Spatial sound reproduction for interactive hearing research

Matthieu Kuntz[1], Bernhard U. Seeber[2]
Audio Information Processing, Technical University of Munich
Arcisstrasse 21, 80333 München, Germany

ABSTRACT

The use of sound field synthesis for hearing research has gained popularity due to the ability to auralize a wide range of sound scenes in a controlled and reproducible way. We are interested in reproducing acoustic environments for interactive hearing research, allowing participants to move freely over an extended area in the reproduced sound field. While the physically accurate sound field reproduction using sound field synthesis is limited to the sweet spot, it is unclear how different perceptual measures vary across the reproduction area and how suitable sound field synthesis is to evaluate them. To investigate the viability of listening experiments and provide a database for modelling approaches, simulations of binaural cues were carried out for the Simulated Open Field Environment loudspeaker array. Results show that the binaural cues are reproduced well close to the array's center, but exhibit more variance than in the corresponding free field case further to the sides. The binaural benefit in speech understanding is reproduced well in the center, with errors lower than 1 dB, but errors increase up to 6 dB close to the loudspeakers. We show that these deviations are driven by errors in better-ear signal-to-noise ratio and that the binaural unmasking contribution is reproduced faithfully.

1. INTRODUCTION

Virtual acoustic environments have gained traction in hearing research due to their ability to recreate complex acoustic scenes under laboratory conditions, ensuring reproducible conditions for every participant, paving the way for more objective assessments of realistic scenarios with respect to speech understanding, hearing aid and cochlear implant development or behavioral research. For these aspects, it is crucial to be able to rely on virtual acoustics and sound field synthesis to present correct sound pressure levels and interaural cues to the participants and in a similar fashion to a measurement device.

With the goal of creating increasingly realistic scenarios, allowing participants to move freely inside a synthesized sound field is the next logical step. The evaluation of physical accuracy of the reproduced sound pressure level and the resulting interaural cues on a wider reproduction area in a loudspeaker array in an anechoic chamber was already the focus of previous work [1, 2]. In particular, it was shown that the interaural coherence (IC) resulting from the auralization of a single virtual source using Higher-Order Ambisonics (HOA) is low at off-center positions, potentially affecting speech understanding or spatial unmasking [2]. Speech intelligibility in virtual acoustic environments was investigated by Ahrens *et al.* [3] with a participant seated in the center of a loudspeaker array in

---

[1] matthieu.kuntz@tum.de

[2] seeber@tum.de

an anechoic chamber, showing that "no significant differences appeared in the highest fidelity reproduction setup"[3]. Other studies investigating speech intelligibility in virtual acoustic environments were also carried out with seated participants in the center, such as [4, 5].

In this paper, we present a simulation of the binaural benefit for speech intelligibility in noise that would occur over a wide area in a virtual acoustic environment, based on the computation of binaural signals and an auditory model.

2. METHODS

2.1. Simulation setup

In order to evaluate speech reception thresholds and the binaural benefit that are to be expected in virtual acoustic environments, we simulated the loudspeaker array of the Simulated Open Field Environment (SOFE, [6, 7]), installed in an anechoic chamber at the Technical University of Munich. For this work, we limited ourselves to the 2D loudspeaker arrangement, consisting of 36 loudspeakers in $10°$ spacing, arranged on a square frame. The loudspeakers are equalized in level, frequency response, and phase response at the center of the array. In this simulation, the loudspeakers are modeled as perfect point sources.

The binaural signals were computed by convolving each loudspeaker signal with a set of HRTFs measured in $0.5°$ steps with a HMS II.3 artificial head (HEAD acoustics GmbH, Herzogenrath, Germany). Each loudspeaker was assigned the closest HRTF based on its relative position to the listener. The binaural signals were simulated on an 11-by-11 points grid, ranging from -1.6 m to 1.6 m around the center of the loudspeaker array in the x and y direction, the same as in previous work [2].

2.2. Stimuli and conditions

We simulated four spatial arrangements of a target source and a masker, defined in reference to the center of the loudspeaker array:

- N_0S_0: The target and the masker are collocated at $0°$ azimuth.
- N_0S_{60}: The target is located at $60°$ azimuth and the masker at $0°$ azimuth.
- N_uS_0: The target is located at $0°$ azimuth and the masker is diffuse.
- N_uS_{60}: The target is located at $60°$ azimuth and the masker is diffuse.

