Toningenieurprojekt

Auditory perception of multiple simultaneous sources in virtual/real environments

Sonja Lutz
Matr. No.: 12336464

Electrical Engineering and Audio Engineering, Master's Program - UV 066 413
University of Music and Performing Arts Graz
Graz University of Technology

Supervisor: Univ.Prof. DI Dr.rer.nat. Franz Zotter

Graz, April 16, 2026

Erklärung

Hiermit bestätige ich, dass mir der Leitfaden für schriftliche Arbeiten an der KUG bekannt ist und ich die darin enthaltenen Bestimmungen eingehalten habe. Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Sonja Lutz, 12336464
Graz, den 16. April 2026

Unterschrift des Verfassers

Zusammenfassung

Diese Studie untersucht, wie gut die Aktivität mehrerer horizontal angeordneter Schallquellen gehört werden kann. Insbesondere wird der Einfluss offener Kopfhörer bei Schallwiedergabe über Lautsprecher sowie die binaurale Wiedergabe über Kopfhörer analysiert. Das Experiment baut auf der Mehrquellen-Studie von Stefan Riedel (DAGA 2025) auf. Es verwendet denselben Versuchsaufbau wie Riedels Studie, einschließlich der (virtuellen) Lautsprechersignale. Neu ist, dass zwei unterschiedliche Wiedergabebedingungen getestet werden: (1) Wiedergabe über Kopfhörer unter Verwendung der HRTFs des KU100-Kunstkopfes und (2) Wiedergabe über Lautsprecher, während die Teilnehmenden offene Kopfhörer tragen. Die wahrgenommene Richtungsaktivität wurde für Azimutrichtungen zwischen -105 und +105 Grad bewertet. Die Ergebnisse zeigen, dass die Teilnehmenden unter dynamischen Hörbedingungen genauere Einschätzungen der Quellenaktivität abgeben konnten.

Abstract

This study investigates how good the activity of multiple horizontal sources can be recognized. In particular, the influence of open headphones when studying physical loudspeakers, as well as binaural rendering on headphones is being investigated. The experiment carried out builds on Stefan Riedel's multi-source study (DAGA 2025). It uses the same experimental setup as Riedel's study, including the auditory stimuli. What is new is, that two distinct playback conditions are being tested: (1) sound played via headphones using the KU100 dummy head HRTFs, and (2) sound presented through physical loudspeakers while participants wear open headphones. The perceived directional activity was assessed for azimuth directions between -105 to +105 degrees. Results indicate that participants were able to give more accurate responses on source activity in dynamic listening conditions.

Contents

Chapter 1

Introduction

In recent years, the virtualization of loudspeaker environments has emerged as a powerful tool in immersive audio reproduction and auditory research. A central question, however, remains to what extent virtualized playback can perceptually approximate its real loudspeaker counterpart.

The perceptual foundations of spatial audio reproduction are well understood for stereophonic and correlated signal scenarios. Early work by Leakey [3]–[5] demonstrated how interaural time differences (ITDs) and interaural level differences (ILDs) interact in the formation of auditory events. These studies showed that controlled time and level differences between loudspeakers can produce stable and predictable phantom source localization. These principles were later further developed and systematized in the context of modern spatial audio reproduction, for example in the work of Pulkki [8], where amplitude-based panning approaches formalize the creation of directional auditory events. In such correlated playback scenarios, spatial perception is comparatively well understood and widely exploited in practical applications.

In contrast, the perceptual behavior of uncorrelated signals in multichannel loudspeaker setups is less well understood. This is particularly relevant for scenarios involving multiple simultaneously active sources, where classical stereophonic cues are no longer sufficient to fully explain perception. An important contribution in this context is the work by Santala and Pulkki [11] which investigates the localization of multiple uncorrelated noise sources. Their results indicate that listeners are capable of identifying multiple simultaneously active sources; however, the perceived spatial distribution exhibits characteristic broadening and interaction effects.

A comprehensive investigation closely related to the present study was conducted by Riedel et al. [10], who analyzed the perceived directional activity of multiple horizontally arranged loudspeakers reproducing uncorrelated noise signals under static and dynamic listening conditions. Their results show that spectral cues play a crucial role in source attribution, and that broadband as well as dynamic listening conditions significantly improve perceptual performance. At the same time, a persistent spatial ambiguity of approximately ±15° in the frontal region was observed, which could not be resolved under any tested condition.

While these findings provide important insights into real loudspeaker reproduction, it remains unclear how such perceptual characteristics translate to virtualized playback scenarios. In particular, it is an open question whether the perceptual organization of multiple uncorrelated sources is preserved under binaural rendering or under modified listening conditions.

The present study addresses this question by extending the experiment of Riedel et al. to virtualized loudspeaker reproduction. Specifically, it evaluates two conditions: (1) loudspeaker playback perceived through the acoustic covering and slight attenuation introduced by open-back headphones, and (2) binaural rendering over headphones using non-individualized head-related transfer functions (HRTFs), such as those obtained from a KU100 measurement.

The newly collected data is compared to the open-ear reference data from Riedel's original experiment. The study aims to answer the following research questions: Does a comparable perceived source activity arise for static and dynamic listening conditions — with high-pass, low-pass, and broadband noise — when

— the loudspeakers are listened to through the slight acoustic covering caused by wearing open-back headphones over the ears?

