NEURAL REPRESENTATION OF DIRECT-TO-REVERBERANT ENERGY RATIO IN RECORDED AND SIMULATED BINAURAL ROOM AURALIZATIONS

Viola Matten[a, *], Rüdiger Stirnberg[b], Steven van de Par[c], Stephan D. Ewert[c], Virginia L. Flanagin[a]
[a]German Center for Vertigo and Balance Disorders, Ludwig Maximilian University Munich, Munich
[b]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
[c]Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, Carl von Ossietzky Universität Oldenburg, Oldenburg
* To whom correspondence should be addressed: Viola.Matten@lrz.uni-muenchen.de

ABSTRACT

In a complex acoustic environment, sound localization involves the extraction and integration of numerous interrelated auditory cues. To understand how these cues are processed in the brain, studies typically isolate a single cue in an artificial experimental framework, to evaluate what brain regions process individual auditory cues. However, multivariate analyses facilitate more complex manipulations with greater ecological validity by providing a method for comparing between brain activity to a quantitative breakdown of the experimental stimuli. Concurrent advancements in virtual acoustics enable a systematic examination of spatial acoustics in complex realistic environments. Although these simulations have a high perceptual plausibility, they still alter auditory reverberation cues in a perceptible way. The impact of these subtle differences on neural processing is unclear. Auditory distance perception is a particularly challenging perceptual process to study, due to the relative nature of the sensory cues. Therefore, we conducted an imaging study to investigate the representation of auditory cues in recorded and simulated acoustic environments, while performing a distance discrimination task. We recorded the actual MRI environment to reduce room divergence, and the auditory simulations modeled reverberation with different degrees of accuracy. We used an acoustic analysis to determine the differences between the acoustic environments and used these quantitative measures to compare to the pattern of brain activity. We found that although the room auralizations were highly similar, it was possible to decode them from brain activity. The ratio of direct-to-reverberant energy level (DRR) was the only acoustic parameter that made a relevant contribution to brain activity. The locus of this activity was in the posterior auditory cortex.

Keywords Spatial hearing · What and where pathway · Auditory cortex · binaural auralization · direct-to-reverberant ratio

Introduction

Auditory perception, unlike vision is not inherently spatial; spatial perception of sound requires the conversion of a one-dimensional change in air pressure into a three-dimensional representation of the environment [1]. Nonetheless, hearing provides spatial information about the environment that vision does not capture. For example when an object is visually occluded, or when vision is impaired sounds, hearing can provide important complementary information about the world around us. Although we have a good understanding of how and where sound location in azimuth is computed by the brain, the cues and computations that the brain performs for auditory distance perception (ADP) are much less well understood.

This is in part because of the nature of the sound cues available for ADP. Multiple auditory cues provide distance information but differ in their effectiveness and reliability (for reviews see [2, 3]). Binaural cues play a role in ADP. For example distance estimates are more accurate for laterally presented sounds [4]. Interaural timing differences (ITD) from distance changes are below just noticeable differences (JNDs), however, interaural level differences (ILDs) are significantly modified by variations in distance at least for distances below $1\,\mathrm{m}$ [5, 6]. For larger distances, ILD becomes approximately independent of sound source distance [7–9].

One of the most important cues for ADP is the intensity of the sound, or sound level [10, 11]. However, sound level is a relative cue. The nervous system cannot determine the distance to a sound source based on intensity alone, but it requires knowledge about the initial level of the sound source itself [2, 3, 12]. Consequently, prior exposure or familiarity with the sound source allows us to better interpret intensity cues, thereby improving distance perception [13–15].

In contrast to sound intensity and binaural cues, reverberation provides a more infallible distance cue over a wide range of distances. While reverberation reduces localization accuracy in the horizontal plane [16], it has been shown to improve auditory distance perception [10]. The ratio of the energy of the direct sound to the energy level of the reflections from the walls is called the direct-to-reverberant energy ratio (DRR). DRR provides a reliable distance estimate of sound source distance [4, 12, 17–21]. Hence, reverberation is crucial for the perception of auditory distance especially when intensity and binaural cues are absent or ambiguous.

