OPEN ACCESS Freely available online

CC BY

M. Kuntz, K. Kaetel, D. Fischer, and B. U. Seeber,
"Loudness Perception of Combined Two-Way and Dipole Loudspeakers,"
J. Audio Eng. Soc., vol. 74, no. 4, pp. 188–198 (2026 Apr.).
https://doi.org/10.17743/jaes.2022.0257.

Loudness Perception of Combined Two-Way and Dipole Loudspeakers

MATTHIEU KUNTZ,[1,*] **KLAUS KAETEL,**[2]
(matthieu.kuntz@tum.de) (klaus.kaetel@molecular-acoustics.com)

DIETMAR FISCHER,[2] **AND BERNHARD U. SEEBER**[1]
(dietmar.fischer@molecular-acoustics.com) (seeber@tum.de)

[1]*Audio Information Processing, School of Computation, Information and Technology, Technical University of Munich, Munich, Germany*
[2]*Molecular Acoustics GmbH, Munich, Germany*

This study investigates the use of dipole loudspeakers in addition to typical two-way loudspeakers arranged as a classical stereo setup in a room. The dipole loudspeakers are placed on top of the two-way loudspeakers, their null oriented toward the listener to only contribute to the reverberant sound, adding strong early reflections. Loudness perception for playback over the two-way loudspeakers only and playback over both the two-way and the dipole loudspeakers were compared using a two-alternative forced-choice paradigm. Different music and noise stimuli, different relative levels between two-way and dipole loudspeakers, and two listener positions in the room were used. Results show that adding the dipole-reproduced sound can increase the perceived loudness despite keeping the sound pressure level constant. That effect can be linked to changes in the resulting spectrum, energy decay, interaural correlation, and spatial perception when adding dipole loudspeakers. A binaural loudness model based on the energy received at both ears but not considering the interaural correlation was used to predict the measured level offsets for equal loudness. The predictions were accurate for most stimuli but failed for impulsive sounds where the room reverberation was more perceptible.

Keywords: binaural hearing, dipole loudspeakers, loudness, stereophony

1 INTRODUCTION

A classical stereo setup consists of two loudspeakers, placed at equal distance of a listener position, 30° to the left and 30° to the right and oriented toward the listener, creating an equilateral triangle. A typical room impulse response (RIR) at the listener position starts with the simultaneous arrival of the direct sound, the first wavefront, emitted by each loudspeaker. The direct sound is not affected by the room and thus arrives coherently at the listeners ears. The early reflections arrive several milliseconds later, usually starting with first-order floor and ceiling reflections, followed by the back wall reflection (depending on the distances of the loudspeakers to each surface). These early reflections lower the interaural coherence (IC) but due to their sparseness do not affect it too much [1]. Early reflections were shown to be beneficial for speech intelligibility [2], and in particular, early lateral reflections contribute to the spatial impression of the room [3]. After 50 to 100 ms, reflections become detrimental to speech intelligibility due to their high number and the resulting loss of IC.

Listening to a loudspeaker stereo pair, several effects come into play, depending on the signals that are played by the individual loudspeakers. If they are coherent, the precedence effect dictates where a sound source is perceived (see [4] for a review). Depending on the time and level differences between the loudspeakers, a sound source can be perceived anywhere between the two loudspeakers or at one of the loudspeakers [5]. If time differences increase beyond 12 to 24 ms, two sound sources can be perceived, so-called split images. However, it has been shown that the precedence effect builds up, and echo thresholds increase as the auditory system adapts to a room-specific reflection pattern (see [6] for a review). If both ear signals are decorrelated slightly, resulting in a lower interaural correlation, the auditory source width (ASW) increases [7, 8] and can even increase beyond the 60° opening angle between the stereo loudspeakers. However, the auditory event is still perceived as a single source.

*To whom correspondence should be addressed, email: matthieu.kuntz@tum.de

In headphone-based reproduction, the loudness of a diotic stimulus is higher than that of its monaural equivalent, the latter one needing about 6 dB of amplification to yield an equal loudness, a phenomenon called binaural summation of loudness [9]. A decorrelation of the ear signals introduces a further increase in loudness. Edmonds and Culling [10] found that interaurally decorrelated noise was perceived as louder than its correlated diotic counterpart, "the difference being equivalent to a 2 dB increase in signal energy" [10].

The use of dipole loudspeakers was studied by Grosse and van de Par in the context of room-in-room reproduction of sound fields [11, 12]. They used a standard stereo setup of loudspeakers combined with two dipole loudspeakers in the back of the listening room specifically to excite the reverberation of the room. Their signals are based on additional microphones in the back of the recording room and optimized to reduce the influence of the listening room on the perceived sound by recreating the spectrum, the interaural cross-correlation, and the reverberation time of the recording room in the playback room. A commercial example of a similar setup is the JBL Synthesis Four system for home theaters. They recommend arranging conventional loudspeakers in a left, center, right triplet in front of the listener, and loudspeakers with a dipole directivity for the midrange and tweeter can be added to the sides and the back, especially for small listening rooms [13].

Linkwitz compared stereo reproduction using a pair of monopole versus a pair of dipole loudspeaker systems. The latter, however, was oriented to radiate toward the listener, the null along the 90° axis [14]. He described the reproduction over the dipole as giving "a much stronger impression of the recording venue space and its depth" [14]. Although he attributes that change to the precedence effect, quoting Hartmann ("Reflections add a sense of 'space filling' and loudness to the sound as a whole, but the reflections are fused with the direct sound" [15]), no further studies were run to quantify those changes.

Loudspeakers with higher-order radiation patterns have also been used for musical applications. By individually controlling multiple loudspeakers, a beamformer can be steered in various directions to create reflections on walls or on specifically installed panels in the room [16] to create different spatial impressions, thus "playing with the acoustical properties of the room" [17]. Adding loudspeakers with a dipole characteristic on top of the typical stereo loudspeaker pair, the null facing the listener, allows the controlled excitation of the room reverberation only, generating strong early reflections without a direct sound component. In empirical listening tests, sound reproduction appeared louder and more enveloping, motivating a comparison at equal reproduced sound pressure levels and leading to the fundamental hypothesis studied in this article.

This work is concerned with sound perception from dipole loudspeakers combined with regular loudspeakers. The authors are testing the hypothesis that stimuli played back over both the regular and the dipole loudspeakers are perceived louder than when played back at equal sound pressure level over the regular loudspeakers only. Through auditory modeling, light is shed on the origins of the effects on loudness. Throughout this project, the terms *two-way loudspeaker reproduction, dipole loudspeaker reproduction*, and *summed reproduction* will be used to describe which loudspeakers are active.



Fig. 1. The loudspeaker stack used in this experiment, with the dipole loudspeaker on top and the two-way loudspeaker at the bottom.

