DOI: 10.61782/fa.2025.0315

FORUM ACUSTICUM EURONOISE 2025

DYNAMIC LOUDSPEAKER EQUALIZATION FOR PARTICIPANT MOVEMENTS IN A LOUDSPEAKER ARRAY

Matthieu Kuntz[1] **Bernhard U. Seeber**[1]
[1] Professorship of Audio Information Processing,
TUM School of Computation, Information and Technology,
Technical University of Munich, Germany

ABSTRACT

Free-field loudspeaker-based sound reproduction allows participants to turn their heads and move around during an experiment, without the need for individualized head-related transfer functions. In this work, we consider a 2D, square, 36-loudspeaker array. Instead of using the center of the array as the reference point for loudspeaker equalization (EQ), we applied a dynamic EQ that follows a participant's movements. We measured word recognition rates (WRR) in a S_0 $N_{diffuse}$ condition with sentences from the Oldenburg sentence test. The signal-to-noise ratio (SNR) at which 50% of words are correctly identified was determined first. This SNR is kept constant throughout the rest of the experiment. Participant WRRs were evaluated at five static positions in the array and two movement conditions, where they walked along a straight path during stimulus playback. The data show that the mean WRRs vary between .52 and .62, which roughly corresponds to a SNR change below 0.7 dB. Since this is in the order of test variance and much lower than the 3-4 dB change observed without dynamic EQ, we conclude that dynamic EQ reduces level-induced errors and can be used for speech intelligibility experiments.

Keywords: *virtual acoustic environment, dynamic loudspeaker equalization, speech intelligibility.*

1. INTRODUCTION

The use of free-field, loudspeaker-based sound reproduction is common in hearing research, as it allows the participants' anatomy to directly influence their perception without needing individualized dynamic binaural synthesis. More so, factors like head-above-torso orientation do not have to be modeled, as they happen physically. Furthermore, headphone-free sound reproduction is useful to run experiments with hearing aid or cochlear implant users [1], as well as young children [2]. Simpler experimental setups, such as speech intelligibility or detection experiments, where a (usually static) target sound is played back with the presence of interfering sounds (*e.g.* noise, one or more other speech sources) can be designed without employing sound field synthesis or panning techniques, by playing each source from a single loudspeaker [3]. However, single loudspeaker reproduction is not adapted to experiments with moving participant. As soon as the signal-to-noise ration (SNR) affects the measured quantity, which is the case for speech intelligibility and masking, position changes directly influence the measurement in an uncontrolled way.
To allow participants to move freely in the loudspeaker array without affecting the SNR, the employed sound reproduction method should be able to either accurately reproduce sound fields over a sufficiently large area, or correctly reproduce within a smaller area that can be steered to follow the listener's movements. Analytical approaches extend the wave field synthesis (WFS) technique and allow the translation of the area of correct reconstruction inside the loudspeaker array. Local WFS by spatial band-limitation [4] recomputes the loudspeaker driving functions based on a circular harmonics expansion around an arbitrary point inside the loudspeaker array. Local WFS by virtual secondary sources [5] creates a virtual loudspeaker

[*] ***Corresponding author***: *matthieu.kuntz@tum.de*

array, ideally placed around the desired position for accurate reproduction. The virtual loudspeakers are then reproduced as focused sources using WFS. Numerical approaches, such as [6, 7], invert previously measured transfer functions to optimize the reproduction accuracy in an arbitrary region.

The approach the authors investigate here and in previous work [8, 9] consists of equalizing the loudspeakers in level, time and phase at a desired point in the loudspeaker array, coinciding with the listener's position. This has the advantage of not influencing the computation of loudspeaker signals, which is often done offline before an experiment, and of intervening only on the reproduction side. Furthermore, the number of loudspeakers only depends on the number of sources to be reproduced, which is interesting for smaller scale measurement setups. Coupled with a motion tracking system, the listener's position can be used to compute the corresponding equalization (EQ) filters for every loudspeaker. This approach has been shown to work well via simulations, with a sweet-spot radius at 2 kHz of at least 22 cm in an area of 1 m × 1 m around the center of the loudspeaker array [8].