— the loudspeaker signals are rendered over headphones using non-individualized head-related transfer functions (HRTFs), e.g., based on a KU100 measurement?

Chapter 2

Methods

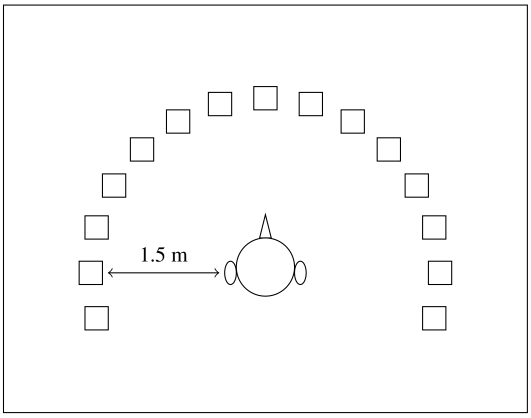


Figure 2.1 – Loudspeaker layout for the listening experiment

Figure 2.2 – Experimental setup in the acoustically damped chamber of IEM

The experiment design is comparable to the original study [10] where the identification of multiple concurrent noise sources was analyzed. Similarly to the original, the present study includes static (no head movements allowed) and dynamic listening conditions (head movements allowed) as well as three signal bandwidths: low-pass (100 Hz to 1.6 kHz), high-pass (3.2 kHz to 24 kHz) and broadband (100 Hz to 24 kHz) pink noise signals.

The listening experiment was conducted in the acoustically damped chamber of the Institute of Electronic Music and Acoustics (IEM) at University of Music and Performing Arts, Graz. Participants were seated in the middle of a circularly arranged arc consisting of 15 Genelec 8020 loudspeakers. The loudspeakers were placed in the horizontal plane from -105 to +105 degrees azimuth with a spacing of 15 degrees. The distance to the listener was 1.5 meters.

The multichannel noise signals were generated independently as uniformly distributed pink noises, which were subsequently high-pass-filtered using an 8th-order Butterworth filter with a cut-on frequency of $f_c = 100$ Hz. The low-pass-filtered condition employed a 12th-order Butterworth with a cut-off frequency of $f_c = 1.6$ kHz, and the high-pass-filtered condition employed a 12th-order Butterworth with a cut-on frequency of $f_c = 3.2$ kHz. The duration of the tested sounds was limited to 2 seconds with sine-squared fade-ins and cosine-squared fade-outs of 10 ms duration. To ensure consistent loudness across all sounds, the multichannel noise signals were divided by the square-root of the active loudspeaker count. The conditions reported in this paper used either 1, 2, or 13 active source directions.

As participants were allowed to move their head in the dynamic listening conditions, a six-degrees-of-freedom (6-DoF) rendering was required for the direct sound to account for not only rotational but also slight translational head movements. Direct-sound rendering was achieved by real-time convolution of the sound signals with diffuse-field equalized Neumann KU100 dummy-head head-related impulse responses (HRIR) [1] via IEM's HRIR Convolver Plugin [1]. Additionally, a 3-DoF 5th-order Ambisonic rendering was employed for the auralization of the room acoustics using the Ambisonic spatial decomposition method (ASDM [12]) to enhance the first-order measurement as in [9]. This approach was considered sufficiently accurate, as participants remained at the central listening position. The experimental setup is illustrated in Figures 2.1 and 2.2.

The noise bursts were either presented directly through the loudspeakers while wearing open headphones [7] or rendered binaurally and played back via these headphones. Head tracking was implemented using an OptiTrack motion capture system. For the static part of the experiment, no head movement was allowed. To ensure static conditions, constraints were implemented in PureData (Pd), so that the playback was automatically muted whenever participants turned their heads more than $\pm$ 5 degrees horizontally or $\pm$ 10 degrees vertically. The experiment was automatically resumed as soon as the head pose did not exceed the boundaries anymore. Participants used a tablet PC equipped with a graphical user interface (GUI) developed in Pd to indicate the perceived locations of the sound sources. This interface is shown in figure 2.3. To support the localization task, each potential sound source on the loudspeaker arc was labeled with its corresponding azimuth angle. These labels visually corresponded to the layout presented in the GUI, ensuring consistency between the physical setup and the user interface. The dynamic experiment was done after the static part, as it was assumed that the dynamic experiment would reveal certain loudspeaker distributions.

