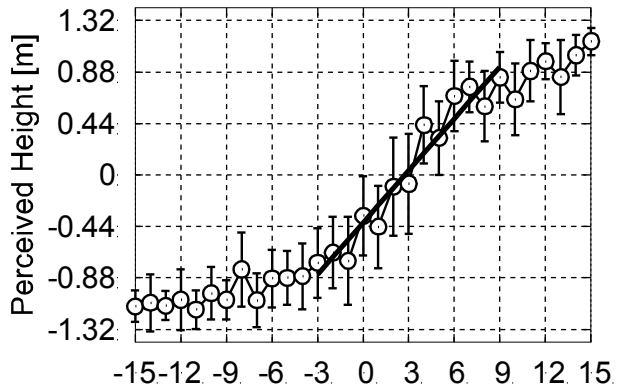
- I_{ans} : Answering index of loudspeakers
- H_{per} : Perceived height
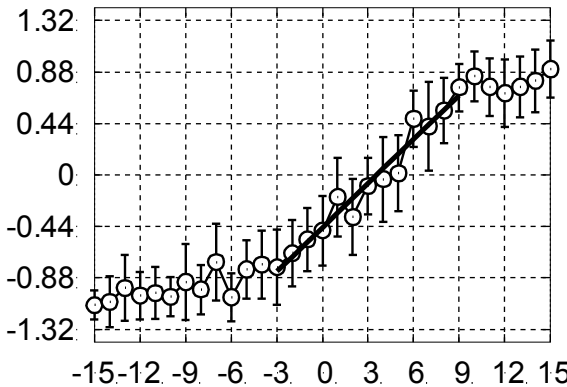 - 1 : -1.43 m
  - 14 : 0 m
  - 27 : 1.43 m

Experimental Results (Panning Condition)

- ## Level difference: -3~9dB
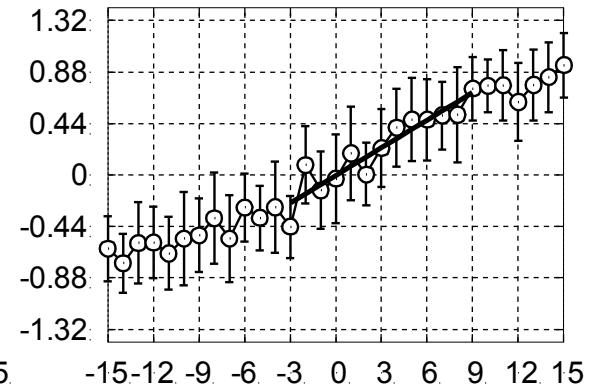 - ### The perceived height linearly changes
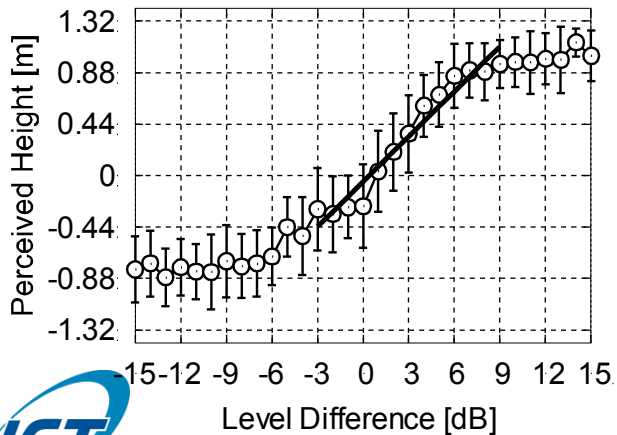

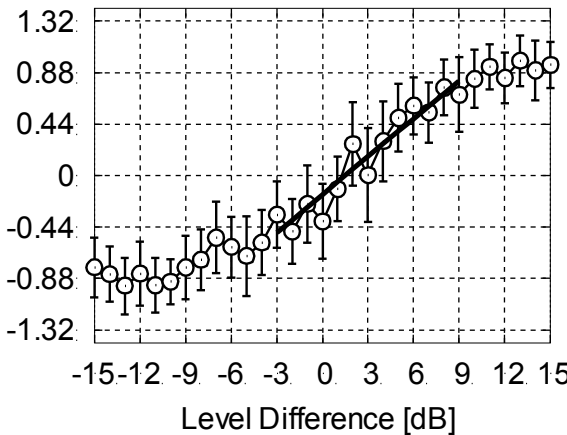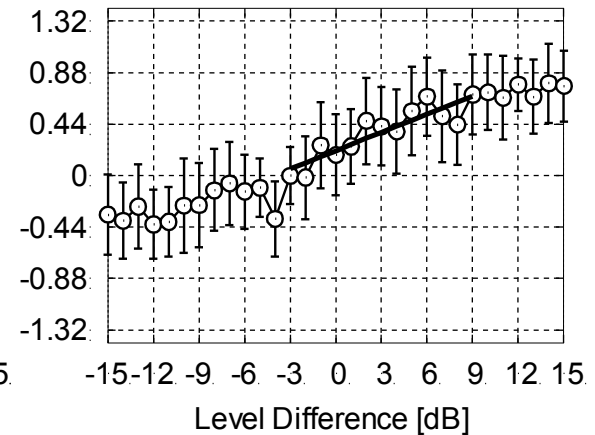White Noise (140ms) · Speech (140ms) · Flute (140ms)
White Noise (1030ms) · Speech (1030ms) · Flute (1030ms)

