Comparing headphone- and loudspeaker-playback using spatial scene analysis in virtual audio-visual reverberant environments

Ahrens, Axel; Neher, Tobias; Dau, Torsten

Published in:
ICA 2022 Pproceedings

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022, in Gyeongju, Korea

Extended abstract

ABS-0481

Comparing headphone- and loudspeaker-playback using spatial scene analysis in virtual audio-visual reverberant environments

Axel AHRENS[1,2]; Tobias NEHER[1]; Torsten DAU[2]

[1] Research unit of Oto Rhino Laryngology, Department of Clinical Research, University of Southern Denmark, Denmark

[2] Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Denmark

ABSTRACT

In multi-talker situations, listeners face the challenge of having to identify a target speech source out of a mixture of interfering sources. To test these realistic situations in the laboratory it is common to employ arrays with many loudspeakers. Allowing clinics to use such experimental paradigms, headphone playback can be an alternative. However, headphone reproduction often leads to different outcomes because of the lack of individual head-related transfer functions (HRTFs). Here, we investigated how normal-hearing listeners analyze virtual audio-visual scenes that differed in terms of the number of concurrent talkers and the amount of reverberation using either head-tracked headphones using generic HRTFs or via a loudspeaker-array. Listeners were asked to identify and locate an ongoing story in a mixture of other stories. The visual components of the scenarios were reproduced via virtual reality headset. Differences between loudspeaker and headphone-based reproduction were evaluated using the response time and the localization accuracy. The number of talkers as well as the amount of reverberation affected the ability to analyze an audio-visual scene, in terms of response time and localization error. Preliminary data suggest only slightly elevated response times when employing headphone reproduction compared to loudspeakers, making them a valid tool for clinical research.

Keywords: Hearing, Speech, Virtual Reality

1. INTRODUCTION

When multiple talkers are present in a scene, listeners face the challenge to perceptually separate them, to be able to understand the speech. In a previous study a method to analyze a listener's ability to analyze a scene consisting of multiple talkers has been presented [1]. The listeners' task was to identify and locate a spoken story in the presence of other stories. Within the experiment, the number of simultaneous talkers (stories) and the amount of reverberation were varied. The acoustic information was presented via a spherical 64-channel loudspeaker array (see [2] for details) and visual information via virtual reality glasses. However, in clinical practice as well as in research such loudspeaker array is often not available. Thus, the sound reproduction via headphones is desirable.

When employing headphones for the reproduction of realistic spatial audio, head-related transfer functions (HRTFs) are needed. HRTFs reflect the acoustical properties of the human torso, head and pinna. While the use of individual HRTFs allow for an accurate spatial reproduction of a sound, the measurement of individual HRTFs needs either a large loudspeaker setup or long time when using few loudspeakers. Thus, using a generic HRTF that are representative of a number of listeners, for example one measured on a dummy head is of advantage. However, using a non-individual HRTF has been shown to affect sound perception (e.g., [3]).

The aim of the current study was to investigate the difference between loudspeaker reproduction and headphone reproduction using generic HRTFs in a speech perception task that closer reflects listening in real-life situations. Visual stimuli were presented via virtual reality glasses. The two outcome measures that were analyzed are the response time to identify and locate a target story and the localization error. The target stories were presented between ±105°, in 15° steps.

2. RESULTS

Figure 1 shows the response time to identify and locate the target story. The response time increases

with the number of simultaneously presented talkers as well as with reverberation. No effect of the reproduction method can be seen. Figure 2 shows the localization error. As for the response time, an effect of the number of talkers can be seen. Furthermore, more and larger errors can be seen for the headphone reproduction than for the loudspeaker reproduction.