2 METHODS

2.1 Loudspeakers

The dipole loudspeaker used in this project (Molecular Acoustics Pope, Molecular Acoustics, Munich, Germany) consists of two membranes arranged back-to-back to each other and connected in opposition of phase, resulting in a joint left-right movement. This results in a dipole directivity characteristic where signals radiated toward the left and the right side show a 180 phase shift. Both dipoles were driven in phase with one another. The dipole loudspeakers are placed on top of regular two-way loudspeakers (Neumann KH 120 A, Georg Neumann GmbH, Berlin, Germany) arranged in a stereo setup, with the dipole's null in the directivity pattern oriented toward the centered listening position of the stereo arrangement. This way, the dipole loudspeakers are not interfering with the direct sound produced by the two-way loudspeakers but create additional reflections

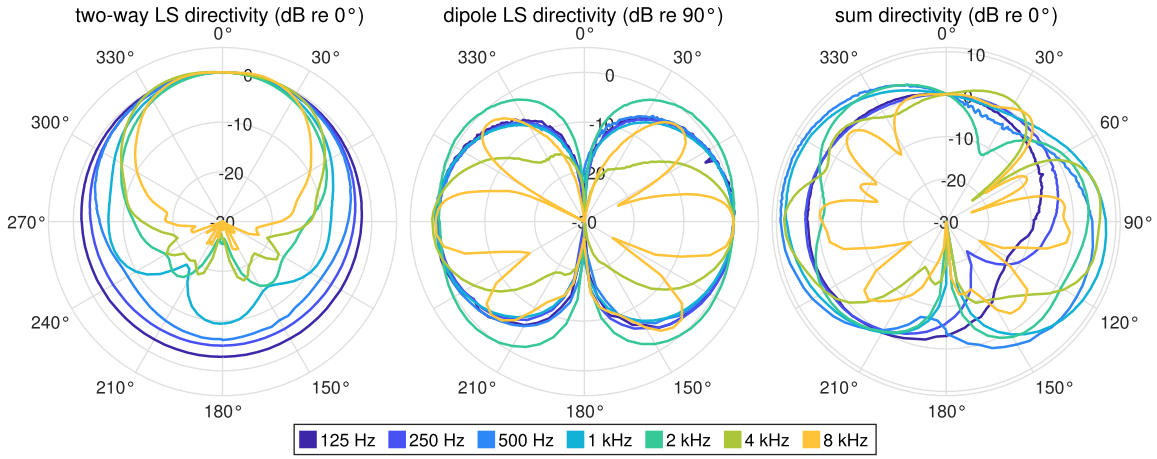


Fig. 2. Directivities of each loudspeaker when arranged in a loudspeaker stack: (a) two-way loudspeaker; (b) dipole loudspeaker; (c) sum of both loudspeakers including the +90° phase shift on the dipole. The values are given relative to the level at 0° for the two-way loudspeaker and the sum of both and at 90° for the dipole loudspeaker, their expected directivity maxima.

on the walls of the reproduction room. The arrangement of one two-way loudspeaker and one dipole as a loudspeaker stack and the combination of two identical loudspeakers in a stereo setup is referred to as a loudspeaker pair.

Signals for the dipole loudspeakers were phase shifted by +90° relative to the two-way loudspeakers, which adds extra delay to the dipole signals, particularly at lower frequencies and thus decorrelates the sum signal at low frequencies. The positive 90°-phase shift thus widens spatial perception, increases perceived envelopment, and makes the sound to be perceived livelier, which was subjectively preferred and hence applied in the experiment.

The directivity measurement of the loudspeaker stack was done in the anechoic chamber of the Professorship for Audio Information Processing at the Technical University of Munich. Impulse responses were measured in 1° steps using a maximum length sequence bandpassed between 20 Hz and 20 kHz for the two-way loudspeaker and between 100 Hz and 20 kHz for the dipole loudspeaker. A GRAS 46AF measurement microphone (GRAS Sound & Vibration, Holte, Denmark) was placed 1.5 m from the acoustic center of the loudspeaker stack, which was defined at the center of the two-way loudspeaker, at tweeter height.

Fig. 2 shows the directivity patterns of both loudspeakers. Like most loudspeakers, the two-way loudspeaker becomes more directive at higher frequencies. The figure-of-eight characteristic of the dipole loudspeaker is quite independent of frequency up to 4 kHz. At 8 kHz, sidelobes emerge in the directivity pattern, which are likely due to higher vibration modes of the dipole membranes. The directivity of the loudspeaker stack, the sum of the two-way and dipole loudspeakers, shows an asymmetry in the directivity. Frequencies below 1 kHz are radiated more toward the left-hand side, relative to the dipole's 0° axis, whereas frequencies above 2 kHz show a directivity maximum toward the right-hand side. The dipoles of both loudspeaker stacks are wired identically, with membranes moving toward the left for a positive voltage, resulting in the same directivity pattern for the left and the right loudspeaker stack.

2.2 Experimental Setup

2.2.1 Room Configuration

The experiment was conducted in a living-room-sized office (34.8 m^2) at the Technical University of Munich, with a reverberation time of about 500 ms, as depicted in Fig. 3. Two loudspeaker stacks were placed at a distance of 1.12 m from the window, spaced at 2.00 m and pointing 30° inward. Their acoustic center was placed 1.08 m above the carpeted floor to be at the ear height of a sitting listener.

Two listener positions were defined: one centered between the loudspeakers in a standard stereo setup with an opening angle of 30° and a distance of 2.00 m to each loudspeaker; the other translated by 1.00 m to the right side to be in front of the right loudspeaker stack. This way, the influence of the dipole loudspeakers could be studied inside and outside of the stereo sweet spot (i.e., in the dipole's null and in the radiation pattern of both dipoles). Participants sat on an office chair with a small headrest attached to it to suppress head movements. Because participants sat on the null of the dipole, even small head movements introduced changes in the overall spectrum, interaural level differences, and IC. This introduces noise into the loudness comparison task.

2.2.2 Room Impulse Responses

RIRs and binaural RIRs (BRIRs) were measured at both listening positions. The RIR measurement used the same hardware settings as the directivity measurements described in SEC. 2.1, the microphone being placed at both listening positions subsequently. For the BRIR measurements, the center of an HMS II.3 artificial head (HEAD Acoustics, Herzogenrath, Germany) was placed at the listening positions, its height adjusted to ensure that the ears were level with the acoustic center of the loudspeaker stacks. The

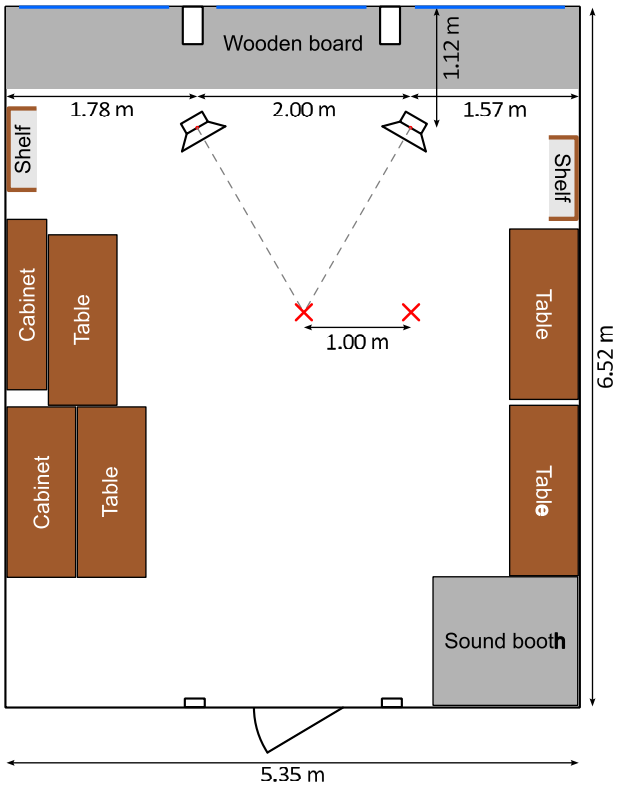


Fig. 3. Top-down view of the experiment room. Red crosses indicate the two measurement positions used for the listening experiment. The ceiling height was 3.2 m.