Calculation of Panning Curve

- Calculate according to the average of regression lines

$$H_{\text{pan}} = \begin{cases} -1.32 & (\Delta A < -11.05) \\ 0.1065\,\Delta A - 0.1437 & (-11.05 \leq \Delta A \leq 13.74) \\ 1.32 & (\Delta A > 13.74) \end{cases}$$

  - ΔA[dB]: level difference

- Calculate the differential limens of the perceived height

$$DL_{\text{pan}}^{+} = \tan\left(\tan^{-1}\left(H_{\text{pan}}/5.5\right) + \varphi\right) \times 5.5$$

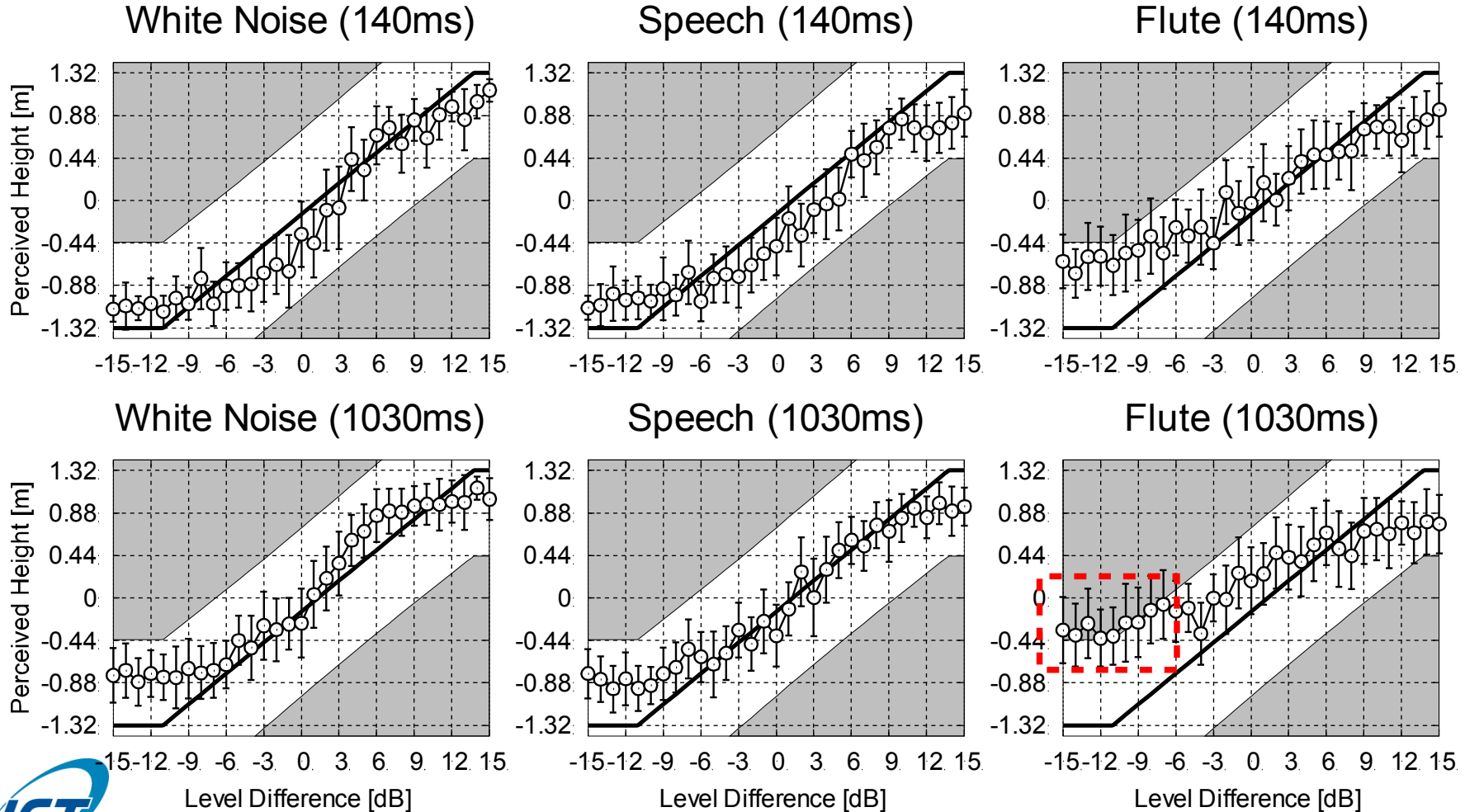$$DL_{\text{pan}}^{-} = \tan\left(\tan^{-1}\left(H_{\text{pan}}/5.5\right) - \varphi\right) \times 5.5$$