Twenty-one participants ($N = 21$: 4 female, 17 male) with a median age of 27 years (min. 22, max. 44) took part in the experiment. All participants were either students or staff members at IEM, each with a minimum of two years of experience in audio engineering or computer music. On average, participants required 28 minutes (min. 12, max. 52) to

1. https://git.iem.at/audioplugins/IEMPluginSuite/-/tree/HRIRConvolver/HRIRConvolver

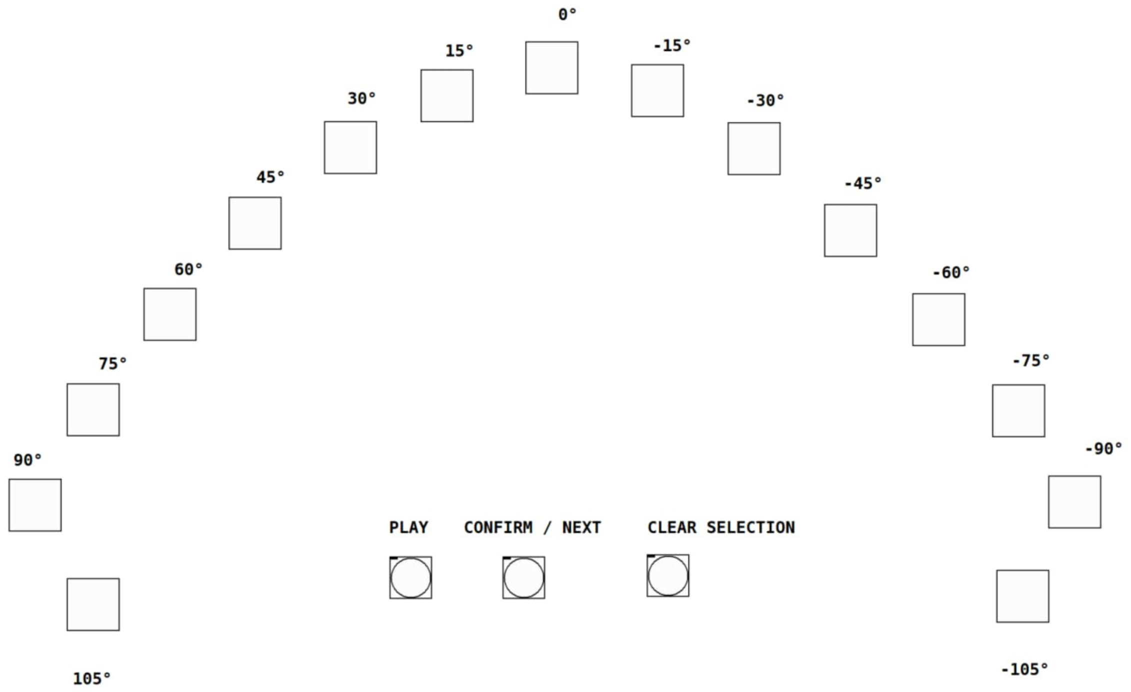


Figure 2.3 – GUI in Pd presented on a tablet PC for the experiment

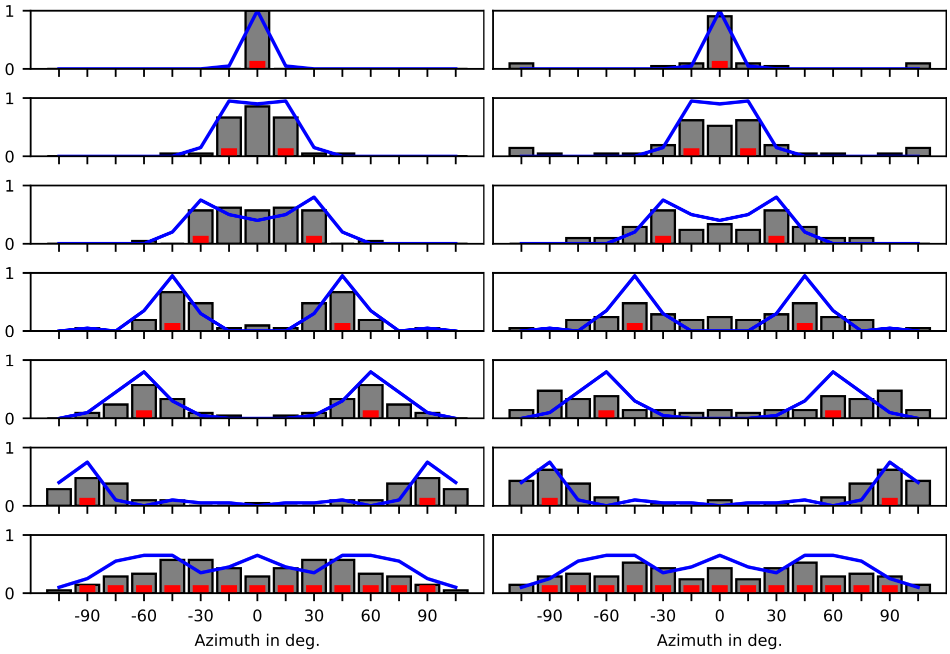
complete all 84 = 7 (source directions) x 2 (playback conditions) x 2 (static/dynamic) x 3 (type of noise) trials.

