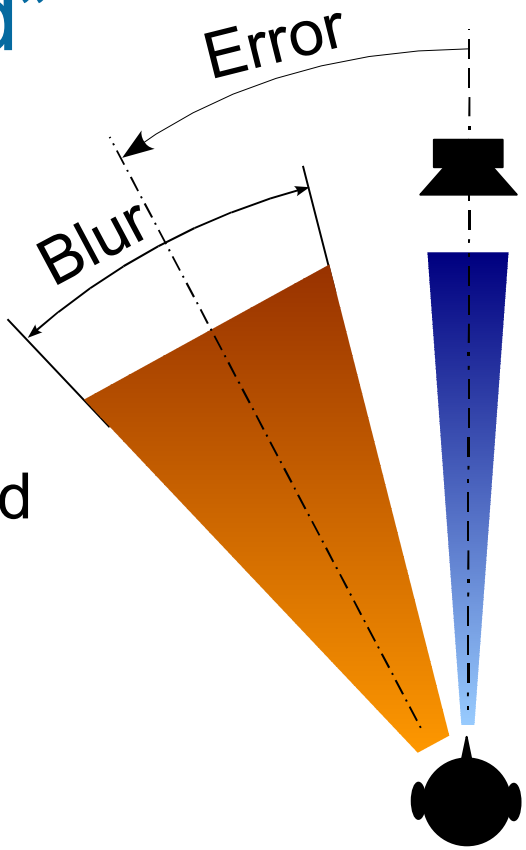
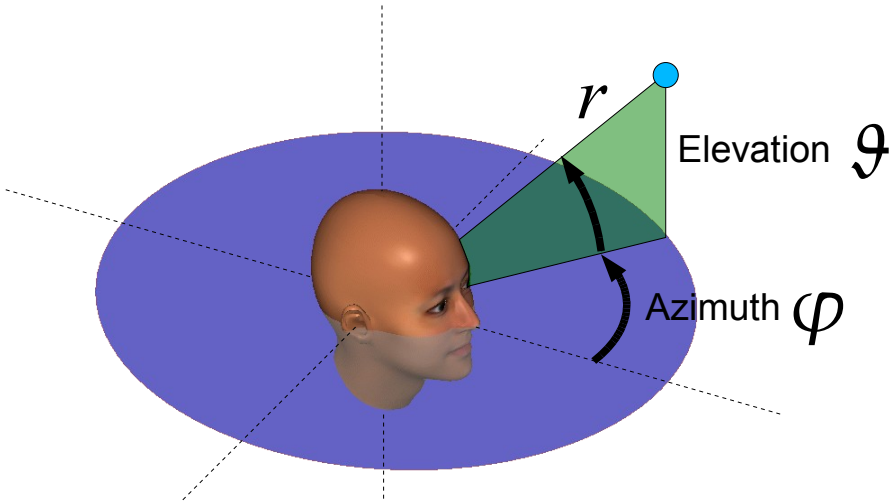
The Accuracy of Localizing Virtual Sound Sources: Effects of Pointing Method and Visual Environment

Piotr Majdak, Bernhard Laback, Matt Goupell, Michael Mihocic
http://www.kfs.oeaw.ac.at
piotr@majdak.com

Requirements for a “Good” Localization Method

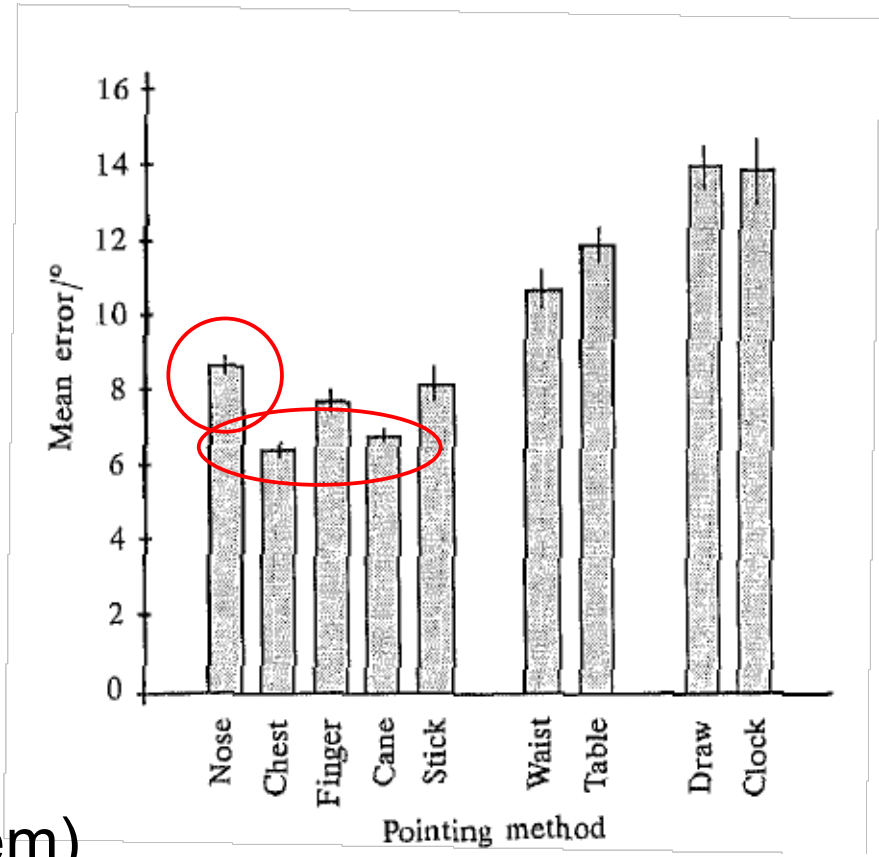
- All positions allowed
- High accuracy in all directions
- No interaction with a response method
- Comfortable for the subjects
- Short familiarization time
- Extendable to a learning procedure