  - ϕ (=9 degrees): differential angle

Experimental Results (Panning Curve)

- There is no average of the perceived height of sound images in the gray area

White Noise (140ms) Speech (140ms) Flute (140ms)

White Noise (1030ms) Speech (1030ms) Flute (1030ms)

Perceived Height [m]

Level Difference [dB]

Experimental Results (Control Condition)

- ## Flute, reverberation 1030 ms
  - ## – There is the average in the gray area

White Noise (140ms) · Speech (140ms) · Flute (140ms) · White Noise (1030ms) · Speech (1030ms) · Flute (1030ms)

Perceived Height [m] · Presented Height [m]

Discussions
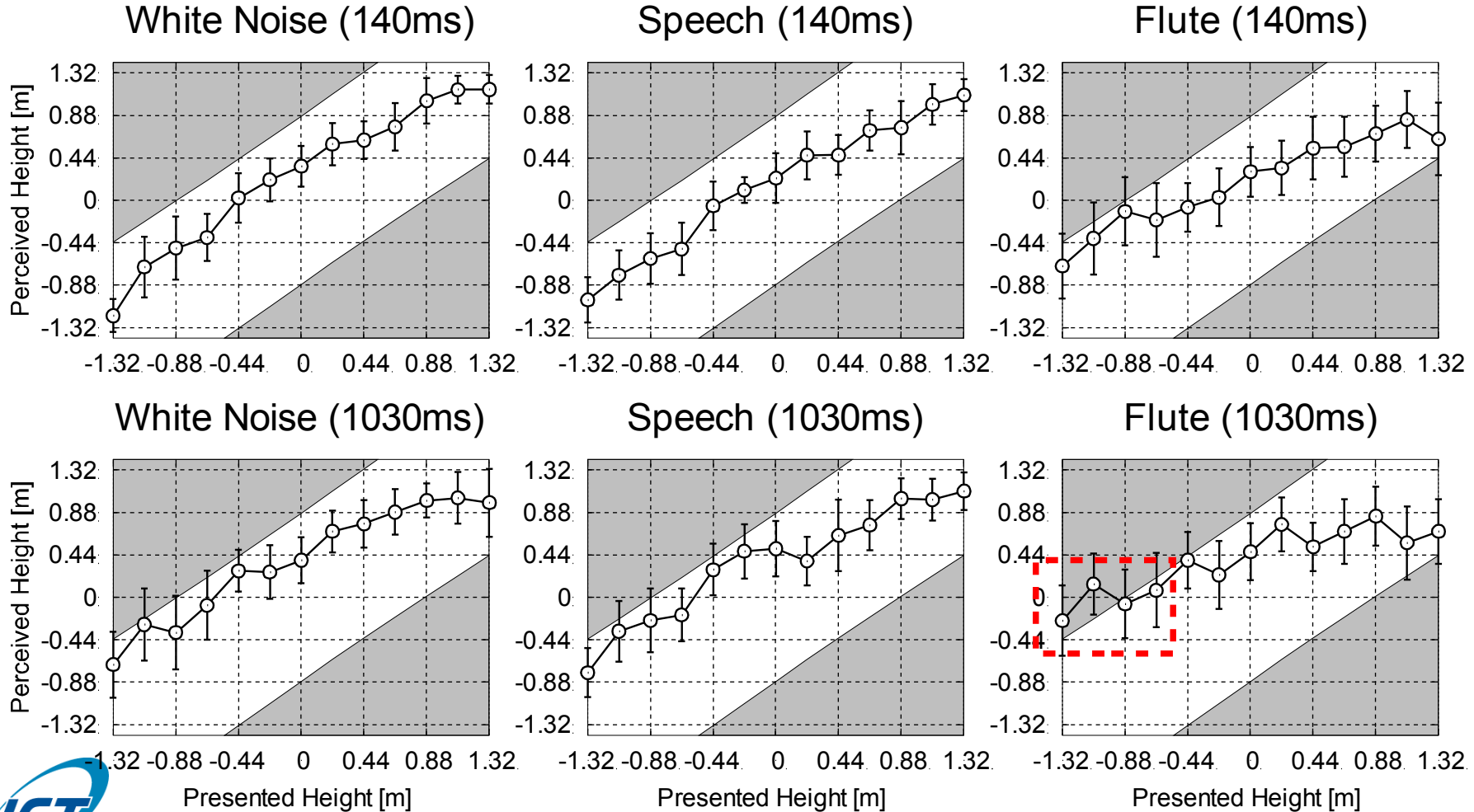
- ## Panning condition
  - ### 5 condition
    - There is no average of the perceived height of sound images in the gray areas
    - The auditory performance of the panning curve is so high that subjects cannot discriminate the difference between the heights
  - ### Flute, reverberation time 1030 ms
    - There is the average of the perceived height of sound images in the gray areas
    - Subjects may be able to discriminate the difference between the heights