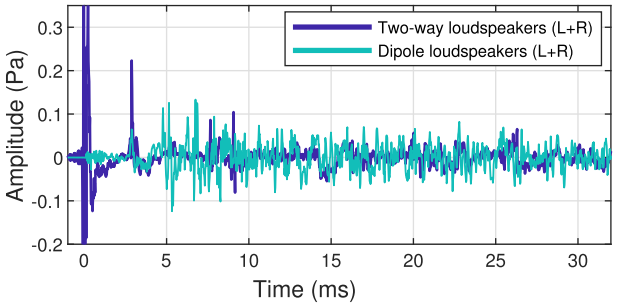


Fig. 4. RIRs for the two-way and the dipole loudspeakers measured at the center position.

in-ear microphone signals were amplified by 20 dB with a GRAS 12AA microphone preamplifier and sent into an RME Fireface UFX II audio interface for another 15 dB of gain and analog-to-digital conversion.

Fig. 4 shows the measured RIRs for the two-way and the dipole loudspeakers in the center position for a broadband level of 65 dB. The RIRs of both two-way loudspeakers and both dipole loudspeakers were summed to represent their individual contribution. The two-way loudspeakers create distinct early reflections, for example at 3, 7.5, and 9 ms. The dipole loudspeakers do not contribute to the direct sound but create dense reflections starting at around 5 ms, compared with the few individual ones coming from the two-way loudspeakers.

Table 1. Stimuli used in the listening experiment.

Denomination	Description
Drums dry	Drum stems from the introduction of the song “Spicy Funk Cake” from the German Acoustical Society (DEGA) stimulus collection [18, 19]. (0:00–0:03 s)
Drums wet	Drum introduction from the song “Well I done got over it” by Buddy Guy (0:00–0:02 s)
“Morph the Cat”	Snippet from the song “Morph the Cat” by Donald Fagan (2:15–2:18 s)
Piano	Snippet from a solo piano recording by Molecular Acoustics (0:23–0:26 s)
Pink noise	1 s of uncorrelated pink noise from 20 Hz–20 kHz, with 50-ms Gaussian envelope fade in and fade out.

2.2.3 Loudspeaker Equalization

Both loudspeakers were equalized at the center position. The measurements in the anechoic chamber showed that the two-way loudspeakers already exhibited a very flat frequency response. Equalizing the full impulse response at the listener position in the room did not lead to satisfactory results, in agreement with the limited literature on this topic (see [18] for a review), which is why the authors opted for a manual equalization: the equalization filter for the two-way loudspeakers was derived with a digital 31-band equalizer by compensating measured one-third-octave band levels for white noise at the listener position in the experiment room. This equalization was applied to compensate for the effect of midfrequency room modes in the center position on the overall spectrum at the listener position.

The equalization filter magnitude spectrum was then scaled by two-thirds, thus not fully compensating the effect of room modes, because the direct sound had a frequency-independent response. This scaling was determined empirically, resulting in a more pleasant sound. Furthermore, reducing both amplification and attenuation takes into account the stronger perceptual weight of the direct sound compared with the reverberated sound [4]. Finally, a 1,024-tap linear-phase finite impulse response filter was derived from the graphical equalizer curve, as depicted in Fig. 5.

The dipole loudspeakers were equalized separately by inverting the full RIR between 100 Hz and 20 kHz and fitting a 512-tap linear-phase finite impulse response filter to its frequency magnitude spectrum. The resulting filter was zero-padded at the beginning and end to create the same group delay as the equalization filter for the two-way loudspeakers. Fig. 5 shows the frequency response of the equalization applied to the loudspeakers and the pre-equalized and postequalized loudspeaker frequency responses.

2.3 Stimuli

The stimuli used in this experiment were chosen based on pilot experiments and are summarized in Table 1. Fig. 6 shows their frequency response in the form of third-octave band levels and their time domain signal. In particular, the drums were included for their impulsive signals. The dry

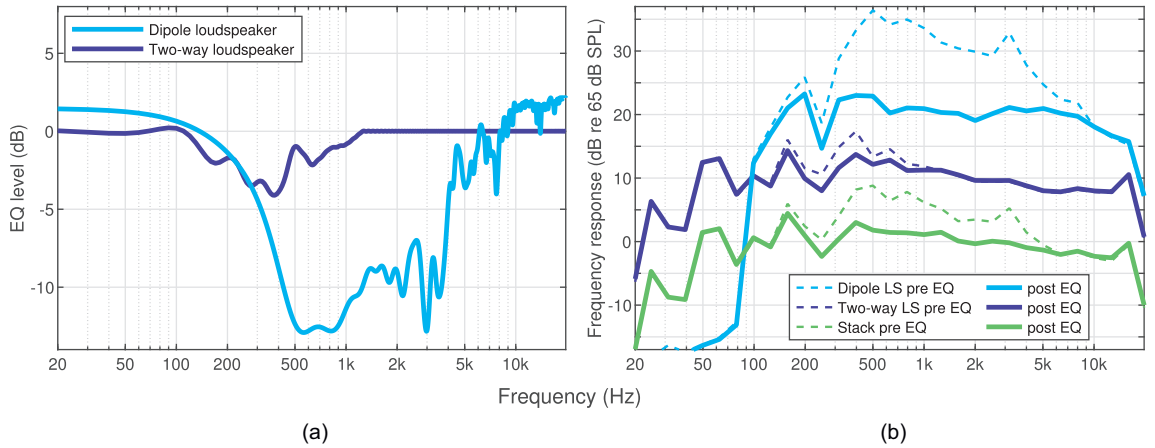


Fig. 5. (a) Frequency response of the equalization filters applied to the individual loudspeakers for the listening experiment in the room. (b) Frequency response at the center listener position in third-octave bands for the individual loudspeakers and the sum of both loudspeakers, pre-equalization and postequalization. The curves for the dipole loudspeaker and the two-way loudspeaker were shifted up in 10 dB increments for better readability.

drums came from a close-mic recording and were not processed with any spatial effects, resulting in a minimally reverberated stimulus [19]. Including dry drums in the experiment illustrates the case where the reverb perceived by the listener is primarily due to the listening room and the balance between two-way and the dipole loudspeakers. "Morph the Cat" and the piano piece were included to cover musical pieces, one with and one without voice. The two two-way loudspeakers received the stereo stimuli. The signals driving the dipole loudspeakers were derived from the signals driving the two-way loudspeakers by applying a high-pass filter at 100 Hz and a +90° phase shift using the Voxengo PHA-979 plug-in to decorrelate the loudspeakers within a stack from one another.