Response Methods

- Haber et al. (1993):
  - Pointing with:
    - Nose
    - Shoulder
    - Finger
    - Cane or short stick
  - Rotating dial:
    - Mounted on a table
    - Attached to the subject
  - Drawing a line
  - Verbal (clockface system)
  - **Blind subjects!**



Effects of the Response Methods

- Head Pointing:
  - Very spontaneous response
  - Visual reference frame close to the acoustic reference frame
  - Comparison to the literature (Middlebrooks, 1999)
- Manual (hand) Pointing:
  - Very intuitive for responses in the range of 360°
  - Better hand is better than head (Pinek & Brouchon, 1992)
  - No underestimation of top positions (Djelani et al., 2000)
  - Trained daily in social interactions

Effects of Visual Feedback

- Indications for importance of visual feedback:
  - Berkinblit et al (1995): Finger pointing with vision better
  - Redon and Hay (2005): better with a structured background
  - Lewald et al (2000): head pointer blind: overestimation of positions; with vision: better
  - Zahorik (2006): fast decrease of acoustic front/back confusions using visual feedback
- **No** study about visual feedback and virtual acoustic stimuli with individualized HRTFs

Hypotheses and Conditions

- Hypotheses:
  - Higher accuracy using visual feedback
  - Higher errors for extreme positions using head pointer
  - Longer familiarization period using manual pointer
- Variable „Visual Environment“ (VE):
  - Dark: in darkness, without any visual cue
  - HMD: with VE presented via head mounted display
- Variable „Pointer“:
  - Head pointer: pointing with the head/nose
  - Manual pointer: pointing with a gun

Virtual Visual Environment (VE)

- Presented via head mounted display (HMD):
  - Stereoscopic view, in color, **without depth**
  - Field of View: 32° x 24° (hor. x vert.)
- Subject's position and orientation:
  - Tracked in real-time
  - Azimuth and elevation for the head (no movements)
  - Five degrees of freedom for the hand pointer (no roll)
- Tests in a dark sound chamber (A-weighted SPL of the background noise: 18 dB)

Platform

- Diameter: 0.8 m
- Supports comfortable 360°-turns
- Head and pointer position and orientation tracked in real-time