Chapter 3

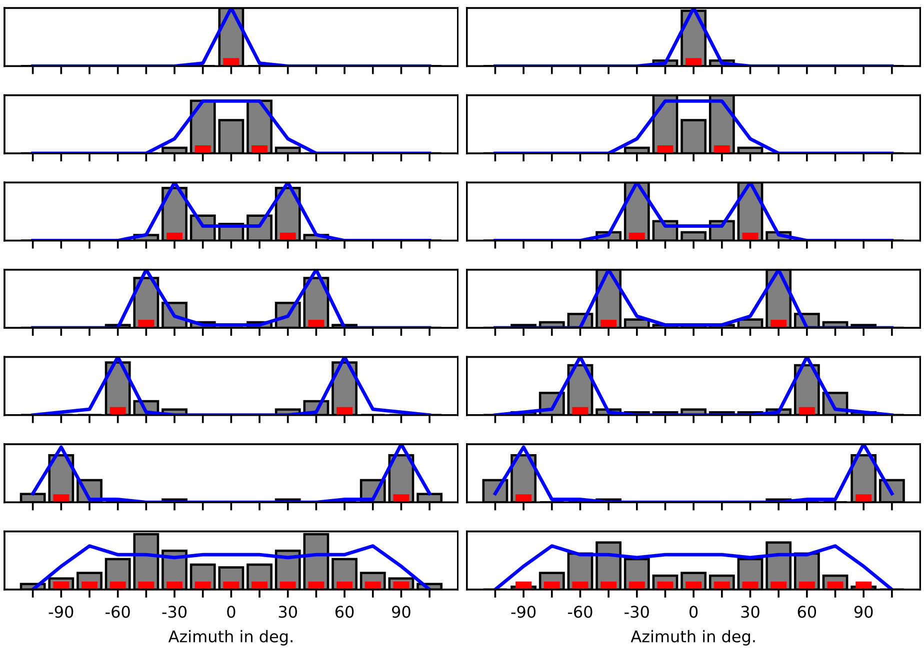
Results

Figures 3.1 to 3.3 show the results of the listening experiment in terms of response histograms (gray bars) with ground truth loudspeaker directions indicated by red squares. Blue curves show the open-ear results from Riedel's study. In this study, the term "open-ear" denotes the reference condition in which participants were not wearing headphones.

Broadband noise Figure 3.1 (broadband noises) shows that the single frontal loudspeaker is identified correctly by all listeners in (a) static and (c) dynamic listening for playback via loudspeakers while wearing open headphones. In the virtualized listening environment, most participants were able to localize the single frontal loudspeaker correctly at both static (b) and dynamic (d) conditions. The ± 15 degree conditions yielded no clear perceptual separation in localization for either condition and neither static nor dynamic listening. The conditions with two active loudspeakers at ± 30 degrees show no perceptual separation in localization for both static conditions. In the dynamic listening conditions, localization ability increases and most participants are able to localize the sources correctly but with additionally perceived sound sources near frontal azimuth directions. Starting with the condition of two active loudspeakers at ± 45, the results indicate a separate localization of the sources for the playback via loudspeakers and also for dynamic listening via headphones. The localization results for the loudspeaker setup show a noticeable bias towards frontal azimuth directions, while playback via headphones shows a noticeable bias towards lateral azimuth directions. The condition with 13 active sound sources leads to a uniform directional perception with noticeable bias towards directions between ± 30 and ± 60 degrees azimuth. Overall, dynamic listening clearly increases localization accuracy both for playback via loudspeakers and for binaurally rendered headphone signals.



(a) Real (Static) (b) Virtual (Static)

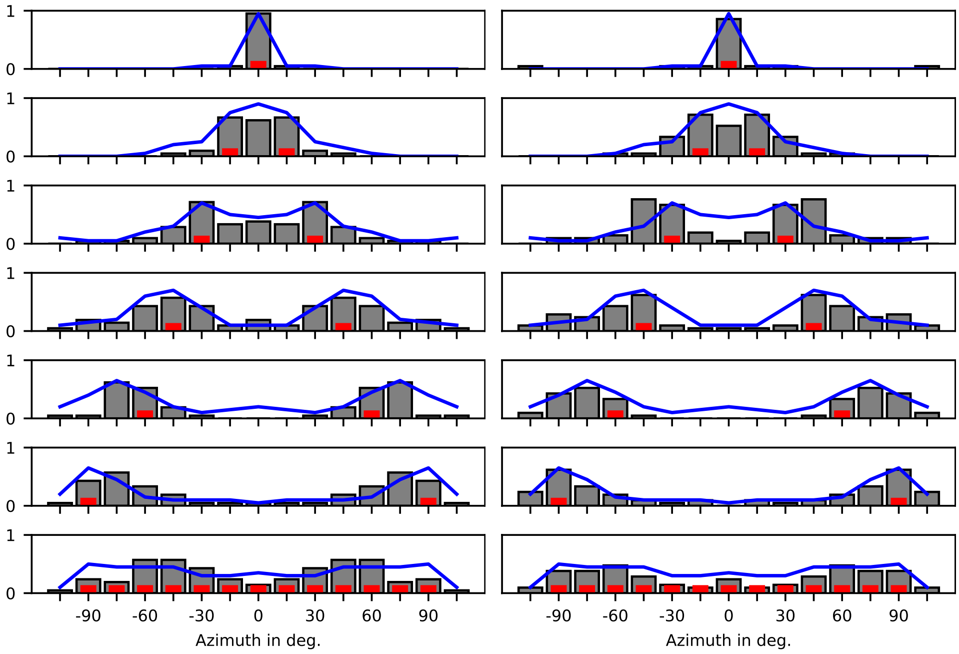


(c) Real (Dynamic) (d) Virtual (Dynamic)