The target stimulus was a sentence from the OLSA sentence test (Hörzentrum Oldenburg gGmbH, Oldenburg, Germany), spoken by a male speaker. The masker consisted of white noise with a bandwidth of 20 Hz to 20 kHz. Both the target and the masker were set to 65 dB SPL. The stimuli were played back using 17th-order 2D Higher-Order Ambisonics [2, 8]. The diffuse noise was generated by playing white noise from all 36 loudspeakers.

In order to assess the results inside the loudspeaker array, a reference condition was also simulated, where the target and masker were simulated as point sources positioned at $0°$ or $60°$, at a distance of 4 m from the center of the loudspeaker array. The diffuse noise was kept constant across all positions.

To investigate the effect of panning angle on the binaural benefit, we simulated the same spatial arrangements, but rotated by $5°$. Thus, the sources had to be panned between two loudspeakers since they were located at $5°$ or $65°$. This usually leads to the strongest panning artefacts, such as increase in source width and coloration [8]. The virtual listener was also rotated by $5°$, to keep the spatial arrangement constant relative to the listener.

2.3. Modelling of speech intelligibility

We used the speech intelligibility model proposed by Jelfs *et al.* [9] as implemented in the Auditory Modeling Toolbox [10]. This model computes the binaural advantage in speech understanding by

combining the monaural better-ear signal-to-noise ratio (SNR) and the binaural masking level differences (BMLD). Since these two effects rely on fundamentally different mechanisms, we will also consider them separately, in addition to the overall binaural benefit in speech intelligibility.

3. RESULTS

We will first introduce the overall binaural benefit in speech intelligibility. The plots in Figure 1 show the difference in overall binaural benefit between the auralization and the reference, for all four spatial arrangements of target and masker.

Difference in binaural benefit in speech intelligibility between auralization and reference (dB)
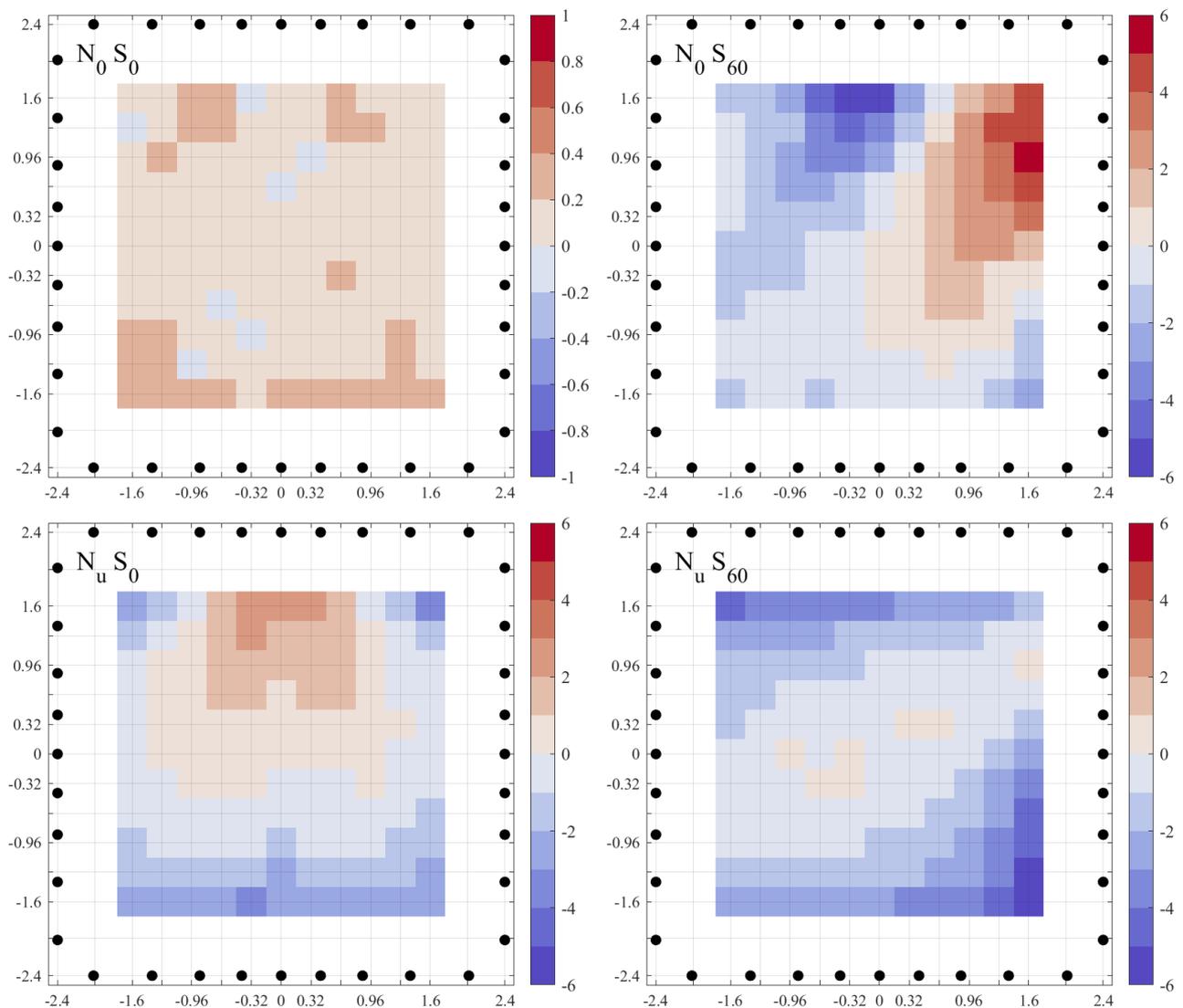