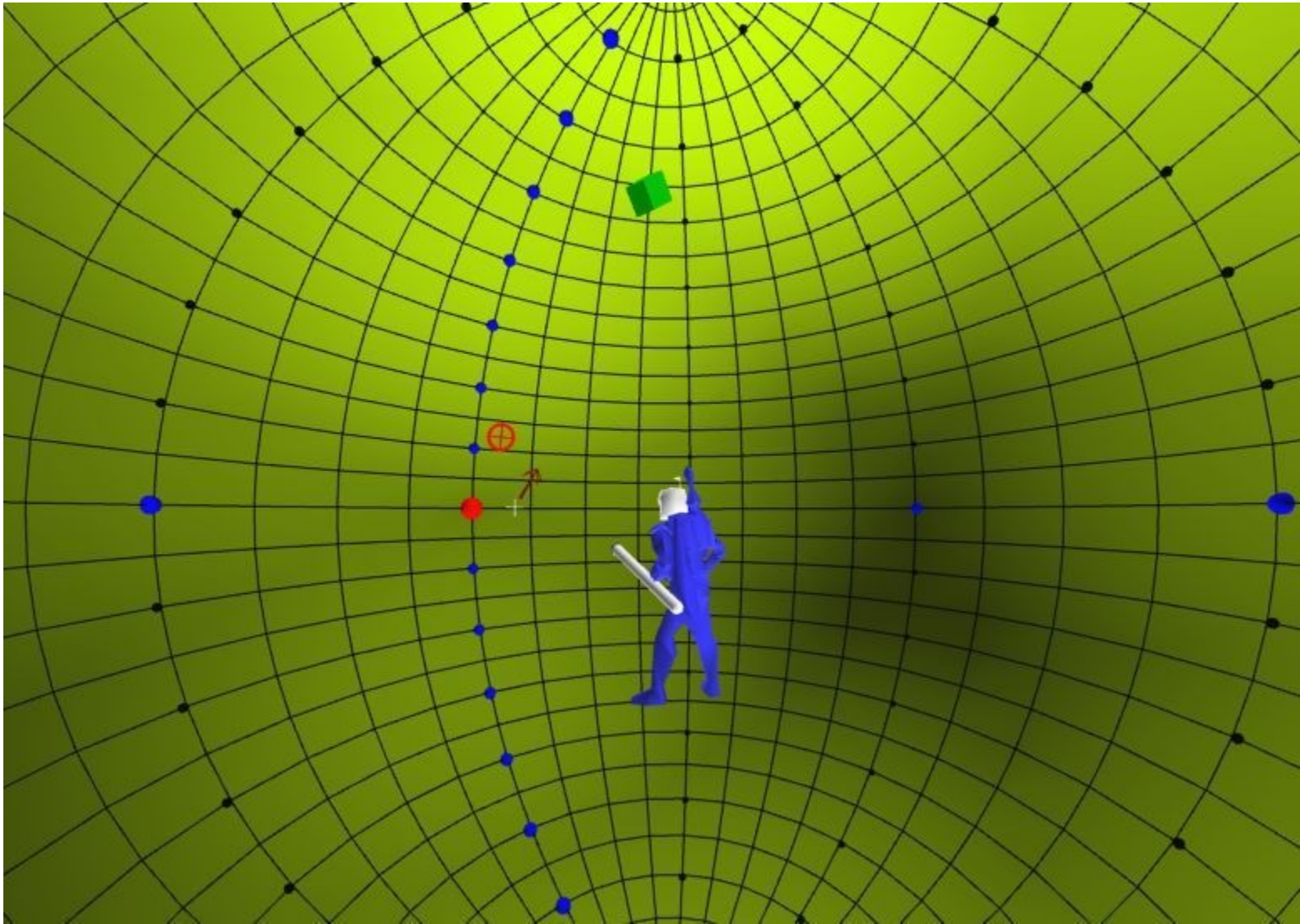
OAW
Austrian Academy
of Sciences

Platform

ARI

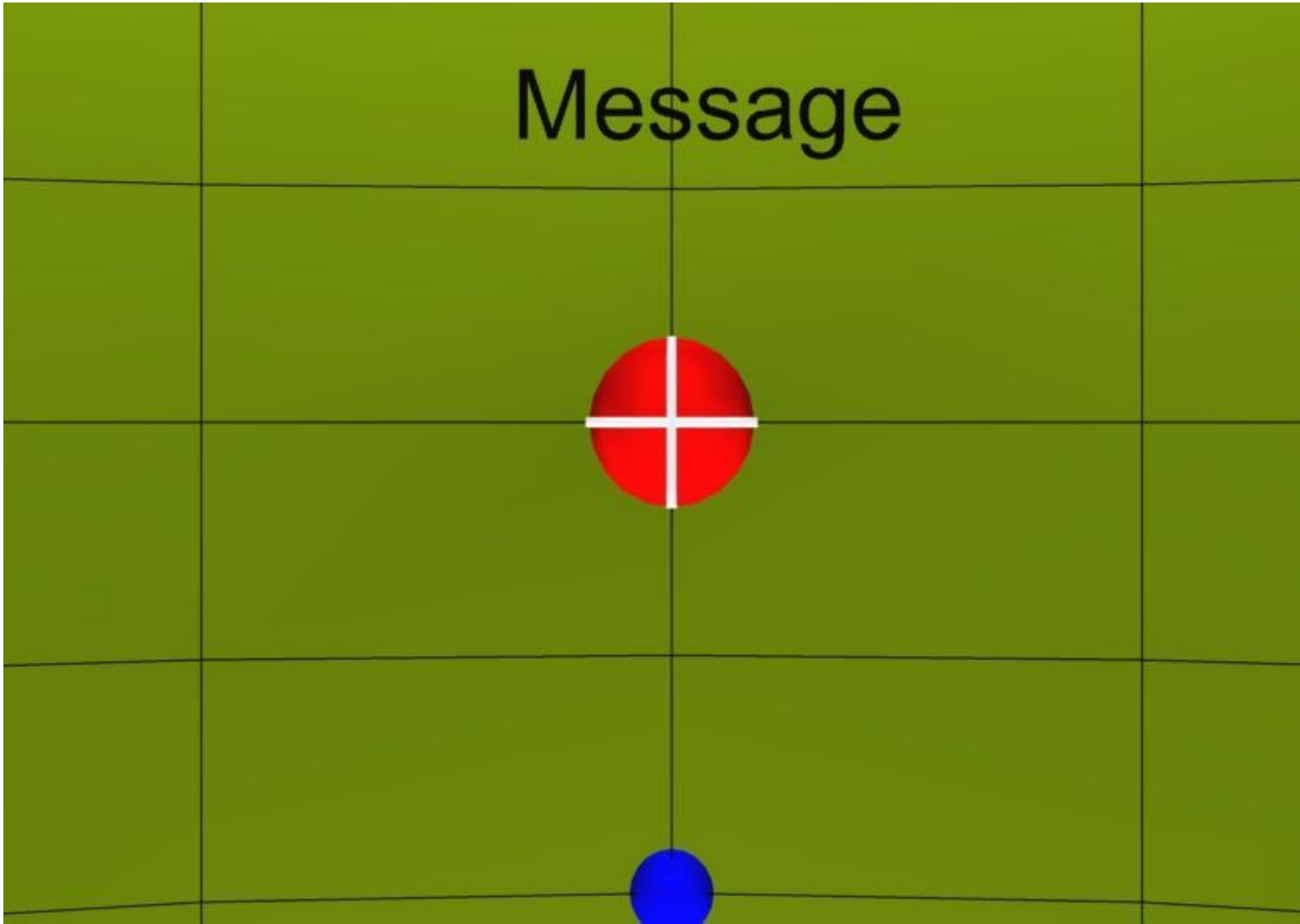
OAW
Austrian Academy of Sciences

Virtual VE: Outside-of-the-box View



ARI

Virtual VE: First-person View



OAW
Austrian Academy of Sciences

Visual Cues

ARI

HRTF Measurements

- System identification method:
  - Multiple Exponential Sweep Method (Majdak et al. 2008)
  - Measured at the entrance of the blocked-ear canal
- Positions:
  - Horizontal plane: -180° to 180°, in 2.5° steps
  - Vertical plane: -30° to +80° in 5° steps
  - 1550 positions in total
  - Subject's position controlled (± 2.5 cm; ± 2.5°)
- Directional Transfer Functions (DTF)

