Comparison of Head and Gaze Behaviours in a Multi-talker Listening Task with a Head-Mounted Display and a Cylindrical Screen

Gerard Llorach[1,2], Maartje Hendrikse[1], Giso Grimm[1,3] and Volker Hohmann[1,2,3]

[1]Medizinische Physik and Cluster of Excellence 'Hearing4all', Universität Oldenburg, Germany; [2]Hörzentrum Oldenburg GmbH, Germany; [3]HörTech gGmbH Oldenburg, Germany

E-mail: g.llorach@hoerzentrum-oldenburg.de

Motivation

- **Why is it important for hearing aid's assessment?**

Performance of hearing aids changes with head movements and orientation. Gaze and head movements change depending on the display and visual cues [1]. Head-mounted displays (HMDs) have not been yet compared and validated in listening tasks.

- **Why head-mounted displays?**

Affordable, easy setup, space independent, immersive.

- **Why use a multi-talker listening task?**

It's a simple scenario close to a possible real-life situation, where the subject has to follow several conversations between 4 speakers.

Conditions

Audiovisual conditions

Audio only

When using the HMD, the background scene is shown without virtual characters or agents. This way the user has a reference of the space (chair, emergency button, room space…).

Virtual characters

Characters resembling the actors of the video recording with speech-driven lip-syncing [2] and automatic eye-head gaze to target speaker.

Video recordings

Recordings of the conversations with real actors. They are presented in the 3D scene as 2D video textures.

Display conditions

Head-mounted display (HMD)

Unity + HTC Vive
Field of view: 110º
Field of regard: all around

Cylindrical screen

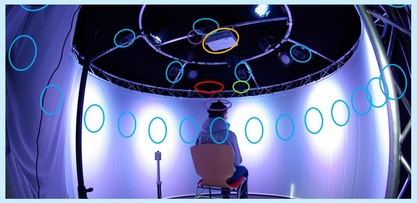
Blender + Projector
Field of view: 120º
Field of regard: static 120º

Method

Setup

The gesture lab

- Loudspeaker array controlled by a real-time audio engine, TASCAR [3].
- Head tracking with an infrared camera and markers or the HMD.
- Eye tracking in the horizontal plane with EOG.



Stimuli

- Conversations of 1-2 minutes between 4 speakers about every day topics (holidays, weather, cooking…) with background noise (SNR ~6dB). The speakers located at 45, 15, -15 and -45 angles in a circle of radius 1.7m.

Physiological Measures

- Number of gaze jumps
- Gaze direction difference with the target speaker position
- Head behaviour differences between conditions

Questionnaires

- Custom questionnaire addressing:
  - Comfort
  - Satisfaction/Enjoyment
  - Perceived listening effort
  - Perceived performance
  - Presence/sense of being there
  - Preference of display device and audiovisual condition
- Simulator Sickness Questionnaire

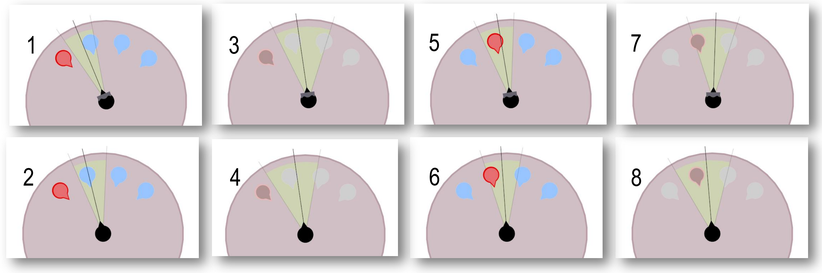
Hypothesis and Questions

- Head and eye behaviours will be similar with both display conditions, but head direction will be more towards the target speaker in HMD due to the smaller field of view.
- The cylindrical screen will be found to be more comfortable.
- The HMD will be found to be more immersive and enjoyable.
- Due to the reduced field of view, the HMD display will induce more head movements.

Preliminary Results (4 pilots)

Target speaker position*	Conditions	Average head direction	Percentiles (5% and 95%)	Figure number
45º	Virtual characters and video with HMD	22.47º	10.70º , 34.71º	1
45º	Virtual characters and video with cylindrical screen	13.82º	-2.51º , 23.85º	2
45º	Audio only with HMD	8.42º	-17.35º , 26.46º	3
45º	Audio only with cylindrical screen	8.18º	-10.52 , 32.62º	4
15º	Virtual characters and video with HMD	9.47º	-2.10º , 25.18º	5
15º	Virtual characters and video with cylindrical screen	4.02º	-10.55º , 18.08º	6
15º	Audio only with HMD	-1.91º	-17.67º , 17.75º	7
15º	Audio only with cylindrical screen	4.80º	-13.99º , 30.60º	8

* When the target speakers where on a negative angle (-45º and -15º), angles of the head direction were flipped for the computation of results.



Figures. From left to right: virtual characters with cylindrical screen, video recordings with cylindrical screen, virtual characters with HMD and video recordings with HMD.

References

[1] Hendrikse, M.M.E. et al., 2017. Head and eye movement behavior in realistic virtual audiovisual environments for hearing aid research. In 20. Jahrestagung der Deutschen Gesellschaft für Audiologie. Aalen.

[2] Llorach, G., Evans, A., Blat, J., Grimm, G., & Hohmann, V. (2016, September). Web-based live speech-driven lip-sync. In Games and Virtual Worlds for Serious Applications (VS-Games), 2016 8th International Conference on (pp. 1-4). IEEE.

[3] Grimm, G., Luberadzka, J., Herzke, T., & Hohmann, V. (2015). Toolbox for acoustic scene creation and rendering (TASCAR): Render methods and research applications. In F. Neumann (Ed.), *Proceedings of the Linux Audio Conference*. Mainz, Germany: Johannes Gutenberg Universität Mainz.