Figure 1: The difference in binaural benefit in dB obtained on an extended area inside a loudspeaker array (indicated by black dots) for different spatial arrangements of target and masker between the auralization and free-field reference. Positive values (red) indicate that the binaural benefit is higher for the auralization than for the reference condition. Note the different scale on the N_0S_0 condition.

The deviations in the N_0S_0 condition are negligible, their absolute value keeping below 0.2 dB at most evaluation points and below 0.4 dB everywhere on the evaluation area. The N_0S_{60} condition yields the highest deviations of the four spatial arrangements. While the errors around the center stay within ± 1 dB of the free-field reference, they reach -6 dB towards the front and 5 dB towards the side. These directions correspond to a movement towards the masker and the target respectively. For the N_uS_0 and N_uS_{60} conditions, the deviations are mostly within ± 2 dB around the center, with slightly higher errors towards the edges of the evaluation area. To better understand where these deviations originate, the monaural and binaural contributions are observed separately in the following sections.

Difference in the contribution of binaural unmasking to the binaural benefit in speech intelligibility between auralization and reference (dB)
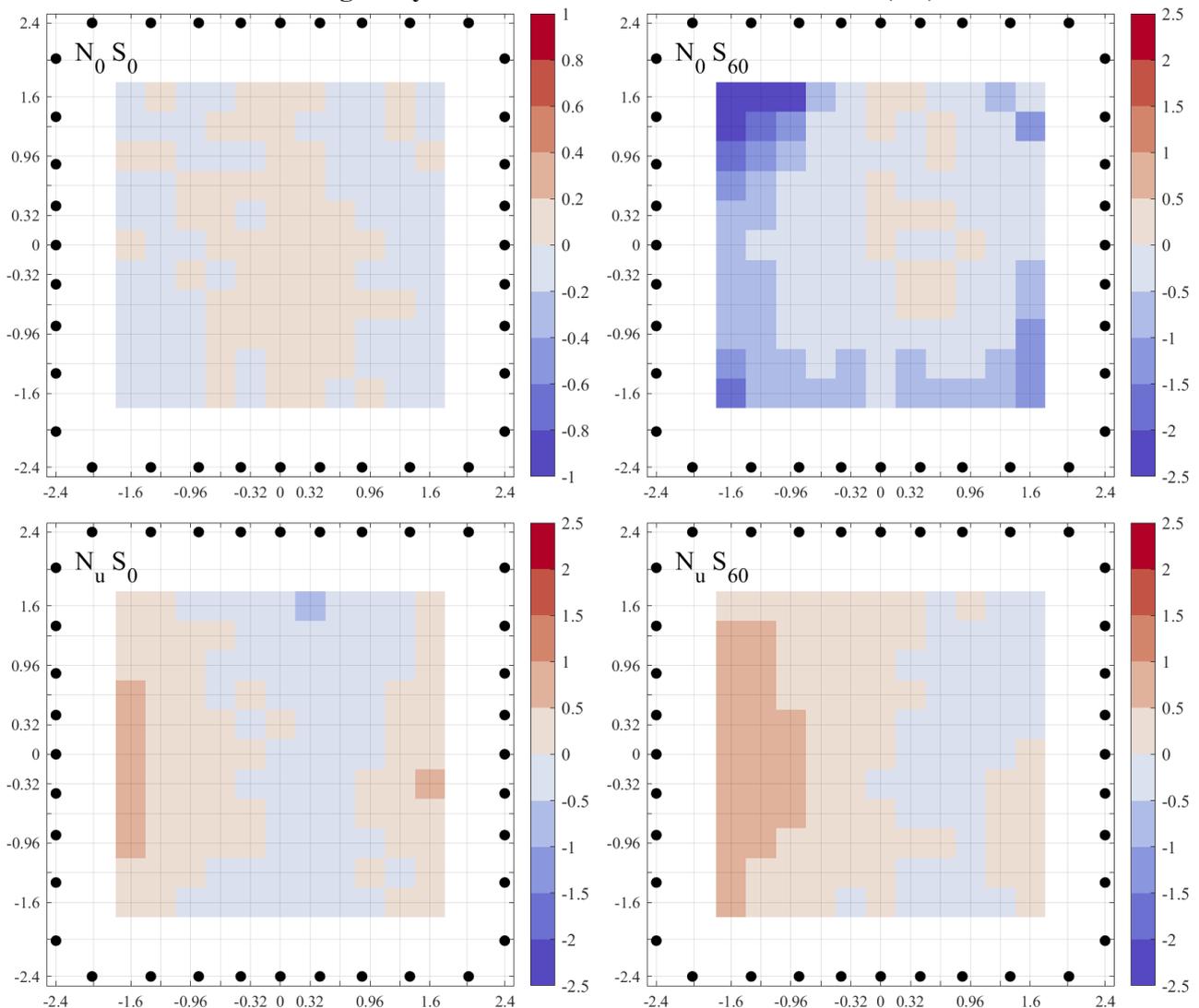

Figure 2: The difference in binaural unmasking in dB obtained on an extended area inside a loudspeaker array (indicated by black dots) for different spatial arrangements of target and masker between the auralization and free-field reference. Positive values (red) indicate that the binaural unmasking is higher for the auralization than for the reference condition. Note the different scale on the N_0S_0 condition.

3.1. Binaural unmasking contribution

The contribution of the binaural unmasking component to the speech intelligibility benefit is shown in Figure 2. The auralization is very close to the reference case, the errors are within ± 0.5 dB up to 0.96 m away from the center and below 1 dB for the N_0S_0, the N_uS_0, and the N_uS_{60} spatial arrangements. While the N_0S_{60} condition shows higher deviation than the others, they appear for distances above 1.4 m from the center.