2. METHODS

The study was carried out using the simulated open field environment (SOFE) loudspeaker array of the Professorship of Audio Information Processing at the Technical University of Munich [10]. This study investigates two-dimensional sound field reproduction, using 36 horizontally arranged loudspeakers, spaced in 10° steps. The participants position was tracked using a camera-based motion tracking device (OptiTrack Prime 17W, NaturalPoint Inc. Corvallis, Oregon, USA).

2.1 Loudspeaker equalization

The EQ filters for the loudspeakers were computed in three steps. 1: All loudspeakers were measured and equalized in the center of the loudspeaker array, to ensure a linear phase, flat frequency response and equal time of arrival of the wave fronts. 2: A relative magnitude EQ was added to compensate for spectral changes due to loudspeaker directivity. It was based on the participants position relative to the 0°-axis of the loudspeaker and computed from a separate directivity measurement of the loudspeakers outside of the setup in 1° steps. 3: A gain factor and delay, based on the participants distance from the loudspeaker, compensating for distance-related time of arrival and level changes. A more detailed description of the EQ routine and its evaluation can be found in [9].

2.2 Listening experiment

2.2.1 Participants

Five participants (2F, 3M) took part in this experiment at the time of writing. All participants had normal hearing (hearing loss of 20 dB or less at the standard audiological frequencies between 125 Hz and 8 kHz, measured with pure tone audiometry) and were native German speakers.

2.2.2 Stimuli

The speech material was taken from the Oldenburger Satztest (OLSA, [11]), a closed-set matrix sentence test. All sentences contain five words (name, verb, number, adjective, noun) each pseudo-randomly picked from a set of ten words. Sentences are arranged in lists of 30, balanced to yield similar speech intelligibility across lists. Sentences were presented in front of the listener, played from a single loudspeaker, and interfering diffuse speech-shaped noise was presented from all loudspeakers, at a level of 65 dB SPL in the center of the loudspeaker array. The speech level changed depending on the measured condition. The noise started 500 ms before and ended 300 ms after each sentence.

2.2.3 Movement conditions

During the experiment, participants had to stand at different points in the loudspeaker array, or walk between two points. Figure 1 shows the two moving and the five static conditions, as well as the loudspeaker playing the speech material. The positions were labeled relative to the target loudspeaker. For the conditions *movement center* (Mov. C, the trajectory from [0m, -1m] to [0m, 1m]), *back center* (BC, [0m, -1m]) and *front center* (FC, [0m, 1m]), the speech material was presented from the loudspeaker at 0°. The lateral distance of the side trajectory was determined by considering the position of the loudspeakers, to ensure picking one that is in front of the participants. For those conditions, *movement side* (Mov. S, the trajectory from [0.82, 1m] to [0.82m, -1m]), *back side* (BS, [0.82m, 1m]) and *front side* (FS, [0.82m, -1m]), the speech material was presented from the loudspeaker at 160°. The *center* (C) indicates the center of the loudspeaker array, [0m, 0m].

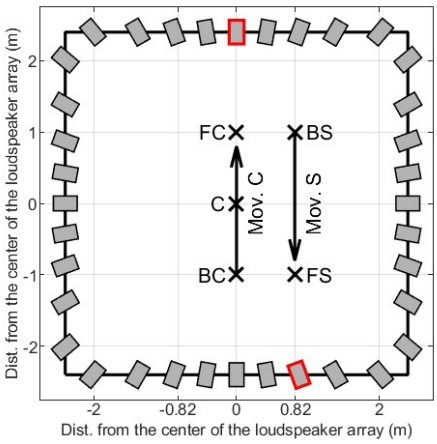

Figure 1. Sketch of the SOFE loudspeaker array and the different participant positions. The crosses indicate the positions at which participants were static, the arrows indicate the paths along which participants were walking. The loudspeakers playing the speech are represented with thicker red edges.

