forum acusticum 2023

THE EFFECT OF TALKER DISTRIBUTION ON SPEECH PERCEPTION WITH BEAMFORMERS IN VIRTUAL AUDIO-VISUAL ENVIRONMENTS

Laurine Dargaud[1] **Quirin Mühlberger**[1] **Sebastian Zeest Rydahl**[1]
Sven-Gustav Thiesen[1] **Axel Ahrens**[1*]
[1] Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Denmark

ABSTRACT

Beamformers have the potential to greatly improve speech intelligibility for hearing aid users. However, most studies investigate beamformers in scenarios with source locations that are ideal for a beamformer and with static listeners. Here, we present a study that investigates the effect of beamformers in two reverberation conditions and with varying number of talkers, positioned either closely or widely spaced. Target locations were varied in the frontal hemisphere and participants could move their head freely. Audio-visual scenes were reproduced using a 64-channel loudspeaker array and virtual reality glasses. The listeners' task was to find a story in a mixture of other stories and correct identification and response time were used as outcome measures. The results show improved speech perception with beamformers in comparison to not using a beamformer. The beamformers' performance was found to be dependent on the distribution of the sources. Larger improvements of speech perception due to the beamformer were found when the reverberation time was higher. These findings demonstrate the potential of beamformers to improve speech perception in environments that are more like real-world scenes in comparison to previous studies.

Keywords: *Speech, Virtual Reality, Beamformer, Virtual Audio, Scene Analysis*

[*] ***Corresponding author***: aahr@dtu.dk

1. INTRODUCTION

Understanding speech in situations with many interfering speech sources can be challenging. Particularly listeners with hearing impairment report to have difficulties in such scenarios. Beamformers in hearing aids have been shown to be able to improve speech perception.

For example, Valente et al. [1] measured the effect of a hearing aid beamformer with a target sentence from the front and a noise from the back and found an average improvement of up to 8.5 dB in speech intelligibility. More recently, Moore et al. [2] evaluated personalized binaural hearing aid beamformers and showed effects of up to 55% better speech perception in comparison to an omni-directional setting. They employed a fixed target positions at either 0° or 67.5° azimuth and a relatively diffuse background noise consisting of eight spatially distributed babble noises. The stimuli were presented over headphones, thus not including head movements. The advantage of the beamformer was reduced for the lateral source at 67.5° in comparison to the source at 0°.

Best et al. [3] investigated the performance of a visually guided hearing aid beamformer using a paradigm closer to real-world listening situations. The background noise consisted of spatially distributed conversations and the target location was either static or dynamic. They showed that the performance of the beamformer was reduced when the target direction was dynamic. However, no head motions were included in the setup.

While these test setups show the advantage of using beamformers for speech perception, they either inform the test participants about the source location and/or employ fixed head positions. Here, we use a dual-task paradigm as in [4] where the test participants are asked to locate and comprehend ongoing speech. This paradigm allows us to investigate the effect of beamformers on speech perception

in more realistic conditions. During the task the participants can freely move their heads, thus, being closer to scenarios as perceived outside the laboratory. Furthermore, we investigate the influence of the spatial distribution of speech sources on the effectiveness of beamformers with different bandwidths.

2. METHODS

2.1 Test participants

Eight native Danish speaking participants took part in the experiment. All participants were part of the course 'Experimental Hearing Science' which is part of the Engineering Acoustics program at the Technical University of Denmark. All participants provided informed consent and the experiment was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

2.2 Virtual audio-visual environment

The experiment was conducted in the audio-visual immersion lab at the Technical University of Denmark. Acoustic stimuli were presented via a 64-channel spherical loudspeaker array. The room acoustics were simulated using Odeon and the loudspeaker signals were created using the nearest-loudspeaker mapping implemented in the LoRA-toolbox [5].
A visual room that matches the room acoustics was presented via HTC Vive Pro Eye Virtual Reality glasses. The visual scenes were controlled via Unity3D. For more details on the implementation see previous publications [4,6].

2.3 Beamformer implementation

In this study a beamformer was simulated by applying spatial gains as in [7]. The beamformer was directed using the eye-gaze direction tracked via the Virtual Reality glasses. Three conditions were tested, an omni-directional setting where no beamformer was applied, a narrow beamwidth and a wide beamwidth. The spatial gains are shown in Figure 1. The narrow pattern (yellow) was chosen to isolate a single talker. The wide one (blue) to allow to also get information on neighboring talkers.