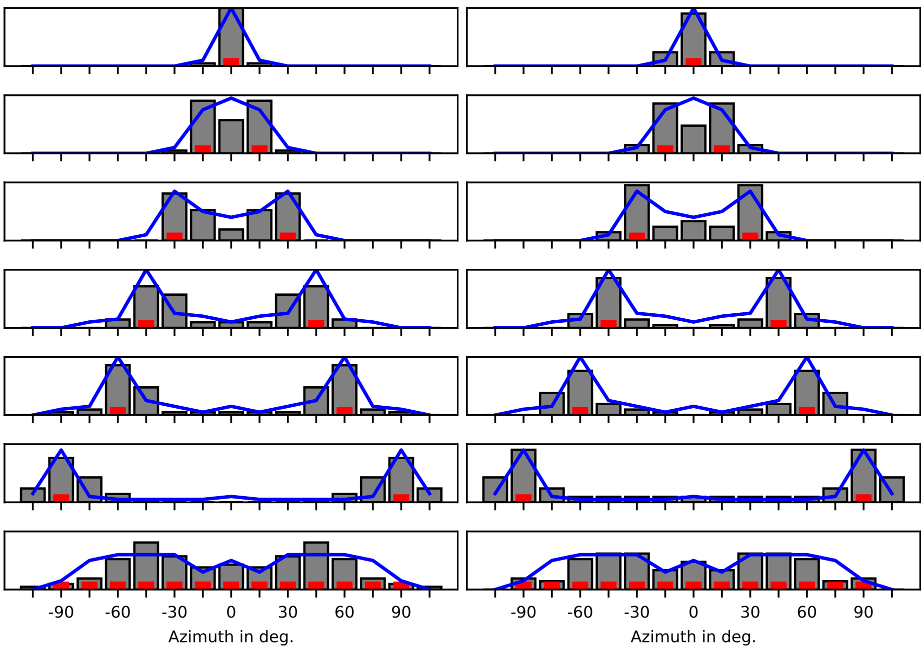
Figure 3.1 – Response histograms (N = 21) for uncorrelated broadband noise signals (100 Hz to 24 kHz). red squares indicate active loudspeaker directions. Blue curves indicate open-ear reference results from Riedel et al. [10].

Low-pass noise Figure 3.2 (low-pass noises) shows that the single frontal loudspeaker is identified correctly by all listeners in static and dynamic listening for playback via loud-speakers while wearing open headphones. In the virtualized listening environment, most participants were able to localize the single frontal loudspeaker correctly in both static and dynamic conditions. The ±15 degree conditions yielded no clear perceptual separation in localization for either condition in static listening, while dynamic listening increased perceptual source separation. The conditions with two active loudspeakers at ±30 degrees show no perceptual separation in localization for static listening and playback via loudspeakers. In the virtualized playback conditions and static listening, a noticeable spread regarding the perceived source direction, as well as a slight (incorrect) maximum for the perceived sound source at ±45 degrees can be seen. In the dynamic listening conditions, localization ability increases and most participants are able to identify the sources correctly with additionally perceived sound sources near frontal azimuth directions. Starting with the condition of two active loudspeakers at ±45, the results indicate a separate localization of the sources for all listening conditions. The results for the real loudspeaker setup show a noticeable bias towards frontal azimuth directions for dynamic listening, while playback via headphones shows a noticeable bias towards farther lateral azimuth directions. The condition with 13 active sound sources leads to a relatively uniform directional perception with noticeable bias towards directions between ±30 and ±60 degrees azimuth, except for the condition with virtual (static) listening, where the bias towards lateral azimuth directions persists, as was recorded for the condition of two active sound sources. Dynamic listening increases the number of sound sources being detected as well as the overall localization accuracy, for playback both via loudspeakers or headphones.

High-pass noise Figure 3.3 (high-pass noises) shows that the single frontal loudspeaker is identified correctly by most listeners in static and dynamic listening for playback via loudspeakers while wearing open headphone as well as for playback via headphones. The ±15 degree conditions yielded no clear perceptual separation in localization for either condition and neither static nor dynamic listening. The conditions with two active loudspeakers at ±30 degrees show no perceptual separation in localization for both static conditions. Furthermore, in both static cases the sound sources are most often perceived at ±15 degrees azimuth. In the dynamic listening conditions, localization ability increases drastically for loudspeaker playback and most participants are able to identify the sources correctly but with additionally perceived sound sources near frontal azimuth directions. In the case of playback via headphones and dynamic listening, localization ability increased slightly. Participants perceived the ±30 condition rather at ±15 degrees azimuth. Starting with the condition of two active loudspeakers at ±45 degrees azimuth, the results indicate a separate localization of the sources for most playback and listening conditions. No condition of two active sound sources at lateral directions of more than ±45 degrees azimuth yielded a localization consistent with the activated loudspeaker positions in the static listening condition. There generally appears to be a strong bias towards frontal azimuth directions in both static listening scenarios. Dynamic listening drastically improved localization accuracy for both headphone and loudspeaker playback.