Several additional factors modify how subjects use and combine reverberation-related cues with other distance cues. For instance when highly familiar sounds are presented, subjects rely less on DRR and more on level cues [19]. Additionally, the properties of the acoustic environment, such as reverberation time (RT30), modify the effectiveness of DRR [22, 23]. Furthermore, although DRR distance cues do not require familiarity with the acoustic environment, prior exposure to the reverberant sound field results in a sustained improvement in the accuracy of distance judgments [3, 18, 24, 25]. This improvement is stable across different listener positions within the same room [3, 25].

Similarly, unexpected changes in room reverberation can have a detrimental effect on auditory spatial perception in general. If a room auralization differs from the room subjects are physically located in, the amount of sounds localized inside their head increases. This increase in internalization as a consequence of a divergence between the acoustics of the recording room and the physical room is called room divergence [26–28]. The relationship between internalization and distance perception is still a matter of debate (for a review see [29]). Nevertheless, it indicates that the expectation of the surrounding room acoustics influences auditory spatial perception, especially in complex acoustic environments [28, 30].

In summary, auditory distance perception involves the interpretation of cues related to the stimulus and the surrounding environment. ADP is also influenced by higher cognitive factors such as knowledge of the environment or expectation of the acoustics. The main challenge for understanding neural processing of auditory distance information is to disentangle the contributions of these different interdependent units [31].

The auditory cortex (AC) located bilaterally along the Sylvian fissure is one of the crucial regions for processing sound. Localization of sound sources requires an intact AC [32, 33]. Spatial features are primarily processed in posterior areas of the AC including the planum temporale and the posterior superior temporal gyrus [34–38]. The posterior non-primary AC responds to changes of the sound source in azimuth [38–42], distance [42, 43], and movement [44, 45]. How auditory spatial information is encoded in the AC is still a matter of debate. It is unclear whether neurons in the non-primary AC are sensitive to basic cues such as ILD, DRR or intensity or whether they represent higher level features of space such as sound source location [46]. While basic processing and extraction of binaural cues occurs already at the level of the brainstem [47–49], binaural cues have been shown to also modulate the neural response in the AC in particular the posterior part of the contralateral superior temporal gyrus [50–55]. Similarly, an increase in intensity enhances the measured response throughout the whole ascending auditory pathway [56–59]. However, the exact locations that are tuned to intensity are still disputed [60–65]. The posterior AC responds to intensity-independent auditory distance cues [43], also in the frontal plane where DRR cues are relevant [42]. However, in order to separate out intensity from reverberation-related cues, the task performed was not related to distance estimation. It still remains unclear whether these neurons respond to reverberation-related cues or whether they reflect an abstract representation of sound location.

Acoustic room simulations provide a technique for separating out distance cues through manipulations of the acoustic properties of the environment. Advancements in spatial acoustic simulations provide a high level of perceptual plausibility, although a perfect physical reproduction might not be necessary for perceptual plausibility [66, 67]. These simulations provide a method for systematically manipulating acoustic parameters that would only be possible with multiple identical rooms that differ only in one parameter. Indeed, it is possible with these simulations to create environments that are physically impossible. These new possibilities open up the ability to investigate auditory spatial perception in a more controlled and complex environment. Instead of manipulating individual parameters like DRR or intensity, it is now possible to create a more naturalistic scenario, by simulating rooms with different acoustic properties and then using advanced fMRI analysis techniques to separate out the individual acoustic parameters involved.

Concomitantly, advanced multivariate analyses of brain data have developed that provide a means to test complex and non-linear relationships between fMRI data and the stimuli presented. For example, representational similarity analysis (RSA), allows for a direct comparison between the activity patterns measured with fMRI and computational models [68]. An abstract representation of the activity maps are computed from the data and then correlated with model predictions based on the stimuli presented and/or behavioral outcomes. With this approach it is possible to quantitatively assess the degree to which each acoustic parameter is

represented in the brain response without the need to explicitly define a mapping between the computational models and the imaging data.