2.4 Procedure

A two-interval two-alternative forced-choice paradigm was used in which the participant indicated which interval was perceived as louder, without considering timbral and spatial changes between the intervals. One interval consisted of the two-way loudspeaker reproduction, the other consisted of the summed reproduction. The experiment was designed with three fixed factors: the mixing ratio between the two-way and dipole loudspeakers in the summed reproduction interval, the stimulus, and the subject's position in the room. To determine the level offset for which both intervals sound equally loud, a constant stimulus design was used, where the level difference between both intervals was varied.

For each trial, a stimulus different from the previous one was picked randomly. For each condition, half of the repetitions were run with the two-way loudspeaker reproduction as the first interval and the summed reproduction as the second, the other half were run with the inverse order. To generate the summed reproduction interval, the dipoles were added to the two-way loudspeakers with a relative level of −12, −9, −6, −3, or 0 dB. Thus, across both intervals, the two-way loudspeaker reproduction and the summed reproduction were directly compared with different mixing ratios for the contribution of the dipole loudspeakers. The intervals were then each normalized to a sum level of 65 dB SPL and roved by applying a level offset randomly picked from a uniform distribution between −2 and 2 dB to both intervals to suppress reference to an across-trial loudness standard. The level difference between the summed reproduction and the two-way loudspeaker reproduction interval was introduced by subtracting half the level difference from the two-way loudspeaker interval and adding half the level difference to the summed reproduction interval. The level difference was varied in five 1 dB steps, distributed around

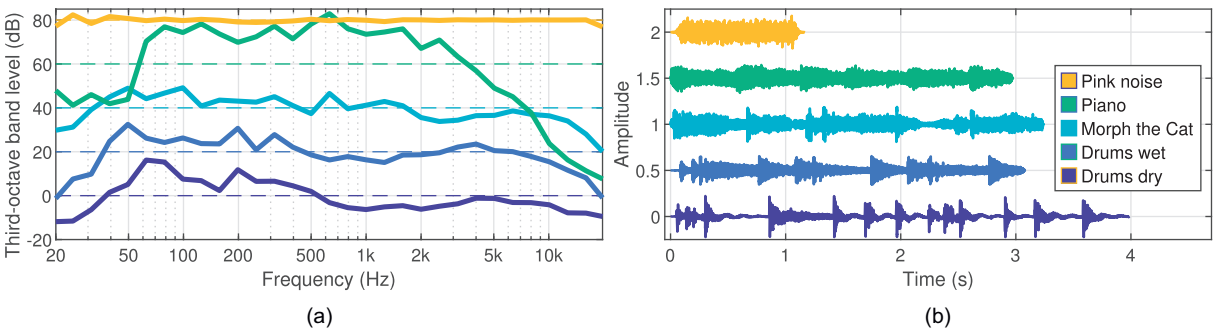


Fig. 6. Third-octave band levels (a) and time signal (b) of the stimuli. The third-octave band levels were shifted up in increments of 20 dB; the time domain signal amplitudes were shifted up in increments of 0.5 dB for better readability.

a value previously estimated in a pilot experiment. In the center, the tested level differences varied from −2 to 2 dB for every stimulus.

On the side, the pilot experiment showed larger level offsets for equal loudness, so the range of measured level differences had to be adapted. A range of −4 to 0 dB was used for both drum stimuli, −3 to 1 dB for "Morph the Cat" and pink noise, and −2 to 2 dB for the piano piece.

The experiment results are given as the level offset between the two-way loudspeaker reproduction and the summed reproduction at which both were perceived as equally loud. For each subject, for each individual condition and each level difference, the authors computed the proportion of repetitions where the summed reproduction interval was chosen as the louder one and fitted a psychometric function through the data points. The 50% point on that curve indicated the level offset for equal loudness. In rare cases, the 50% point of the fitted curve lay outside of the measured range. These conditions were inspected to ensure a reliable fit was obtained. A negative level offset indicates that the summed reproduction is perceived as louder than the two-way loudspeaker reproduction and that its level can be reduced by that value to match the loudness of the two-way loudspeaker reproduction.

A total of 2,000 trials (five level differences × five mixing ratios × five stimuli × two positions × eight repetitions) were presented in random order and split into 24 runs of around 12 minutes each. Participants took a short break after each run. Before the experiment, participants carried out one training run of 84 trials to familiarize themselves with the experiment routine. The 24 runs were split over 2 to 4 separate days. The experiment lasted an average of 5 h and 25 minutes, with subjects taking an average of 7.5 h to complete it, including the training run and breaks.

2.5 Participants

Eleven participants (9 male, 2 female) with a mean age of 26.5 years took part in the experiment. All participants had less than 20 dB hearing loss at 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz, as measured with a Madsen Astera 2 audiometer (Natus, Middleton, WI, USA). One subject had to be excluded from the analysis because their responses were not suited to fit a psychometric function. The participant selected the first interval in 80% of the trials, resulting in most of the data points located around chance performance of 50%. The following analysis is based on the data from the ten remaining participants.

2.6 Binaural Loudness Model

The stimuli used in this experiment were also analyzed using the binaural loudness model described by Moore et al. [20], on which the ISO 532-2:2017(E) norm for binaural acoustic loudness is based [21], and implemented in the Auditory Modeling Toolbox version 1.2 [22]. Specific loudness is calculated for each ear and critical band separately from the ear signals. To obtain binaural loudness, specific loudness contributions from both ears are summed, yielding a maximum binaural loudness gain of two. However, if specific loudness at both ears differs, the effectiveness of that summation is reduced by a binaural inhibition factor reducing the specific loudness contribution of the softer ear such that the model defaults to the monaural loudness of the dominant ear in case of strong specific loudness asymmetries. The model does not consider interaural phase and interaural correlation.

To estimate the level offset needed for loudness equality with the model, the loudspeaker signals were computed as explained in SECS. 2.3 and 2.4, without the random level roving. They were then convolved with the BRIRs measured at both listener positions to simulate the ear signals of the participants. For every stimulus and mixing ratio, the level offset for equal loudness was estimated by computing the loudness of the two-way loudspeaker reproduction ear signals and then iteratively changing the level of the summed reproduction ear signals until the estimated loudness matched the two-way loudspeaker reproduction loudness within 0.1 sone. Although the sound pressure levels are slightly different than those during the experiment (due to random level roving), the differences are small enough that the change in upward spread of masking is negligeable, even more for broadband sounds.