Difference in the contribution of better-ear SNR to the binaural benefit in speech intelligibility between auralization and reference (dB)
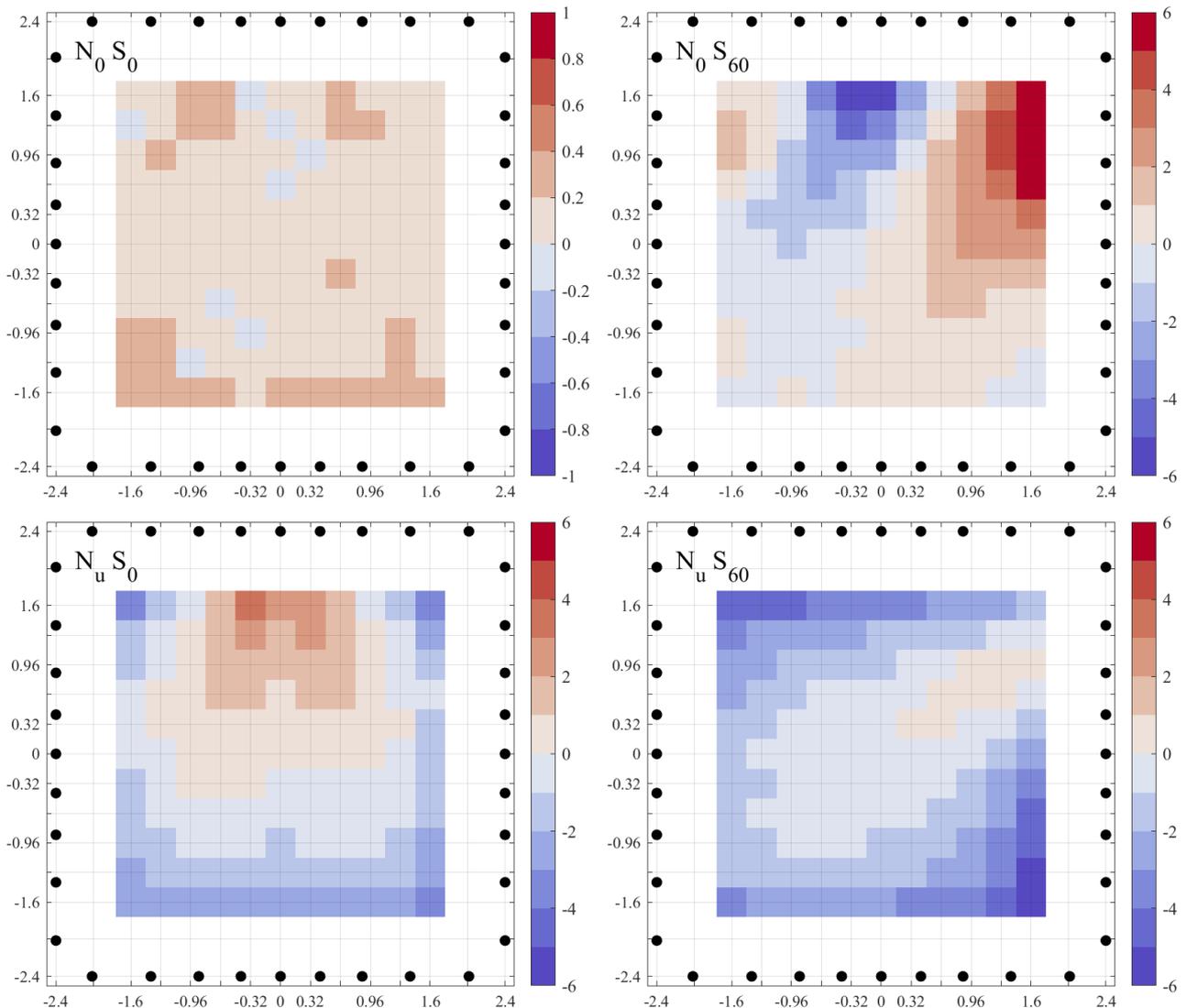

Figure 3: The difference in monaural better-ear SNR in dB obtained over an extended area inside a simulated loudspeaker array (indicated by black dots) for different spatial arrangements of target and masker between the auralization and free-field reference. Positive values (red) indicate that the monaural better-ear SNR is higher for the auralization than for the reference condition. Note the different scale on the N_0S_0 condition.

3.2. Better-ear signal-to-noise ratio contribution

The contribution of the better-ear signal-to-noise ratio component to the speech intelligibility benefit is shown in Figure 3. Larger deviations from the reference case are observable than for its binaural counterpart. The deviations for the N_0S_0 condition are very low. In the N_uS_0 and N_uS_0 conditions, the deviations are within ± 2 dB in a radius of 0.96 m around the center of the loudspeaker array. The N_0S_{60} condition shows the highest deviations (up to 6 dB) in predicted better-ear SNR benefit, especially for positions closer to the loudspeakers in directions of the masker ($0°$) or the target ($60°$).