2.2.4 Procedure

Participants were holding a tablet PC displaying a GUI with the 5x10 OLSA word matrix. They had to select one word of each category in order to complete the trial. Speech reception thresholds (SRTs) were defined as the level of the speech material at which 50% of the words were correctly understood. The experiment was conducted in three stages: familiarization, measurement of SRTs at the array center and measurement of word recognition rates (WRR) at off-center positions. The *familiarization* run consisted of two OLSA lists presented at 65 dB SPL, for participants to get used to the experimental procedure and the speech material. They were instructed to follow the movement conditions to familiarize themselves with walking on the predefined paths in the anechoic chamber setup. The *SRT measurement* was carried out in the center of the loudspeaker array, facing the 0° loudspeaker. The level of the speech material started at 65 dB SPL and was adapted using a 1-up 1-down procedure [12]. If three or more words were identified correctly, the speech level was decreased; if two or fewer words were identified correctly, the speech level was increased. We started with a step size of 3 dB, which was decreased to 2 dB after two reversals, and decreased to 1 dB after two more reversals. The measurement was run on two OLSA lists. A psychometric function was fitted to the participants data to estimate the SRT, using the psignifit toolbox [13].

The *WRR measurement* was run with the speech material at the previously estimated SRT. Participants were presented with six OLSA lists, one for each off-center position condition. The off-center conditions were presented as a block, randomized across participants. The two movement conditions were interleaved, alternating between movement in the center and at the side, allowing participants to walk back and forth in the loudspeaker array. After the six off-center positions, participants ran an additional measurement in the center with a seventh OLSA list. This measurement allows to quantify potential changes across SRTs along the experiment. Participants had to take mandatory breaks every 60 sentences, which corresponded to blocks of about 12 minutes. The whole experiment lasted an average duration of 2.5 hours from the measurement of the hearing threshold to the end of the WRR measurement. The measured WRRs are discussed as an evaluation metric for the dynamic EQ of the loudspeakers. The more homogenous they turn out across all off-center positions, the better the participants movements are compensated by the dynamic EQ.

3. RESULTS AND DISCUSSION

Figure 2 shows the measured WRRs. The WRR is related to an equivalent change in speech level. Within participants, the equivalent level change was computed using the individual slope of the psychometric function at the threshold. The right y-axis shows the average equivalent level change across participants: a WRR change of 0.18 per dB.

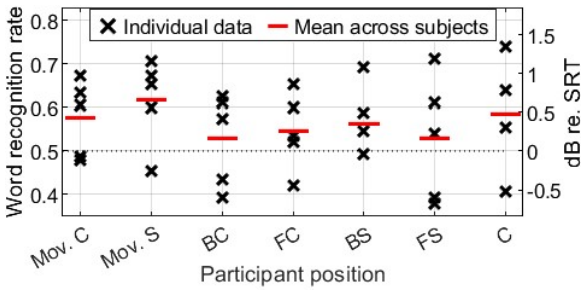

Figure 2. Measured word recognition rate for all participants across positions. Individual participant data are plotted as crosses, the mean across participants is shown as a red line. The target line of 0.5 is shown as a thin dotted line. The right axis indicates the speech level relative to the measured SRT that corresponds to the given correct word rate.

Overall, most measured WRRs lie above 0.5, indicating a slight performance increase compared to the initially measured SRT in the center. The overall mean value shows a WRR of 0.56. The equivalent level changes across all measurement positions are between -0.4 dB and 1.1 dB. For comparison, measurements of reproduction accuracy in the SOFE loudspeaker array show level changes of ± 4 dB SPL for single loudspeaker playback in a 93 cm ×93 cm area around the center [14]. The resulting level errors using dynamic EQ are much lower.

Without dynamic EQ, measurements show that the absolute SNR changes between the center and the other static conditions lie around 3-4 dB, which would quickly push participants into the ceiling performance for speech intelligibility. While such wide movements are not expected if participants are instructed to stand still it is expected to observe small variations in position across multiple trials, which can accumulate to significant changes.