(a) Real (Static) (b) Virtual (Static)



(c) Real (Dynamic) (d) Virtual (Dynamic)

Figure 3.2 – Response histograms (N = 21) for uncorrelated low-pass noise signals (100 Hz to 1.6 kHz). red squares indicate active loudspeaker directions. Blue curves indicate open-ear reference results from Riedel et al. [10].

The condition with 13 active loudspeakers leads to a strong directional perception with noticeable bias towards directions between ±15 and ±45 degrees azimuth for playback via loudspeakers in static listening and a slightly less directional perception for dynamic listening. Playback via headphones lead to a uniform directional perception in static listening and a strong directional perception at ± 45 degrees azimuth for dynamic listening. For most playback conditions, dynamic listening clearly increases localization accuracy for both playback via loudspeakers as well as for listening through headphones.

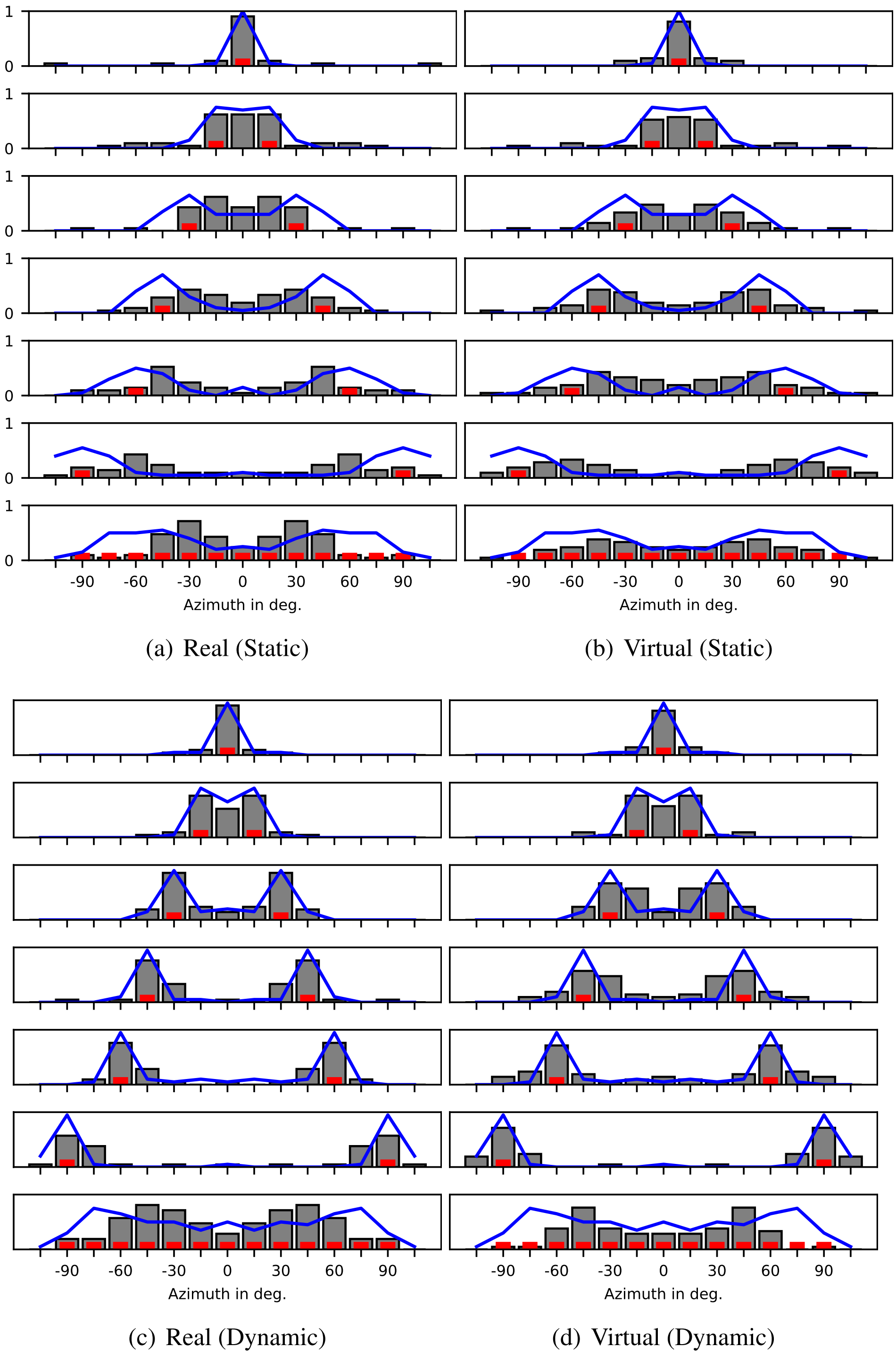


Figure 3.3 – Response histograms (N = 21) for uncorrelated high-pass noise signals (3.2 kHz to 24 kHz). red squares indicate active loudspeaker directions. Blue curves indicate open-ear reference results from Riedel et al. [10].