3.3. Influence of panning direction

Figure 4 compares the deviations of overall binaural benefit, binaural unmasking, and better-ear SNR between the auralization and the reference for the $0°$ and the rotated $5°$ panning situations. The deviations in binaural unmasking (panel 2) are very similar. The median deviations in better-ear SNR are a little larger in the rotated condition, which also carries over to the overall binaural benefit. A spread of the values towards higher deviations is also observed.
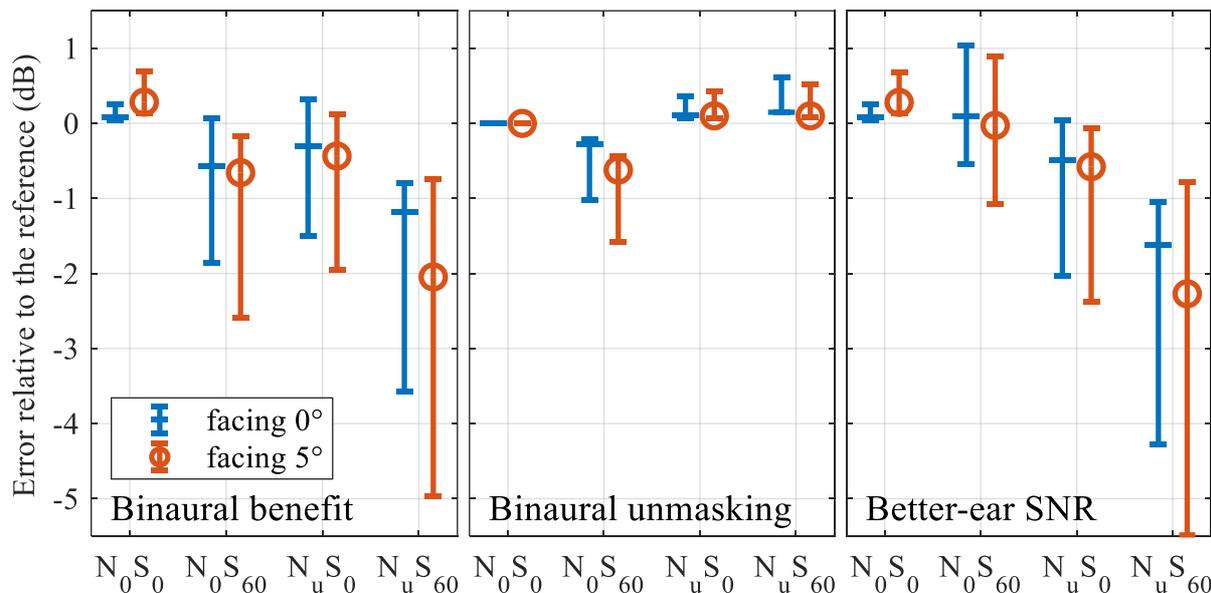

Figure 4: Binaural benefit for speech intelligibility (left panel) an the contributions of the binaural unmasking (center panel) and better-ear-SNR (right panel) components in dB. Data are pooled over an extended area inside a simulated loudspeaker array and are expressed as the difference between a Higher-Order Ambisonics reproduction and the corresponding free-field reference. Colors indicate the regular $0°$ and the $5°$ rotated panning situation. The markers and whiskers indicate the median and quartile values.

4. DISCUSSION

We presented a simulation of the binaural benefit for speech intelligibility across an extended area inside a loudspeaker array using the model by Jelfs *et al.* [9], allowing us to visualize the monaural and binaural contributions to the binaural benefit separately. The total binaural benefit is the sum of both contributions.

Considering the overall binaural benefit (see Figure 1), we observe that the errors are below 1 dB for distances below 0.32-0.64 m from the center of the loudspeaker array for the spatial arrangements tested here. The much lower errors in the N_0S_0 condition are coherent with the findings of Ahrens *et*

al. [3]. For listener positions further away from the center, the deviations increase, especially in the N_0S_{60} condition, where they reach 6 dB at the edge of the evaluation area towards the target or interferer position, 1.6 m away from the loudspeaker center. Note that this position is already very close to the loudspeakers, where participants would rarely stand if prompted to move around the center of the array.