Targets and Subjects

- Targets:
  - 400 random positions per condition (out of possible 1550)
  - Subjects cannot build-up a mental map of the spatial positions (Butler et al. 1990; Hammershoi and Sadvad, 1994; Perret and Noble, 1995)
  - Statistical analysis easy by having a well-defined distribution (ANOVA)
- Subjects:
  - 10 **naïve**, right-handed subjects
  - Normal or corrected-to-normal vision
  - No auditory deficits

Experiments

- Visual Training:
  - Familiarization with equipment and procedure
  - Improving the degree of immersion in the VE
  - Reducing differences in experience across the subject
- Visual Test
  - Test the response accuracy with visual targets
- Acoustic Test:
  - Test the localization ability to virtual acoustic stimuli
  - No feedback provided

OAW
Austrian Academy of Sciences

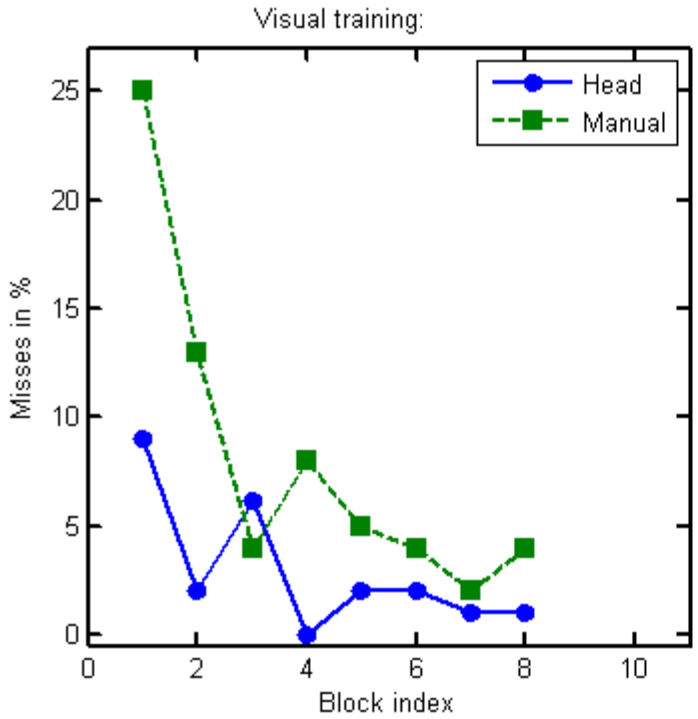
Visual Training: Method

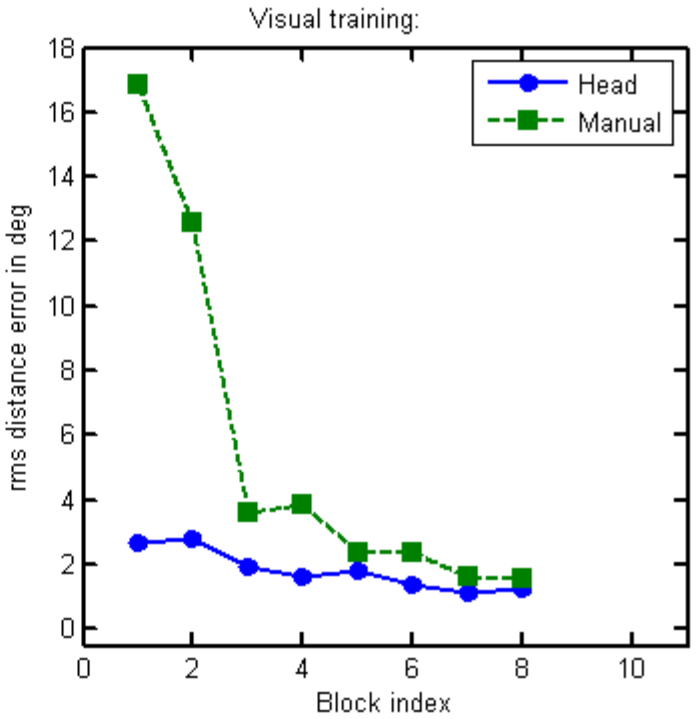
ARI

Visual Training: Method

Visual Training: Results

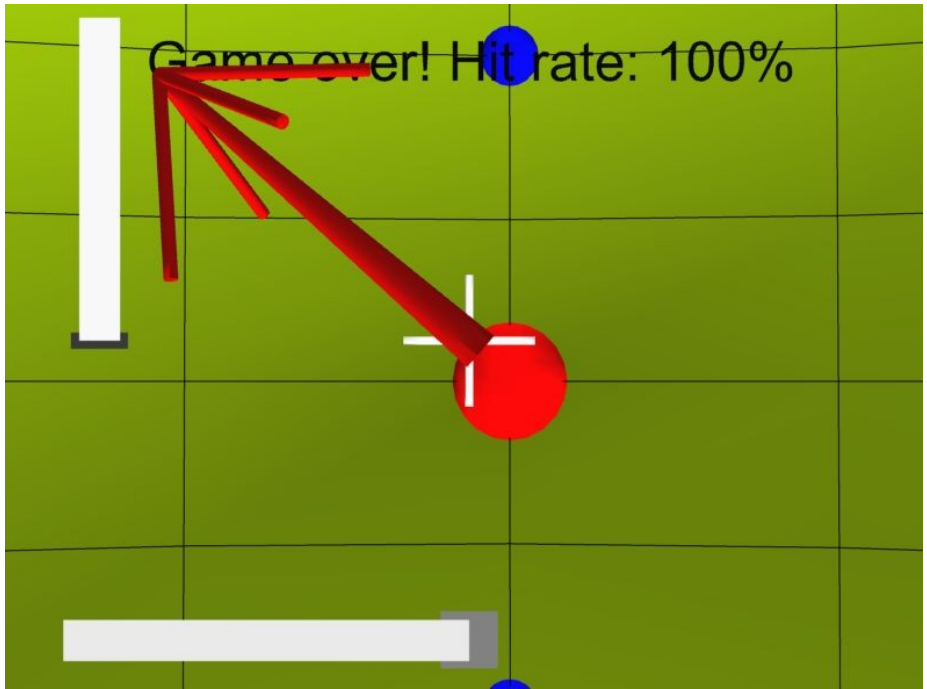
- Training until:
  - Hit rate of > 95% in one block of 100 targets
  - Distance error of < 2° in any direction





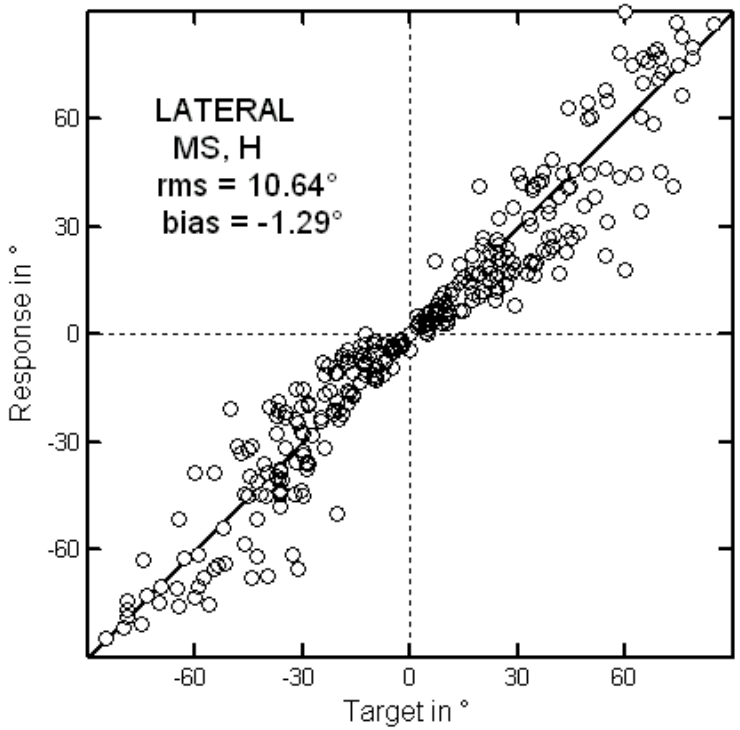
Visual Test

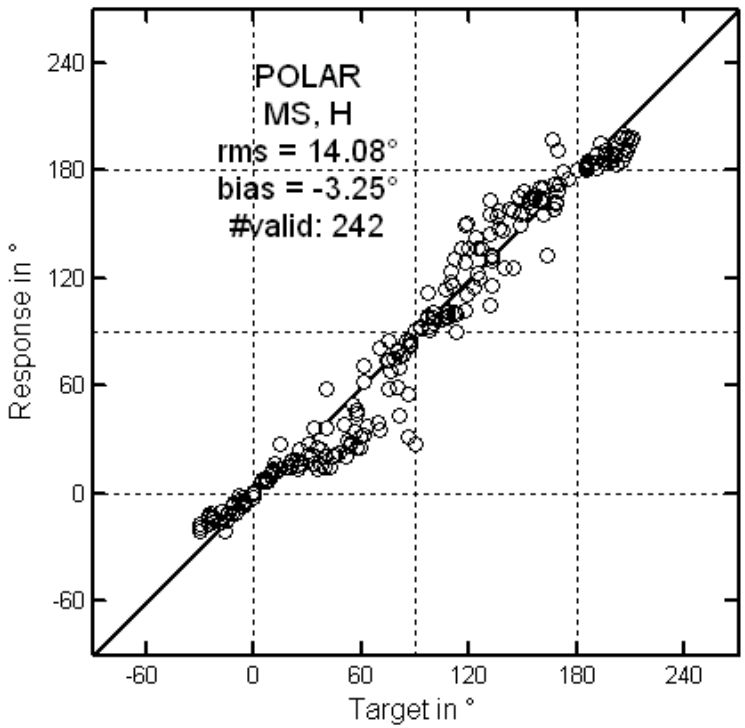
- Visual target **only**
- Presentation duration: 700 ms