3 RESULTS

Fig. 7 shows the measured individual and average level offsets for each condition. A first look at the data shows that most level offsets are negative. A one-sided t test on the pooled results across all stimuli, positions, and mixing ratios verifies that the measured level offsets are significantly below 0 dB ($p < 0.01$) and thus rejects the null hypothesis that the summed reproduction is as loud as the two-way loudspeaker reproduction. However, there are clear differences between the stimuli; for example, the level offsets for the piano piece at the center position are slightly positive, as opposed to the dry drums. The following sections discuss the individual factors.

3.1 Effect of the Stimulus

Fig. 8(a) shows the mean and the SE of the individual level offsets for the different stimuli pooled across all mixing ratios. Both drum tracks showed the largest level offset values, indicating the largest effects of the added dipole loudspeakers on perceived loudness, with −1.0 dB for the reverberated drums and −1.5 dB for the dry drums. Level offsets for "Morph the Cat" and pink noise were smaller, with values of −0.8 and −0.6 dB, respectively. The piano piece was the only stimulus to be perceived as louder in the two-way loudspeaker reproduction case, its level offset reaching 0.2 dB. A first group of double-sided, Bonferroni-corrected t tests was run to investigate significant differences from 0 dB. All level offsets were significantly different from 0 dB ($p < 0.01$). A second set of double-sided, Bonferroni-corrected t tests compares the groups among themselves, finding significant differences between stimuli for all pairs except between "Morph the Cat" and pink noise.

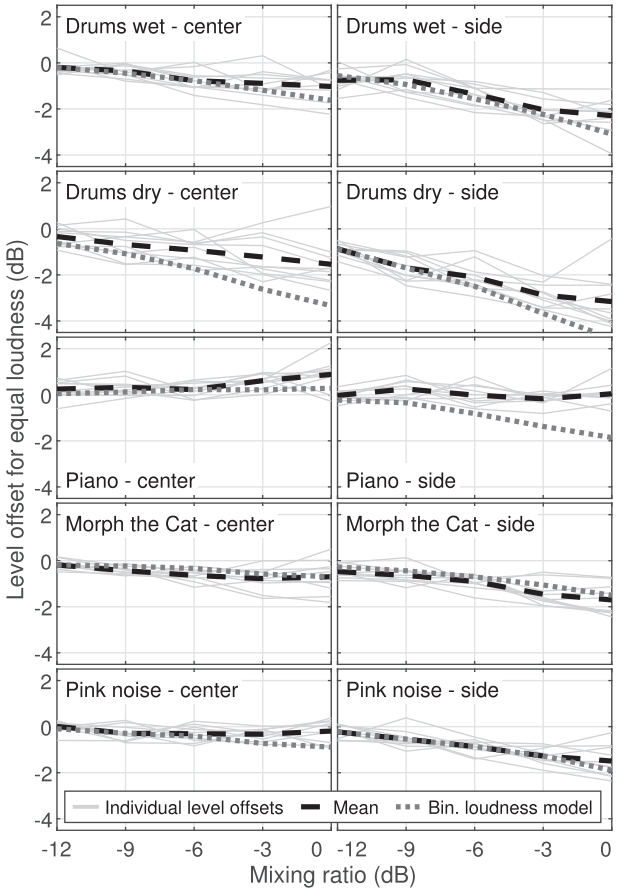


Fig. 7. Results of the listening experiment. Each panel shows the measured level offsets of each participant (gray lines) and their mean value (dashed line) across mixing ratios. The dotted line indicates the level offsets estimated with the binaural loudness model. Data collected at the center position are shown on the left side, and data from the side position are on the right side. Each row corresponds to a different stimulus. Negative level offsets indicate the summed reproduction to be perceived louder than the two-way loudspeaker despite identical sound pressure levels.

3.2 Effect of the Mixing Ratio

Fig. 8(b) shows the mean and the SE of the individual level offset values for the different mixing ratios between the two-way and the dipole loudspeakers. The downward trend (larger perceived loudness) observed in the raw data is highlighted here; the level offsets increase in absolute value from −0.3 dB for a mixing ratio of −12 to −1.1 dB for a mixing ratio of 0 dB. All conditions were significantly lower than 0 dB, so even the addition of the dipoles with a −12 dB mixing ratio introduced a perceivable and statistically significant loudness difference compared with the two-way loudspeakers only. Running the same statistical tests as before, it was found that all stimuli were significantly different from 0 dB. Between mixing ratios, no difference could be observed between −12 and −9 dB and between −3 and 0 dB, indicating that the evoked effect can best be tuned for a relative level of the dipole loudspeakers between −9 and −3 dB.

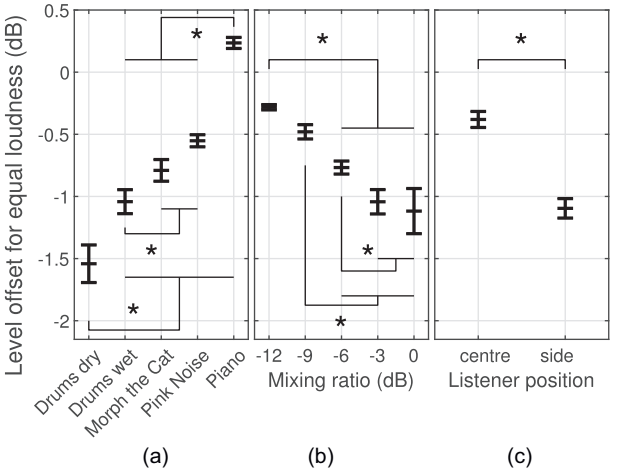


Fig. 8. Mean and SE of the measured level offset values pooled to express the different factors of the experiment. (a) Level offsets are grouped by stimuli. (b) Level offsets are grouped by mixing ratio. (c) Level offsets are grouped by listener position. Significant differences between groups ($p < 0.01$) are indicated by an asterisk. Additionally, all groups are significantly different from 0 dB ($p < 0.01$).

3.3 Effect of the Listener Position

Fig. 8(c) shows the mean and the SE of the individual level offsets for the two listener positions pooled across stimuli and mixing ratios. Listeners reported a stronger loudness increase when sitting on the side compared with the center, level offsets going from −0.4 to −1.1 dB, respectively. The same statistical tests as before show both level offsets are significantly different from 0 dB and significantly different from each other.

3.4 Loudness Model

The level offsets for equal loudness estimated with the binaural loudness model are represented as dotted lines in Fig. 7. Overall, they are fairly accurate for low mixing ratios, with deviations from the experimental data below 0.3 dB for a mixing ratio of −12 dB. As the mixing ratio increases, the predictions of the binaural loudness model deviate more, but the deviation is still below 1 dB for a mixing ratio of 0 dB in most cases. The dry drums at the center and at the side and the piano piece at the side show the biggest deviation: −1.8 dB, −1.6 dB, and −1.8 dB, respectively. For these impulsive sounds with room reverberation, the binaural loudness model overestimated the perceived loudness, resulting in a lower level offset than measured experimentally. Because the model only considers energetic effects including monaural masking, the deviations suggest contributions from binaural processes.