2.4 Experimental paradigm

The experimental paradigm was as in Ahrens & Lund [4]. Participants were instructed to locate a single story out of a mixture of other spatially distributed stories as fast as possible. Responses were collected by pointing and clicking the controller of the Virtual Reality glasses. Before the experiment, participants were familiarized with all stories. The main outcome measure was the response time, i.e., the time between audio onset and the participant's response.

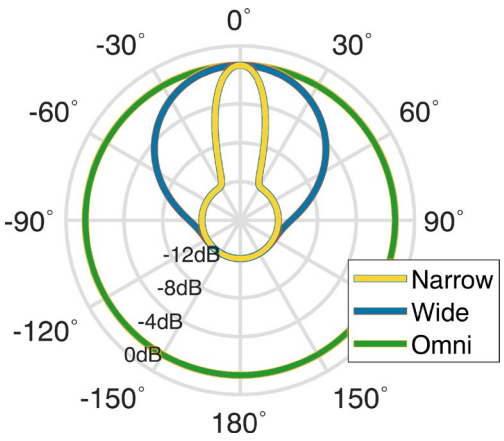

Figure 3. Polar patterns of the three beamformer conditions. The design is based on [7].

2.5 Experimental conditions

Four factors were tested in all test participants and repeated three times. The reverberation was varied between low reverberation and high reverberation. The low reverberation condition had a reverberation time of approximately 0.4 seconds and the high reverberation time of approximately 1.8 seconds. The number of simultaneous talkers was either 2, 4, or 6, including the target talker. The beamformer was either 'off'/omni-directional, narrow-beam or wide-beam.
Lastly, the spatial distribution of the sources was varied. The target story was located randomly on a discrete grid, spaced 15°, between ±105° on the horizontal plane. The distribution of the interferers was either compact or widely spread. In the compact setting, all sources were placed next to each other with a spacing of 15°. In the widely spread setting, the sources were placed far from the target story.

3. RESULTS

3.1 Response time in low reverberation

Figure 2 shows the response time in the low reverberation condition. The top panel shows the compact spacing of the sources and the bottom panel shows the spread-out spacing. The colors indicate the beamformer condition. The general trend of an increase in the response time with an increase of the number of talkers can be observed. This increase is

larger in the omni-directional condition in comparison to the conditions with the narrow and wide beamformer.
In the low reverberation condition an effect of the beamformer can only be seen when six talkers are presented but not with two or four talkers. No difference between the two beamwidths can be seen. The response time with the beamformer is on average about 20 seconds shorter in comparison to the omni-directional setting.
Comparing the compact and the spread-out spacing of speech sources shows only an effect for six talkers, where the response time is faster when the talkers are widely spread. This effect can only be observed in the omni-directional condition.

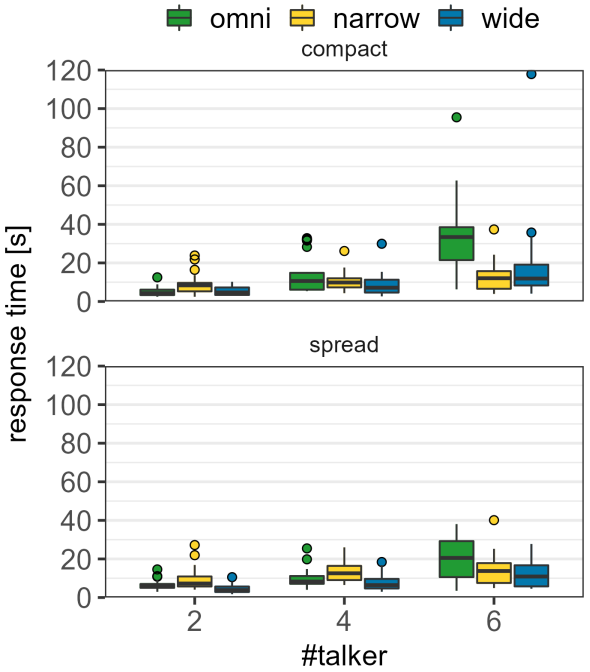

Figure 2. Response time in seconds in the condition with low reverberation. The colors indicate the beamformer configuration. The panels indicate the conditions with a compact (top) and spread-out (bottom) source distribution. Boxplots indicate the median and 1st/3rd quartile. The whiskers include 1.5 times the interquartile range.

3.2 Response time in high reverberation

Figure 3 shows the response time in the high reverberation condition. Generally, higher response times can be observed in comparison to the low reverberation condition.