Visual Test: Results

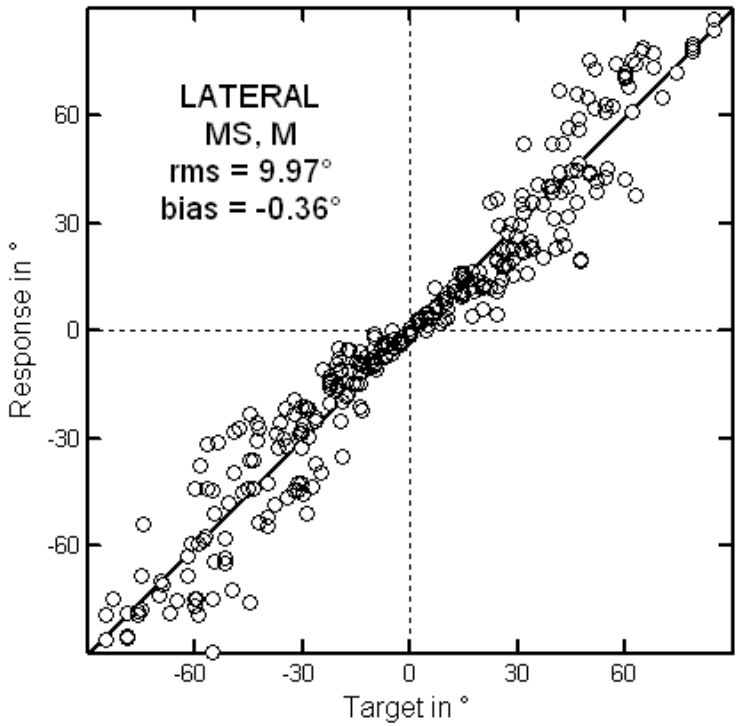
- Head pointer:

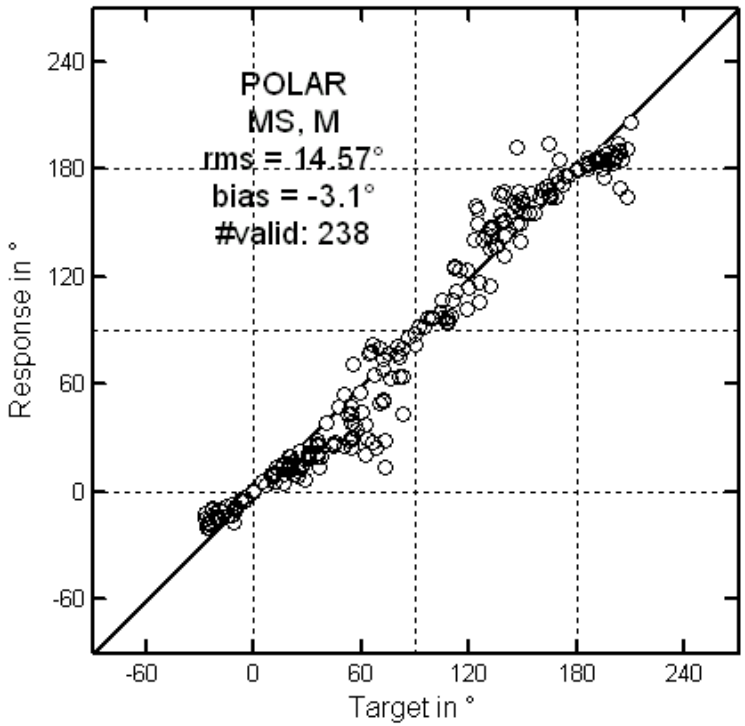




Visual Test: Results

- Manual pointer:

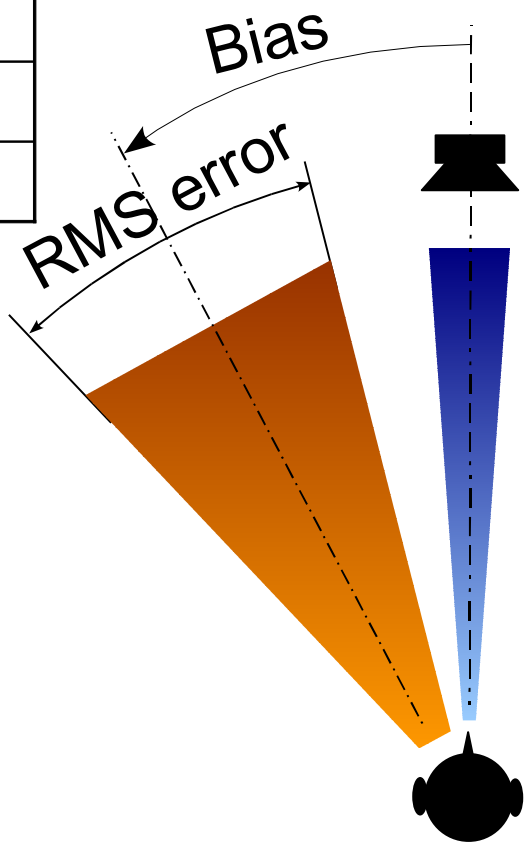




Visual Test: Statistics

Error	Head	Manual	p-value
Azimuth bias	0°	0.96°	0.034
RMS azimuth error	10,1°	9.44°	0.034
Elevation bias	0.9°	-0.1°	<0.001
RMS elevation error	8°	7.4°	0.005

- Small differences only
- Horizontal precision of about 10°
- Vertical precision of about 8°