4 DISCUSSION

The experimental results indicate that the playback combining dipole and two-way loudspeakers can be perceived as louder than the playback via two-way loudspeakers only, at equal sound pressure level. The following sections discuss the changes occurring in the spectrum, energy decay,

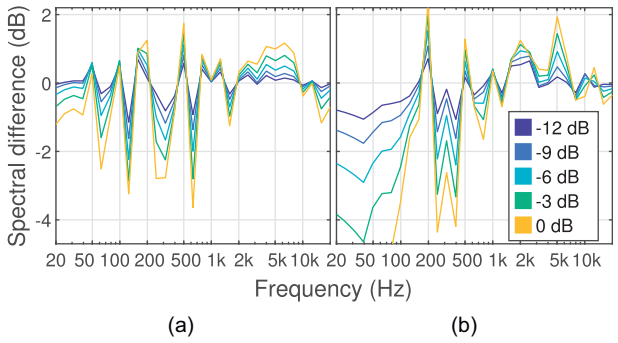


Fig. 9. Spectral difference in third-octave bands, defined as the frequency response in dB of the summed reproduction RIRs minus the frequency response in dB of the two-way loudspeaker reproduction RIRs. Each panel corresponds to a different listener position: (a) center, (b) side.

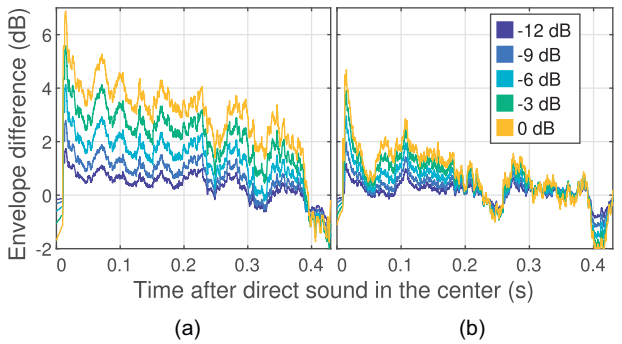


Fig. 10. Decay envelope difference, defined as the envelope of the RIRs in decibels of the summed reproduction minus the envelope of the RIRs in decibels of the the two-way loudspeaker reproduction for different mixing ratios; each panel corresponds to a different listener position: (a) center, (b) side.

and spatial perception when dipole loudspeakers are added to explain the effects observed in the experiment and the deviations observed from the model.

4.1 Spectral Differences

An increase in the relative level of the dipole loudspeakers also introduced a timbral change in the stimuli: they became brighter and thinner. Fig. 9 shows the spectral differences between the sum of both loudspeakers and the two-way loudspeakers in third-octave bands. The higher the mixing ratio was, the higher the deviations were from the two-way loudspeaker reproduction. Most notably, the summed reproduction contains more energy in the frequency band between 2 and 8 kHz, covering eight critical bands (BARK) and thus contributing most to the perceived loudness [23]. Here, the differences are more pronounced at the side position, indicating a strong contribution of the overall spectrum to the increased loudness perception.

In the case of the piano piece, which showed no or even an inverted loudness effect, that timbre change subjectively resulted in a less deep and impressive sounding piano. Because the dipole loudspeakers were high-passed at 100 Hz, increasing their relative contribution effectively high-passes the summed reproduction stimulus as the contribution of the two-way loudspeaker is reduced to keep the overall level constant. However, the lower energy for frequencies above 2 kHz in the piano signal reduces the effect the added dipoles can produce compared with the other stimuli.

The binaural loudness model employs a purely energetic approach, and its predictions show that the spectral changes account for most of the loudness changes. However, the deviations, especially for impulsive sounds with room reverberation, indicate that there must be other factors at play.

4.2 Energy Decay

Fig. 10 shows the difference between the decay of the summed reproduction RIR in decibels and the decay of the two-way loudspeaker RIRs in decibels (summed reproduction minus two-way loudspeaker). It was computed from the measured RIRs with a 20-ms sliding root-mean-square window. As shown in the Fig. 10(a), a higher mixing ratio resulted in more energy of early reflections after the direct sound, reaching up to 5 dB for a mixing ratio of 0 dB. The energy of the direct sound impulse tends to be smaller with increasing mixing ratio, as is visible in the starting points of the decay difference curves. This is due to the level reduction on the two-way loudspeakers as the dipole loudspeakers are added to keep the sound pressure level constant.

When comparing the two drum tracks with the other stimuli, it seems that impulsive excitation of listening room reverberation, particularly for the dry drums, leads to a higher contribution of the added dipole loudspeakers to the perceived loudness. For both of these stimuli, the level of the summed reproduction can be reduced by more than 1 dB to yield equal loudness with the two-way loudspeaker reproduction. These stimuli contain strong transients and short pauses between them, where the reverberation of the room and the higher energy in the decaying impulse response of the summed reproduction can be heard and thus has a stronger effect on loudness perception.

On the other hand, the piano piece is perceived as slightly quieter when adding the dipole loudspeakers. This stimulus has a low dynamic range, leaving less room between the transients for the reverberation to be heard in. Different to the drum recordings, much of the added reflected energy would be masked by the direct sound and its reverberation. Because the recording is already quite reverberant and ear signals are decorrelated, the additional decorrelation of the signals through the addition of the dipole loudspeakers is small and should have a limited effect on loudness. “Morph the Cat” can be considered as a mix of strong transients, provided by the drums in the song and wider elements like the background vocals, which could explain a smaller level offset from 0 dB than is observed for the drum stimuli.

However, the pink noise stimulus had the lowest dynamic range of all but still showed a slight increase in loudness perception when adding the dipole loudspeakers. Furthermore, the decay difference was higher in the center than at the side, whereas the level offsets for equal loudness were greater at the side.

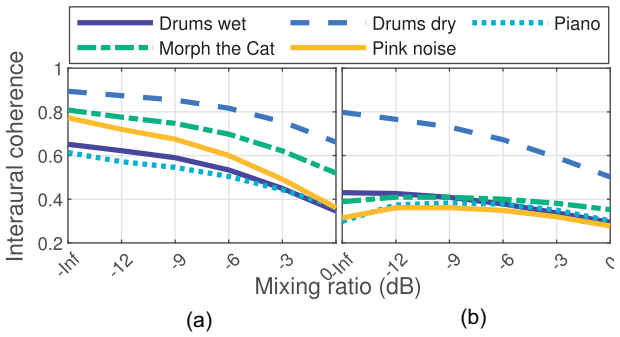


Fig. 11. IC for the different stimuli in the center (a) and at the side (b). A mixing ratio of negative infinity (−Inf) indicates the two-way loudspeaker reproduction.

Loudness increases when early reflections are fused with the direct sound [24, 25]. The temporal integration of individual reflections with the direct sound or with the late reverberation has been shown to depend on the interaural phase difference [2] among other cues for auditory scene analysis. The binaural loudness model does not account for fusion and separation of reflections with the direct sound nor for binaural processes related to binaural unmasking based on interaural phase differences, which play a crucial role in the audibility and thus loudness of partially masked sound sources [26]. In return, this is an indication that those binaural processes are at play here and influence loudness perception [27].