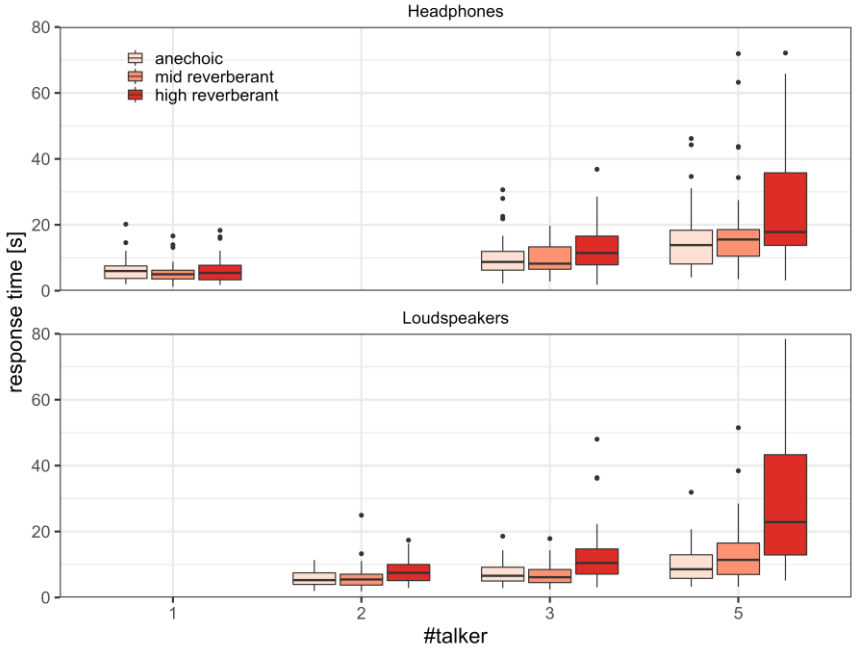

Figure 1. Response time from audio onset until the listener identified and located the target story. The colors indicate the reverberation conditions. Top panel: Headphone reproduction. Bottom panel: Loudspeaker reproduction.

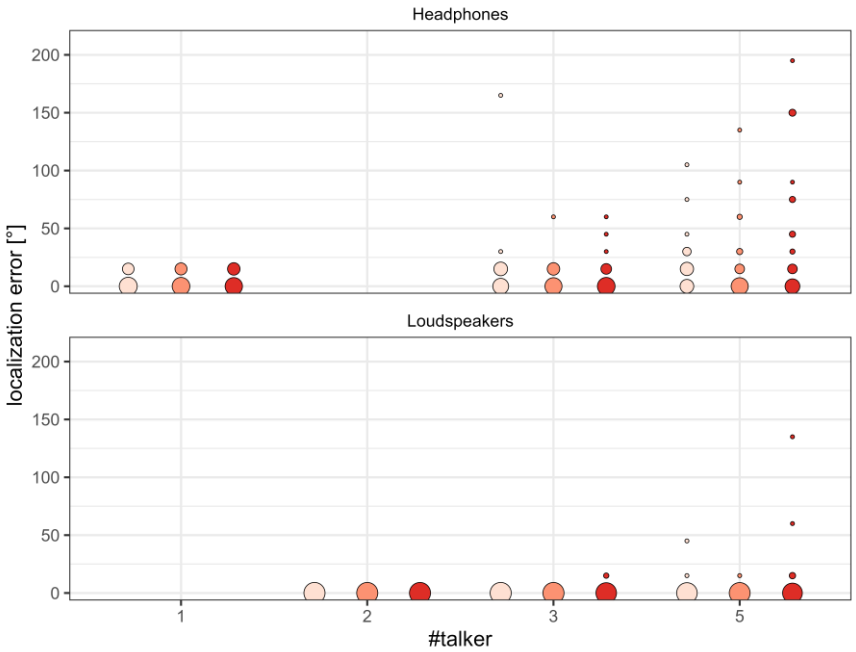

Figure 2. Absolute localization error of the target story in the azimuthal plane. The bubble sizes in each column visualize the error distribution and add up to the same size. The colors indicate the reverberation conditions (see Figure 1). Top panel: Headphone reproduction. Bottom panel: Loudspeaker reproduction.

3. CONCLUSIONS

Headphone- and loudspeaker-reproductions were analyzed within a speech comprehension and localization task. Two outcome measures showed different findings. The response time to identify and locate a speech source was similar with both reproduction methods. The localization error was worse with the headphone reproduction compared to the loudspeaker reproduction.

REFERENCES

1. Ahrens A, Duemose Lund, K. Auditory spatial analysis in reverberant audio-visual multi-talker environments with congruent and incongruent visual room information. bioRxiv 2022. 2022.04.30.490125. https://doi.org/10.1101/2022.04.30.490125.

2. Ahrens A, Marschall M, Dau T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. Hearing research 2019; 377. p. 307-317.

3. Møller H, Sørensen MF, Jensen CB, Hammershøi D. Binaural technique: Do we need individual recordings?. Journal of the Audio Engineering Society 1996; 44(6). p. 451-464.