Discussions

- Control condition
  - Flute, reverberation time 1030 ms
    - There is the average of the perceived height of sound images in the gray areas

- Subjects may not perceive the height of sound sources itself due to the reverberation time when the sound sources are flute

- It needs to evaluate the effect of the reverberation time by the additional listening test

Conclusion

- ## Novel 3D audio system is proposed
  - Based on Multiple Vertical Panning (MVP)
  - Match to the large glasses-free 3D video display system

- ## Evaluation of the auditory performance of the proposed system by the listening test
  - 27 loudspeakers were aligned on the vertical line
  - 5 conditions except the condition (Flute, Reverberation time 1030 ms)
    - Subjects cannot discriminate the difference between the perceived heights of sound images

Future Works

- Evaluation of the effect of the reverberation time by the additional test
  - flute and the reverberation time varies
- Evaluation of the effectiveness of the proposed system in an audio-visual system
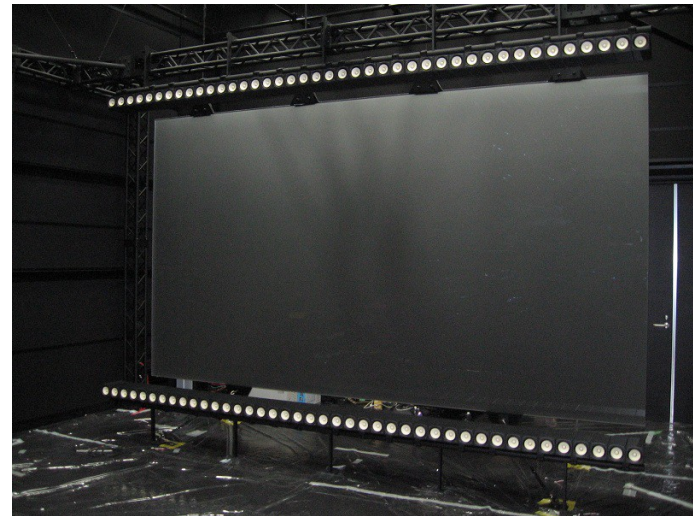
Technical Requirement of 3D Audio System

- Observers can feel sound images at the position of 3D objects in the free position
  - (NOT) stereophonic, 5.1ch system
- Observers don't wear a sound device
  - (NOT) binaural
- There are no devices between the projector array and the screen
  - (NOT) 22.2ch system, higher order ambisonics, wave field synthesis
- There are no microphones between the screen and the viewing position
  - (NOT) transaural, boundary surface control

Motivation of Listening Test

- Proposed system
  - The minimum component is 2 loudspeakers placed at the top and bottom of the screen
    - Auditory performance can be represented by the superposition of minimum components
    - It is enough to evaluate the perceived height of synthesized sound images of 2 loudspeakers
  - Listening test to evaluate the perceived height of synthetic sound images of 2 loudspeakers
    - It is assumed that loudspeakers are placed at the top and bottom of the screen of the large glasses-free 3D video display system
    - The vertical panning curve is also evaluated

Rates of Answers of One Sound Image

- ## Panning condition
  - Range: level difference -3~9dB
  - There are no significant difference between control conditions
    - Except the condition (speech, RT 140 ms)
  - Rates are always more than 93.9%

RT 140ms

RT 1030ms