4.3 Interaural Coherence and Spatial Characteristics

In addition to the different decay and timbral changes, blending in the dipole loudspeakers also changes the sound's spatial characteristics. Indeed, adding more early reflections increases diffuseness and decreases the IC, leading to a greater apparent source width [8], increased loudness [10], and a more enveloping sound [28]. Some participants also reported these sensations.

Fig. 11 shows the IC, computed using the recorded BRIRs convolved with the same stimuli that were used with the loudness model, by taking the maximum absolute value of the interaural correlation function on the evaluation interval of $\pm$ 1 ms. Fig. 11(a) shows the IC at the center. For the two-way reproduction, the coherence varies between 0.89 for the dry drums and 0.61 for the piano. The low coherence already observed without the dipole loudspeakers in the piano and wet drums recording comes from the high amount of reverberation (and hence decorrelation) in the stimuli. The drop in IC between the two-way loudspeaker reproduction and the summed reproduction with a mixing ratio of 0 dB is about 0.25 for all stimuli and 0.4 for pink noise.

However, changes in IC are perceived as stronger for high ICs, meaning the drop in IC was perceived as more significant for the dry drums or "Morph the Cat" than the piano stimulus. At the side, the dry drums still show an IC of 0.80 for the two-way reproduction, the other stimuli yielding values between 0.45 and 0.32. Here, the drop in IC between the two-way loudspeaker reproduction and the summed reproduction with a mixing ratio of 0 dB is largest for the dry drums, with 0.30, followed by 0.14 for the wet drums. The other stimuli showed only minimal changes in IC.

At the side position, the effects of the additional dipole loudspeakers were more pronounced than in the center of the stereo setup, which is contrary to the change in IC just observed. A possible explanation is the smaller distance to the nearest wall, enhancing the level of the reflections from that side and the contribution of the dipole loudspeaker on the right side. On top of that, listeners at the side position were not sitting on the null of the dipole characteristic of the closest loudspeaker, meaning it also contributed to the direct sound. The result was that the auditory image of the summed reproduction extended further upward than compared with the two-way loudspeaker reproduction, a change also mentioned by one participant. Finally, in some cases of the two-way loudspeaker reproduction interval, the right loudspeaker was perceived as the only source in the room due to the precedence effect. This may have led to a narrower source width in the two-way loudspeaker reproduction, which could have increased the contrast to the summed reproduction more than in the center.

5 CONCLUSION

In a listening experiment with ten subjects in a living-room-sized office, loudness was perceived higher with dipole loudspeaker reproduction added to two-way loudspeaker reproduction in a stereo setup than with the two-way loudspeakers alone, despite keeping the sound pressure level constant. The loudness increase depended on the characteristics of the input signal. Songs featuring strong transients and more focused mixes benefited more from the addition of dipole loudspeakers. The stimuli containing drums showed equivalent loudness increases of 0.7 to 1.5 dB, whereas the effect on loudness was negligible for pink noise and a piano stimulus. The piano stimulus was perceived slightly quieter with the added dipoles, with a level offset for equal loudness of 0.2 dB. Increasing the relative level of the dipole loudspeakers increased the loudness difference.

Informal listening and participant feedback showed that increasing the level of the dipoles broadens the acoustic scene and creates a more enveloping listening experience but can lead to a thinner sound lacking low-end when approaching the 0 dB mixing ratio. For the side position, loudness was higher on average by an equivalent of 0.5 dB compared with the center position, where the dipoles did not contribute direct sound. Because loudness is increased with added dipole reproduction, the reproduced sound pressure level can be lowered by up to 1.5 dB compared with standard two-way loudspeakers. Although this was not investigated further, the use of dipole speakers as an addition to standard stereo pairs in home hi-fi setups or for sound reinforcement in closed concert venues is possible.

Loudness effects were well predicted by the loudness model, indicating that spectro-temporal changes in the ear

signals explain a large part of the effect. However, model deviations of up to 2 dB for sounds with strong transients indicate that changes in interaural phase differences and IC due to different excitation of the room reverberation also play a role, likely related to the mechanism of binaural unmasking that are not captured by the model.

Future work could focus on the signal characteristics that facilitate the loudness effect and also on how audio engineers could use dipoles to purposefully change spatial characteristics during sound reproduction. Furthermore, the integration of binaural processes based on interaural correlation and phase into binaural loudness models could be the scope of following studies.

6 ACKNOWLEDGMENT

We thank the two reviewers for their insightful comments throughout the revision process. This study was funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (Innovationsgutschein 2021-5664-MO-03_SK) and Molecular Acoustics GmbH.

7 REFERENCES

[1] F. Menzer and C. Faller, "Investigations on an Early-Reflection-Free Model for BRIRs," *J. Audio Eng. Soc.*, vol 58, no. 9, pp. 709–723 (2010 Sep.). https://aes.org/publications/elibrary-page/?id=15517.

[2] J. Rennies, A. Warzybok, T. Brand, and B. Kollmeier, "Measurement and Prediction of Binaural-Temporal Integration of Speech Reflections," *Trends Hear.*, vol 23, p. 233121651985426 (2019 Jun.). doi:https://doi.org/10.1177/2331216519854267.

[3] M. Barron, "The Subjective Effects of First Reflections in Concert Halls—The Need for Lateral Reflections," *J. Sound Vib.*, vol 15, no. 4, pp. 475–494 (1971 Apr.). https://doi.org/10.1016/0022-460X(71)90406-8.

[4] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The Precedence Effect," *J. Acoust. Soc. Am.*, vol. 160, no. 4, pp. 1633–1654 (1999 Oct.). https://doi.org/10.1121/1.427914.

[5] N. V. Franssen, *Stereophony* (Philips Technical Bibliography, Eindhoven, The Netherlands, 1963).

[6] B. U. Seeber and S.W. Clapp, "Auditory Room Learning and Adaptation to Sound Reflections," in J. Blauert and J. Braasch (Eds.), *The Technology of Binaural Understanding*, pp. 203–222 (Springer Nature Switzerland, Cham, Switzerland, 2020). https://doi.org/10.1007/978-3-030-00386-9_8.

[7] K. Ueda and M. Morimoto, "Estimation of Auditory Source Width (ASW): I. ASW for Two Adjacent 1/3 Octave Band Noises with Equal Band Level," *J. Acoust. Soc. Jpn. (E)*, vol. 16, no. 2, pp. 77–83 (1995 Feb.). https://doi.org/10.1250/ast.16.77.

[8] W. M. Whitmer, B. U. Seeber, and M. A. Akeroyd, "The Perception of Apparent Auditory Source Width in Hearing-Impaired Adults," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. 3548–3559 (2014 Jun.). https://doi.org/10.1121/1.4875575.