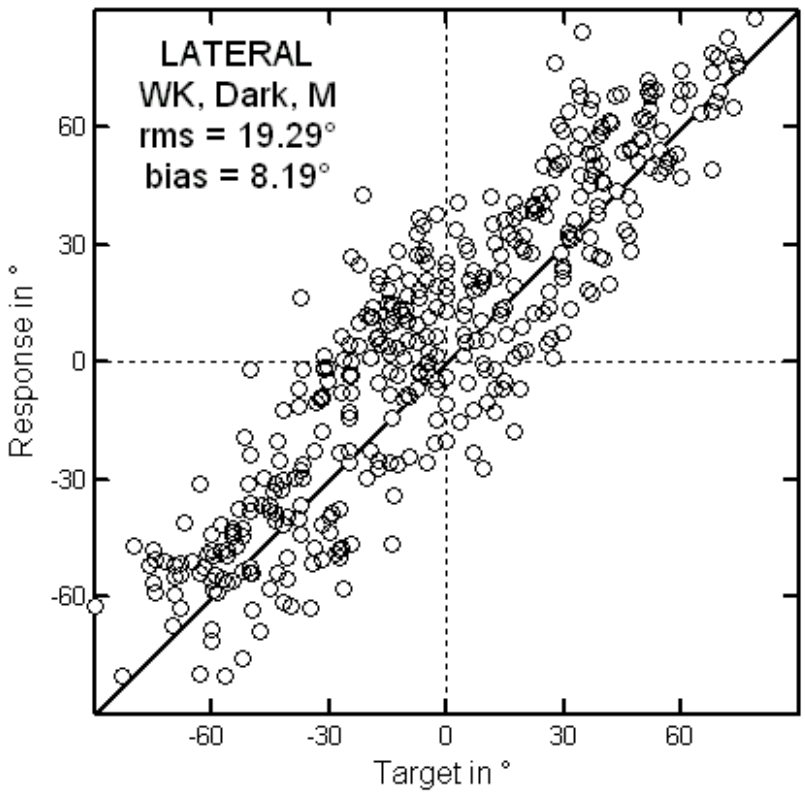


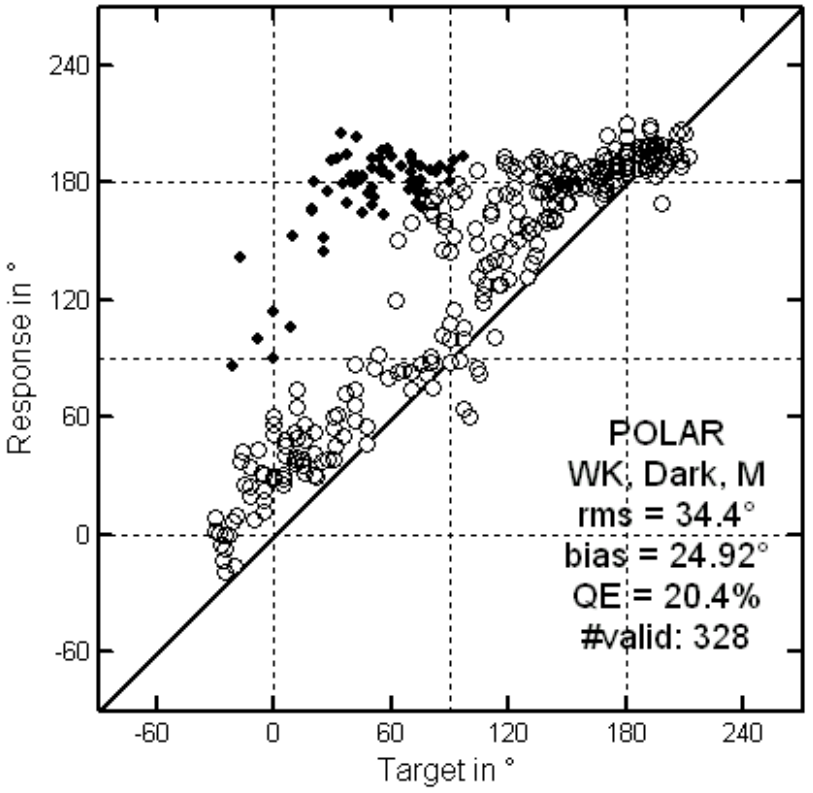
Acoustic Test: Methods

- Virtual acoustic stimuli (VAS):
  - Gaussian white noise, duration: 500 ms
  - Filtered with individualized DTFs
  - Presented via headphones
- Level:
  - Comfortable level of 50 dB SL
    - Hearing threshold: manual „one-up-two-down“ procedure resulting in a 73%-threshold for the VAS at the frontal position (0°; 0°)
  - Level roving in the range of 5 dB (trial-to-trial)
- Procedure: similar to the visual test