Chapter 4

Discussion

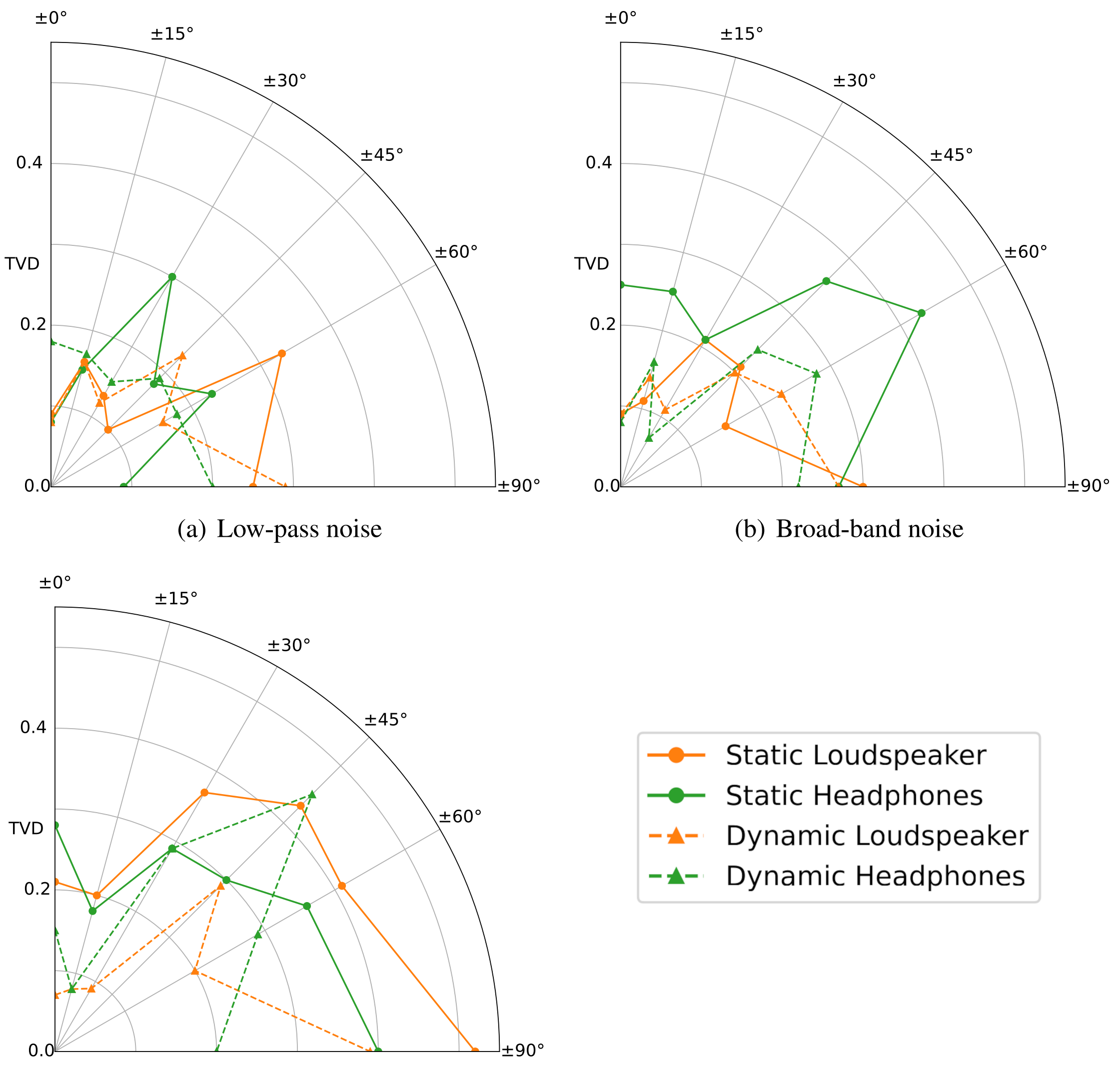
Probability mass functions (PMF) $P(\alpha)$ were calculated for all experimental conditions as well as for the open-ear reference data:

$$P(\alpha) = \frac{s(\alpha)}{\sum_\alpha s(\alpha)}\,, \tag{4.1}$$

where s_α denotes the peak-normalized values of the participants' answers depending on the azimuth angle α. To compare the similarity between the reference condition and the experiment data, the total variation distance (TVD) between their PMFs P_ref and P_exp was calcuated:

$$\text{TVD} = \frac{1}{2}\sum_\alpha \left|P_\text{exp}(\alpha) - P_\text{ref}(\alpha)\right|\,. \tag{4.2}$$

The TVD values for each listening condition are shown in Figure 4.1. For low-pass signals there is not much difference between the experiment data and the open-ear reference. For broad-band noise only static listening via headphone playback yields noticeably different results. High-pass signals show the strongest variation between the experiment data and the open-ear reference for every playback condition.



(a) Low-pass noise (b) Broad-band noise

(c) High-pass noise

Figure 4.1 – Total variation distance (TVD) between experimental results and open-ear reference data [10].