[9] G. S. Reynolds and S. S. Stevens, "Binaural Summation of Loudness," *J. Acoust. Soc. Am.,* vol. 32, no. 10, pp. 1337–1344 (1960 Oct.). https://doi.org/10.1121/1.1907903.

[10] B. A. Edmons and J. F. Culling, "Interaural Correlation and the Binaural Summation of Loudness," *J. Acoust. Soc. Am.*, vol. 125, no. 6, pp. 3865–3870 (2009 Jun.). https://doi.org/10.1121/1.3120412.

[11] J. Grosse and S. van de Par, "Evaluation of a Perceptually Optimized Room-in-Room Reproduction Method for Playback Room Compensation", in Proc. *EuroNoise Conf.*, pp. 2071-2076 (Maastricht, Netherlands) (2015 May/Jun.).

[12] J. Grosse and S. van de Par, "Perceptually Accurate Reproduction of Recorded Sound Fields in a Reverberant Room Using Spatially Distributed Loudspeakers," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 5, pp. 867–880 (2025 Aug.). https://doi.org/10.1109/JSTSP.2015.2402631.

[13] JBL, *Synthesis Four S4VC, S4HC & S4Ai Flush-Mount Loudspeakers: Owner's and Installer's Manual*. (Harman Consumer Group, Inc., Woodbury, NY, 2008).

[14] S. Linkwitz, "Room Reflections Misunderstood?," presented at the *123rd Conv. Audio Engineering Society* (2007 Oct.). https://www.linkwitzlab.com/AES123-final2.pdf.

[15] W. M. Hartmann, "Listening in a Room and the Precedence Effect," in R. Gilkey and T. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments,* pp. 191–220 (Psychology Press, New York, NY, USA, 1997). https://doi.org/10.4324/9781315806341.

[16] GmbH sonible, "IKO," https://iko.sonible.com/ (accessed 16 Jul. 2024).

[17] N. Misdariis, F. Nicolas, O. Warusfel, and R. Caussé, "Radiation Control on Multi-Loudspeaker Device: La Timée," in Proc. *Intl. Computer Music Conf.*, pp. 306–309 (Havana, Cuba) (2001 Jul.).

[18] S. Cecchi, A. Carini, and S. Spors, "Room Response Equalization—A Review," *Appl. Sci.*, vol. 8, no. 1, p. 16 (2017 Dec.). https://doi.org/10.3390/app8010016.

[19] Deutsche Gesellschaft für Akustik e.V., "Stimulus-Datenbank," https://www.dega-akustik.de/va/stimulus-datenbank (accessed 21 Jul. 2022).

[20] B. C. J. Moore, B. R. Glasberg, A. Varathanathan, and J. Schlittenlacher, "A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition," *Trends Hear.*, vol. 20, pp. 1–16 (2016 Sep.). https://doi.org/10.1177/2331216516682698.

[21] ISO, *Acoustics—Methods for Calculating Loudness. Part 2: Moore-Glasberg Method*," Standard 532-2:2017 (2017 Jun.).

[22] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1.x: A Toolbox for Reproducible Research in Auditory Modeling," *Acta Acust.*, vol. 6, p. 19 (2022 May). https://doi.org/10.1051/aacus/2022011.

[23] H. Fastl, and E. Zwicker, *Psychoacoustics: Facts and Models* (Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2007), 3rd ed.

[24] H. Haas, "Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache" [On the Influence of a Single

Echo on the Intelligibility of Speech], *Acustica*, vol. 1, no. 2, pp. 48–58 (1951).

[25] C. Visentin, M. Pellegatti, and N. Prodi, "Effect of a Single Lateral Diffuse Reflection on Spatial Percepts and Speech Intelligibility," *J. Acoust. Soc. Am.*, vol. 148, no. 1, pp. 122–140 (2020 Jul.). https://doi.org/10.1121/10.0001527.

[26] N. F. Bischof, p. G. Aublin, and B. U. Seeber, "Fast Processing Models Effects of Reflections on Binaural Unmasking," *Acta Acust.*, vol. 7, p. 11 (2023 May). https://doi.org/10.1051/aacus/2023005.

[27] E. Zwicker and U. T. Zwicker, "Dependence of Binaural Loudness Summation on Interaural Level Differences, Spectral Distribution, and Temporal Distribution," *J. Acoust. Soc. Am.*, vol. 82, no. 2, pp. 756–764 (1991 Feb.). https://doi.org/10.1121/1.1894635.

[28] J. S. Bradley and G. A. Soulodre, "Objective Measures of Listener Envelopment," *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2590–2597 (1995 Nov.). https://doi.org/10.1121/1.413225.

THE AUTHORS


Matthieu Kuntz


Klaus Kaetel


Dietmar Fischer


Bernhard U. Seeber

Matthieu Kuntz is a research associate and Ph.D. candidate at the Professorship for Audio Information Processing at the Technical University of Munich, Munich, Germany. His research focuses on the assessment of sound field synthesis techniques and audio reproduction for psychoacoustic and hearing research. He received his engineering diploma in vibrations and acoustics from the École Nationale Supérieure d'Ingénieurs du Mans (ENSIM), Le Mans, France in 2018.

•

Klaus Kaetel studied electronic engineering at the Technical University of Munich where he received his Master of Engineering. His experience in sound reinforcement led him to focus on sound transmission with maximal authenticity. He holds numerous patents in capturing and reproducing sound fields with rotating sound velocity vectors. Klaus Kaetel is a founding member and managing director at Molecular Acoustics GmbH.

•

Dietmar Fischer studied Mathematics and Physics at the University of Ulm, got his Dr.-Ing. degree with distinction in Astronautics and Aeronautics from the University of the German Federal Armed Forces, Neubiberg, Germany. After working for ESG Electronik System-Gesellschaft, Munich, Germany; Daimler Group, Ottobrunn, Germany; and Airbus, Bremen, Germany, he founded his first start-up in Iris Technology. He has been involved in start-up businesses ever since and is a consulting coach for the European Union in this field. He is experienced in satellite business, biometrics, traffic management, logistics, and others. In 2020 he started Molecular Acoustics GmbH as a founding member, together with Klaus Kaetel.

•

Bernhard U. Seeber has been a professor of audio information processing at the Technical University of Munich (TUM), Germany, since 2012. He obtained a Dr.-Ing. degree in electrical engineering with distinction from TUM in 2003. After a postdoc in psychology at University of California Berkeley (Berkeley, CA, USA), he joined the Medical Research Council (MRC) Institute of Hearing Research, Nottingham, UK, in 2007 to lead the Spatial Hearing Laboratory. He has worked on spatial hearing and virtual acoustics, real-time room acoustic auralization, signal processing for hearing devices, and acoustic testing. Prof. Seeber heads the Technical Committee (TC) on Hearing Acoustics of the Association for Electrical, Electronic & Information Technologies (ITG/VDE), served on the board of the German Acoustical Society, and is a member of the Acoustical Society of America and the Association for Research in Audiology. He received the Lothar-Cremer Award, the ITG Doctoral Thesis Award, the ITG Publication Award, and multiple best paper and poster awards.