In the condition with high reverberation, an effect of the beamformer was observed with four and six talkers. The reduction in the response time was larger with six than with four talkers.
Comparing the spatial distributions of the talkers shows that in the omni-directional setting, the response time decreases when the talkers are spread-out for four talkers but increases for six talkers. Furthermore, an interaction between the beamformer and the source distribution was found. The wider beamwidth is advantageous when the sources are spread out. When the sources are close together, the narrow beam is advantageous.

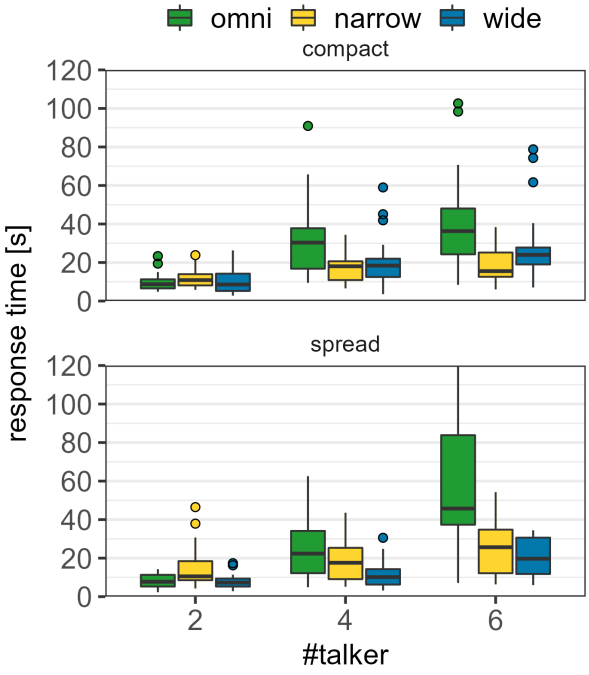

Figure 3. Response time in seconds in the condition with high reverberation. The colors indicate the beamformer configuration. The panels indicate the conditions with a compact (top) and spread-out (bottom) source distribution. Boxplots indicate the median and 1st/3rd quartile. The whiskers include 1.5 times the interquartile range.

4. CONCLUSIONS

We employed a scene analysis task, comprised of speech localization and comprehension, to investigate the effect of beamformers on speech perception with spatially distributed sources. The results revealed an advantage of the

beamformer in more difficult situations in terms of reverberation and number of talkers. Furthermore, we found an interaction between the beamformer configuration and the spatial distribution. Thus, the effectiveness of beamformers depends on the spatial configuration and needs to be evaluated in more realistic and complex scenarios including head motion.

5. ACKNOWLEDGMENTS

The authors would like to thank the responsibles of the course 'Experimental Hearing Sciences' at the Technical University of Denmark, Abigail Kressner and Jeremy Marozeau. A.A. was supported via a grant from the Independent Research Fund Denmark awarded to Tobias Neher, Torsten Dau, Virginia Best, and Adam Westermann.

6. REFERENCES

[1] M. Valente, D. Fabry, and L.G. Potts: "Recognition of speech in noise with hearing aids using dual microphones", Journal of the American Academy of Audiology, vol. 6, no. 6, 1995.

[2] A.H. Moore, J.M. de Haan, M. Syskind Pedersen, P.A. Naylor, M. Brookes, and J. Jensen: "Personalized signal-independent beamforming for binaural hearing aids", The Journal of the Acoustical Society of America, 145 (5), 2019. https://doi.org/10.1121/1.5102173

[3] V. Best, E. Roverud, T. Streeter, C.R. Mason, G. Kidd: "The Benefit of a Visually Guided Beamformer in a Dynamic Speech Task", Trends in Hearing 21, 2017. https://doi.org/10.1177/2331216517722304

[4] A. Ahrens, and K. Duemose Lund: "Auditory spatial analysis in reverberant multi-talker environments with congruent and incongruent audio-visual room information", The Journal of the Acoustical Society of America, 152 (3), 2022. https://doi.org/10.1121/10.0013991

[5] S. Favrot, and J.M. Buchholz: "LoRA: A loudspeaker-based room auralization system", Acta Acustica united with Acustica, 2010.

[6] A. Ahrens, K. Duemose Lund, M. Marschall, and T. Dau: "Sound source localization with varying amount of visual information in virtual reality", PLoS ONE 14(3): e0214603, 2019. https://doi.org/10.1371/journal.pone.0214603

[7] L. Hladek, B. Porr, G. Naylor, T. Lunner, and W.O. Brimijoin: "On the Interaction of Head and Gaze Control With Acoustic Beam Width of a Simulated Beamformer in a Two-Talker Scenario", Trends in Hearing 23, 2019. https://doi.org/10.1177/2331216519876795