The experiment results show that for all signal bandwidths and all listening conditions, most participants were able to localize the single frontal loudspeaker correctly. Perceptual separation of two active sound sources was only possible for sound sources located at a minimum of ± 30 degrees azimuth with more accurate localization through dynamic listening for wider angles. Binaural rendering of the high-pass filtered noise signals on headphones revealed an impaired perceptual separation and lateralization for the ± 30 degrees and ± 45 degrees 'stereo setups' in both static and dynamic listening conditions, which is likely due to non-individual spectral cues (see modeling results of the previous study [10]). For low-pass filtered signals or broadband noise signals, a noticeable bias towards lateral azimuth directions could be found for playback in the virtualized environment, which can be explained by dummy heads often exaggerating the interaural time difference compared to individual listeners' natural time differences [2]. In contrast, playback via loudspeakers yielded an obvious bias towards frontal directions when wearing the open headphones in the static listening conditions, likely caused by spectral through the slight covering which the headphones introduce at the sides (±90 degrees) [6]. For the condition of 13 active loudspeakers, broadband and low-pass filtered signals lead to relatively uniform distributions in the localized directions for both binaural and loudspeaker playback as well as static and dynamic listening. The overall localization performance is similar when comparing dynamic listening to loudspeakers outside open headphones and non-individual dynamic binaural rendering. Both of these conditions contain distorted HRTF cues, which become particularly noticeable with the high-pass noise signals. Further research could include individual HRTFs or study vertical (front-back) source arrangements for a more comprehensive evaluation.

Chapter 5

Conclusion and Outlook

The present study successfully addressed how well listeners can identify which simultaneous uncorrelated-noise sources are activated in a frontal surround layout, in dynamic and static conditions, both when wearing open-back headphones to listen to loudspeakers and when binaurally rendering these sources. The study hereby extends Riedel's study that did not involve wearing open-back headphones or using binaural rendering [10]. It demonstrates that both listening through open-back headphones and rendering via non-individualized HRTFs produce broadly comparable perceptual outcomes, albeit with notable deviations. Open-back headphones diminish the detectability of lateral sources and enhance frontal mislocalizations for high-pass signals, particularly in static conditions. Non-individualized HRTFs exhibit similar, though generally milder, distortions, largely affecting lateral perception, most likely due to exaggerated short-term ITDs in some listeners.

Crucially, despite these limitations, the core patterns of directional activity remain largely intact, especially under dynamic listening. This shows that even with open-back headphones or non-individualized HRTFs, the essential features of spectrally guided source activity identified by Riedel can be faithfully reproduced. These findings not only deepen our understanding of binaural perception for multiple uncorrelated sources but also provide practical guidance for the design of virtual loudspeaker environments, highlighting how perceptually robust spatial cues can be preserved in realistic listening scenarios.

Future research could extend these findings by investigating individualized HRTFs, to determine whether they eliminate the lateral and frontal distortions observed here, and by studying vertical (front-back) source arrangements to enable a more comprehensive evaluation of spatial perception in multi-source virtual environments.

Bibliography

[1] B. Bernschütz, “A spherical far field HRIR/HRTF compilation of the Neumann KU100”, in *Fortschritte der Akustik, DAGA/AIA Merano*, Mar. 2013.

[2] P. Gutierrez-Parera, J. J. Lopez, J. M. Mora-Merchan, and D. F. Larios, “Interaural time difference individualization in HRTF by scaling through anthropometric parameters”, *Eurasip Journal On Audio, Speech, And Music Processing*, vol. 2022, no. 1, p. 9, 2022.

[3] D. M. Leakey, “Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems”, *Journal of the Acoustical Society of America*, vol. 29, no. 2, pp. 284–285, 1957.

[4] D. M. Leakey, “Binaural fusion of low- and high-frequency sounds”, *Journal of the Acoustical Society of America*, vol. 30, no. 3, p. 222, 1958.

[5] D. M. Leakey, “Some measurements on the effects of interchannel intensity and time differences in two channel sound systems”, *Journal of the Acoustical Society of America*, vol. 31, no. 7, pp. 977–986, 1959.

[6] A. Mülleder, M. Romanov, N. Myer-Kahlen, and F. Zotter, “Do-it-yourself headphones and development platform for augmented-reality audio”, in *AES 2023 Int. Conf. on Spatial and Immersive Audio, Huddersfield*, Audio Engineering Society, Aug. 2023.

[7] A. Mülleder and F. Zotter, “Ultralight circumaural open headphones”, in *Audio Engineering Society Convention 154, Espoo*, Audio Engineering Society, May 2023.

[8] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning”, *JAES*, vol. 45, no. 6, pp. 456–466, 1997.

[9] S. Riedel, M. Frank, and F. Zotter, “Effect of HRTFs and head motion on auditory-visual localization in real and virtual studio environments”, *Acta Acustica*, vol. 9, p. 21, 2025.

[10] S. Riedel, M. Frank, and F. Zotter, “Towards models for listener envelopment based on binaural multi-source localization”, in *Fortschritte der Akustik, DAS/DAGA, Copenhagen*, Mar. 2025.

[11] O. Santala and V. Pulkki, “Directional perception of distributed sound sources”, *JASA*, vol. 129, pp. 1522–1530, Mar. 2011. [Online]. Available: `https://doi.org/10.1121/1.3533727`.

[12] M. Zaunschirm, M. Frank, and F. Zotter, “Binaural rendering with measured room responses: first-order ambisonic microphone vs. dummy head”, *Applied Sciences*, vol. 10, no. 5, 2020. [Online]. Available: `10.3390/app10051631`.

List of Figures