A first glance at the data seems to indicate that the WRRs are higher for the movement conditions than for the static conditions. A single-sided, unequal variances *t*-test fails to find a statistically significant difference between the two groups ($p = 0.10$). Participant WRRs were on average better at the end of the test than in the beginning: 0.59 instead of 0.5, corresponding to a level change of 0.5 dB. We attribute this to a residual training effect, even after 120 sentences of familiarization and SRT measurement. Notably, one participant achieved a WRR of 0.74. Another participant had a WRR of 0.41, indicating some variability due to measurement uncertainty, potential concentration fluctuations and individual differences in training effects. The so far limited number of participants prevents assessing measurement variance.

4. CONCLUSION

This study investigated the potential of compensating sound field reproduction at off-center positions using a only simple, dynamic compensation of loudspeaker EQ for directionality and distance. Speech reception in noise was measured at static and dynamic off-center positions. Preliminary data from five subjects show that WRR deviations across positions are within 0.1. The largest observed WRR change across positions within a participant was 0.23, corresponding to an equivalent level change of 1.3 dB, which is comparable to the variance in speech intelligibility measures. These results show that dynamic loudspeaker EQ can be added to loudspeaker-based free-field reproduction to compensate for SNR changes induced by participant movements.

5. ACKNOWLEDGMENTS

This study was funded by TUM and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330 C5. The rtSOFE system was funded by BMBF grant 01GQ1004B (Bernstein Center for Computational Neuroscience Munich).

6. REFERENCES

[1] S. Kerber, B. U. Seeber: "Psychoacoustic abilities of CI users in relation to speech understanding and localization in a reverberant room," in *Fortschritte der Akustik -- DAGA '13*, (Merano, Italy), pp. 227–230, 2013.

[2] D. A. McCartney: "Development of the AnimalSeek method to evaluate the localisation ability of children under five;" Ph.D. dissertation, University of Nottingham, Nottingham, UK, 2013.

[3] S. Kerber, B. U. Seeber: " Sound localization in noise by normal-hearing listeners and cochlear implant users;" *Ear & Hearing,* vol. 33, no.4, pp. 445–457, 2012

[4] N. Hahn, F. Winter, S. Spors: "Local wave field synthesis by spatial band-limitation in the circular/spherical harmonics domain," in *Proc. of the 140th AES Convention*, (Paris, France), p. 9596, 2016.

[5] S. Spors, J. Ahrens: "Local sound field synthesis by virtual secondary sources," in *Proc. of the 40th AES International Conference*, (Tokyo, Japan), p.6–3, 2010.

[6] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, H. Hamada: "Local sound field reproduction using digital signal processing," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, 1996.

[7] M. Kolundžija, C. Faller, M. Vetterli: "Reproducing sound fields using MIMO acoustic channel inversion," *Journal of the Audio Engineering Society*, vol. 59, no. 10, pp. 721–734, 2011.

[8] M. Kuntz, B. U. Seeber: "Moving the ambisonics sweet-spot by adapting loudspeaker equalization and ambisonics decoding," in *Proc. of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, (Turin, Italy), pp. 1805–1808, 2023.

[9] M. Kuntz and B. U. Seeber: "Adapting sound reproduction to listener position with dynamic loudspeaker equalization," in *Fortschritte der Akustik -- DAGA '24*, (Hannover, Germany), pp. 1649–1651, 2024

[10] B. U. Seeber and S. W. Clapp: "Interactive simulation and freefield auralization of acoustic space with the rtSOFE," *The Journal of the Acoustical Society of America*, vol. 141, p. 3974, 2017.

[11] K. Wagener, V. Kühnel and B. Kollmeier: "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test)," *Zeitschrift für Audiologie*, vol. 38, no. 1, pp. 1–32, 1999.

[12] H. Levitt: "Transformed up-down methods in psychoacoustics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 467–477, 1971.

[13] H. Schütt: "Toolbox for bayesian psychometric function estimation," https://github.com/wichmann-lab/psignifit.

[14] M. Kuntz, N. F. Bischof, B. U. Seeber: "Sound field synthesis for psychoacoustic research: *In situ* evaluation of auralized sound pressure level" *The Journal of the Acoustical Society of America*, vol. 154, no. 3, pp. 1882–1895, 2023.