Considering the contribution of the binaural unmasking component to the binaural speech intelligibility benefit, Figure 2 shows that the expected errors in binaural unmasking are low for all spatial arrangements. The error remains within ±0.5 dB in a circular area with a radius below 0.96 m, indicating that a participant's position inside the loudspeaker array only has a small effect on overall binaural unmasking. Hence, we do not expect the participant's position in that area to affect binaural speech unmasking to a problematic degree. This finding could be surprising considering the low IC of about 0.75 observed on the sides of the evaluation area in a previous study [2]. However, in the N_0S_0 condition, target and masker are collocated for every evaluation position, reducing the contribution of binaural unmasking to 0 dB. In the N_uS_0 and N_uS_{60} conditions, Kolotzek and Seeber [11] found the effect of the source's position in diffuse noise on binaural unmasking to be below 2 dB. While their observation is based on measured detection thresholds of a 500 Hz tone, this direction independence could potentially also be present for speech intelligibility.

We do observe larger deviations in the better-ear SNR prediction (see Figure 3), that could become problematic in speech intelligibility experiments where some conditions only differ by a couple decibels. These differences come from the mismatch between the loudspeaker (the physical sound sources) distances and the virtual source distance. This leads to a different 1/R distance dependent attenuation that creates different target-to-masker level ratios.

This behaviour does not change much for virtual sources panned exactly between two loudspeakers, as shown in Figure 4, which leads us to believe that this will not be different for other panning directions.

5. CONCLUSION

This work investigated the reproducibility of the binaural benefit in speech intelligibility over a loudspeaker array by comparing four spatial arrangements of a target speaker and a masking noise to a free-field reference. We observe that the overall binaural benefit shows errors below 1 dB around the center. The larger errors at the edges of the reproduction area are mainly driven by the better-ear SNR contribution. It was shown that the binaural unmasking contribution is reproduced accurately, with errors below 0.5 dB for distances to the center of the loudspeaker array of up to 0.96 m. The reproduction of better-ear SNR is less accurate, with errors below 1 dB for distances to the center of the loudspeaker array of up to 0.64 m. Larger errors observed at the edges are due to the distance mismatch between the array loudspeakers and the virtual source. The panning direction of the sources does not influence these results much, although there is a tendency towards slightly higher deviations when sources are panned between loudspeakers compared to the on-loudspeaker panning case.

This evaluation yields promising results about the feasibility of speech intelligibility experiments in the free-field in a loudspeaker array with moving participants.

6. ACKNOWLEDGEMENTS

REFERENCES

1. Kuntz, M., Kolotzek, N., Seeber, B.U. Gemessener Schalldruckpegel im Lautsprecherarray für verschiedene Schallfeldsyntheseverfahren. *Fortschritte der Akustik – DAGA '21*, 1437-1440 (2021)
2. Kuntz, M., Seeber, B.U. Sound field synthesis: Simulation and evaluation of interaural cues over an extended area. *Proceedings of the Euronoise 2021 Conference,* 1830-1839 (2021)
3. Ahrens, A., Marschall, M., Dau, T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. *Hearing Research,* **377**, 307-317 (2019)
4. Mansour, N., Marschall, M., May, T., Westermann, A., Dau, T. Speech intelligibility in a realistic virtual sound environment. *The Journal of the Acoustical Society of America*, **149(4)**, 2791-2801 (2021)
5. Au, E., Xiao, S., Hui, C.T.J., Hioka, Y., Masuda, H., Watson, C.I. Speech intelligibility in noise with varying spatial acoustics under Ambisonics-based sound reproduction system. *Applied Acoustics*, **147**, 107707 (2021)
6. Seeber B.U., Clapp, S. Interactive simulation and free-field auralization of acoustic space with the rtSOFE. *The Journal of the Acoustical Society of America*, **141(5)**, 3974 (2017)
7. Seeber, B.U., Kerber S., Hafter E.R. A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hearing Research,* **260(1-2)**, 1-10 (2010)
8. Zotter, F., Frank, M. *Ambisonics: A practical 3D audio theory for recording*, Springer (2019)
9. Jelfs, S., Culling, J.F., Lavandier, M. Revision and validation of a binaural model for speech intelligibility in noise. *Hearing Research,* **275(1-2)**, 96-104 (2011)
10. Majdak, P. Hollomey, C., Baumgartner, R. AMT 1.0: The toolbox for reproducible research in auditory modeling, *http://amtoolbox.org/*
11. Kolotzek, N., Seeber, B.U. Binaurale Entmaskierung zirkulär bewegter Schallquellen. *Fortschritte der Akustik – DAGA '19*, 850-851 (2021)