Acoustic Test: Results

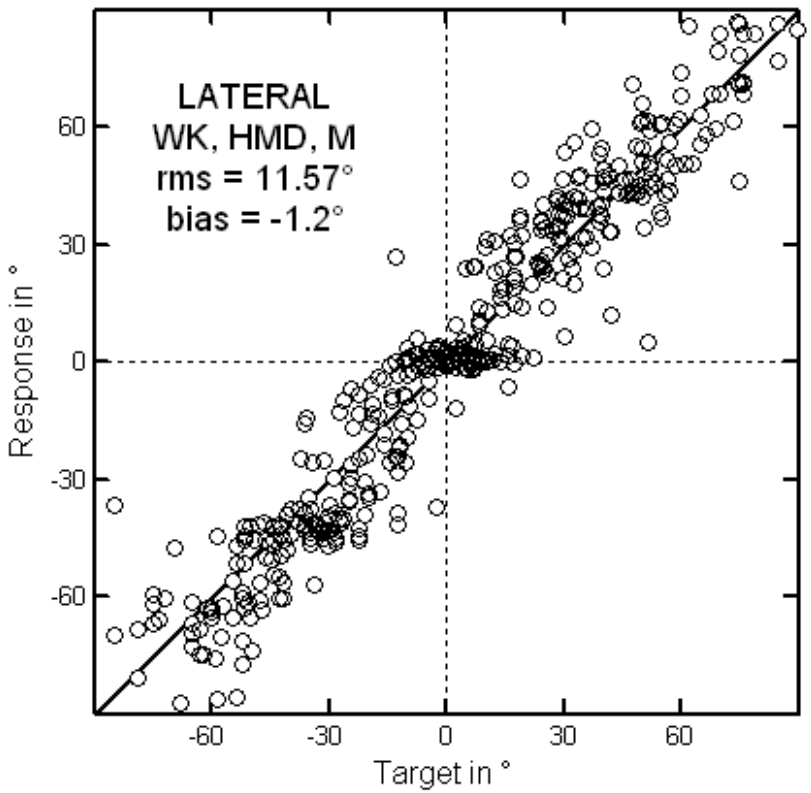
- Dark, Head pointer

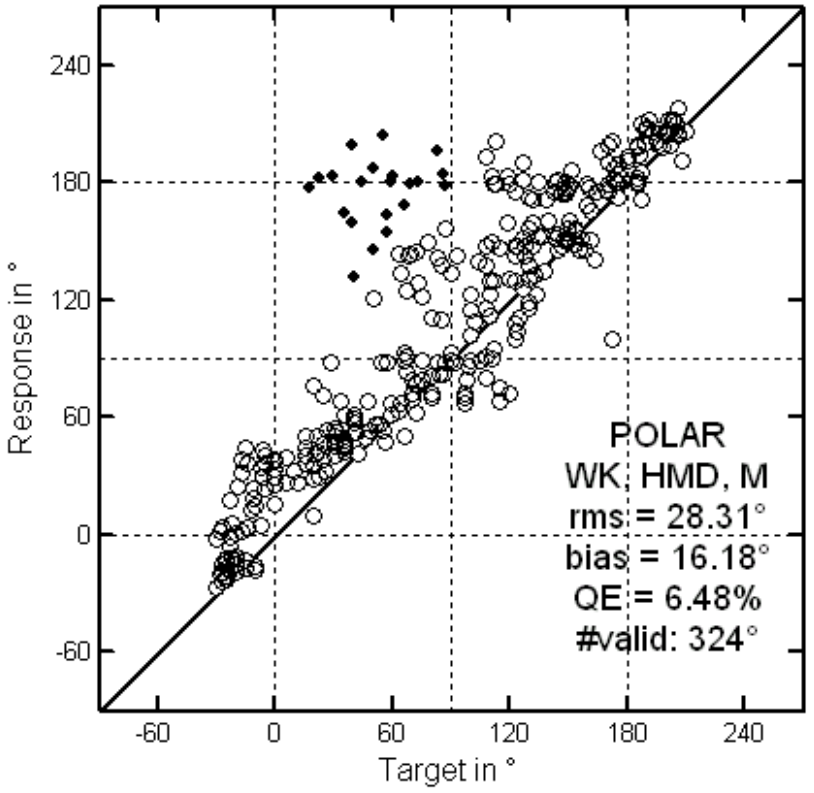




Acoustic Test: Results

- HMD, Manual pointer





Acoustic Test: Statistics

	Head Dark	Head, HMD	Manual, HMD	Manual, Dark
RMS lateral error	15.9°	13.3°	13.5°	16.7°
Lateral bias	6.5°	4.9°	5.5°	6.4°
quadrant errors, front (%)	30.3 %	26.2 %	23.1 %	30.2 %
quadrant errors, back (%)	15.6 %	15.6 %	14.4 %	14.4 %
Raw RMS polar error	52.5°	51.7°	48.2°	52.5°
Corrected RMS polar error	32.3°	30.5°	30.5°	32.3°

- Statistical analysis (RM ANOVA):
  - Visual environment: with HMD better ($p < 0.004$)
  - Head pointer: better in the horizontal plane ($p < 0.02$)
  - Manual pointer:
    - better precision in the vertical plane ($p < 0.047$) with HMD
    - less front/back confusions for rear-upper sounds ($p = 0.032$)

Summary: Visual

- Procedural training requires at least 600 targets:
  - Precision of 2° is possible
  - Shorter training period for head pointer
  - After 700 targets no differences between pointing methods
- Visual test:
  - Precision of 7° to 10° is possible
  - Head pointer: better in the horizontal plane
  - Manual pointer: better in the vertical plane

Summary: Acoustic

- Effect of the visual environment:
  - Smaller errors with visual environment
  - Limitations not because of visual deficits
- Effect of the pointing method:
  - Very small effect only
  - Head pointer: better in the horizontal plane
  - Manual pointer: sometimes better in the vertical plane
- However:
  - Errors higher than comparable literature probably...
    ...because the listeners were not trained or selected!