Here we use recorded and simulated virtual acoustic environments to test the role of different acoustic parameters on brain activity during a distance discrimination task, using functional magnetic resonance imaging (fMRI). First, we auralized the room where the experiment was conducted, to reduce room divergence and thereby reduce in-head localizations and increase familiarity (for more details regarding room divergence see [28]). To measure and record the acoustics of the MRI scanner room we built an MRI compatible head and torso model, adapted from a Kemar simulator [69] to capture the binaural impulse response (BRIR). Two additional BRIRs were generated from room acoustic simulations of the same MRI room with varying degrees of accuracy, for a total of three BRIRs: the recording of the actual MRI room, a perceptually plausible simulation, and a simplified simulation of the same room. Subjects were then presented with speech sounds via MRI compatible headphones at one of two lateral positions convolved with one of the three BRIRs and asked to discriminate whether the sound came from the close or far position, to test for differences in brain activity in response to the three BRIRs during distance judgements. We chose an explicit distance task to focus subjects' attention on the relevant auditory cues, to enhance brain activity in non-primary cortices [59, 70–74]. Data were collected with a novel MRI scanning sequence that provides interleaved silent steady-state (ISSS) [75] fMRI data acquisition with silent periods for auditory stimulus presentation, but using a 3D echo planar imaging (3D-EPI) sequence with high spatial and temporal resolution and improved tSNR in subcortical brain regions [76]. By using a combination of traditional data analysis methods and pattern extraction methods such as multi-voxel pattern analysis (MVPA) and representational similarity analysis (RSA), we searched for evidence of separation of room auralization and distance in brain activity. Then, by analyzing the auditory properties of these sounds and correlating them with the brain activity patterns, we were able to determine the acoustic parameters that drive brain activity. With this approach we were able to decode the three room simulations from brain activity. The acoustic analysis and subsequent RSA revealed that only DRR contributed to separation of both BRIR and distance. Both the decoding and the RSA analyses were driven by signals in the posterior AC as well as auditory association areas relevant for processing reverberation-related cues. The results suggest the presence of neurons sensitive to DRR in the posterior AC, and demonstrate the usefulness of virtual acoustic simulations and RSA in understanding how we process auditory spatial information.

Results and Discussion

The reduction of auralization complexity alters distance perception

Two sound source locations were convolved with three BRIRs that capture the acoustics of the MRI magnet room. The receiver was positioned approximately in the middle of the MRI room in front of the MRI (Fig. 1a). The high magnetic field and constant background noise generated by cooling systems and helium (re-)compression have inhibited previous acoustic measurements of the MRI magnet room. However, we developed a system that enabled us to overcome these issues and record BRIRs inside the MRI scanner room (see Materials and Methods for details).

Distance discrimination was compared between the recorded BRIR (**Recorded**) and two simulations using the room acoustical simulator RAZR [77]: One simulation modeled the magnet room as closely as possible (**RAZR**), but neglected objects inside the room. The other simulation was further simplified by auralizing only first order specular reflections and removing the late reverberation, reducing the simulation to a 1st-order image source model (**1st-order ISM**). This led to slight differences in the **Recorded** and **RAZR** BRIRs while the removal of the late reverberation substantially modifies the reverberant part of the impulse response (Fig. 1b).

To investigate how the observed differences in auditory cues influence the neural processing of spatial acoustics, we used a task that focused the subjects' attention on room acoustical properties by means of a distance discrimination task. For each auralization, BRIRs were generated at one of two distances (Fig. 1c). The distances were chosen to maximize the distance difference between the sound sources given the geometry of the magnet room, but close enough to the JNDs to challenge the participants and maintain a high level of attention (see Supplemental information). Subjects indicated whether they located the sound at their feet (**Close**) or in the corner of the room (**Far**). Additionally, to look for differences in externalization between BRIRs, subjects could alternatively indicate whether sounds appeared inside their head and they were not able to not assign a distance to the sounds (**In head**).

Subjects reported perceived differences in auralizations, but, as expected, struggled to correctly identify the distances as near or far. The percentage correct for discriminating distances is not significantly above chance (Fig. 1d). Note, that we used an alpha value of 0.05 for all statistical tests. To investigate differences between the response distributions of the three auralizations, We conducted a mulitnomial logistic regression (see Materials and Methods). If we compare the **1st-order ISM** with **Recorded**, while the distance is held constant, the likelihood of subjects' responding **In head** was significantly increased compared to the responses **Close** (Wald-$z(1.35) = 11.034, p < 0.001$) or **Far** (Wald-$z(-2.24) = -17.4, p < 0.001$). Similarly, the comparison between the **1st-order ISM** and **RAZR** revealed that subjects pressed significantly more often **In head** in the ISM instead of **Close** (Wald-$z(1.62) = 11.89, p < 0.001$) or **Far** (Wald-$z(2.72) = 19.23, p < 0.001$). Together these observations showed that

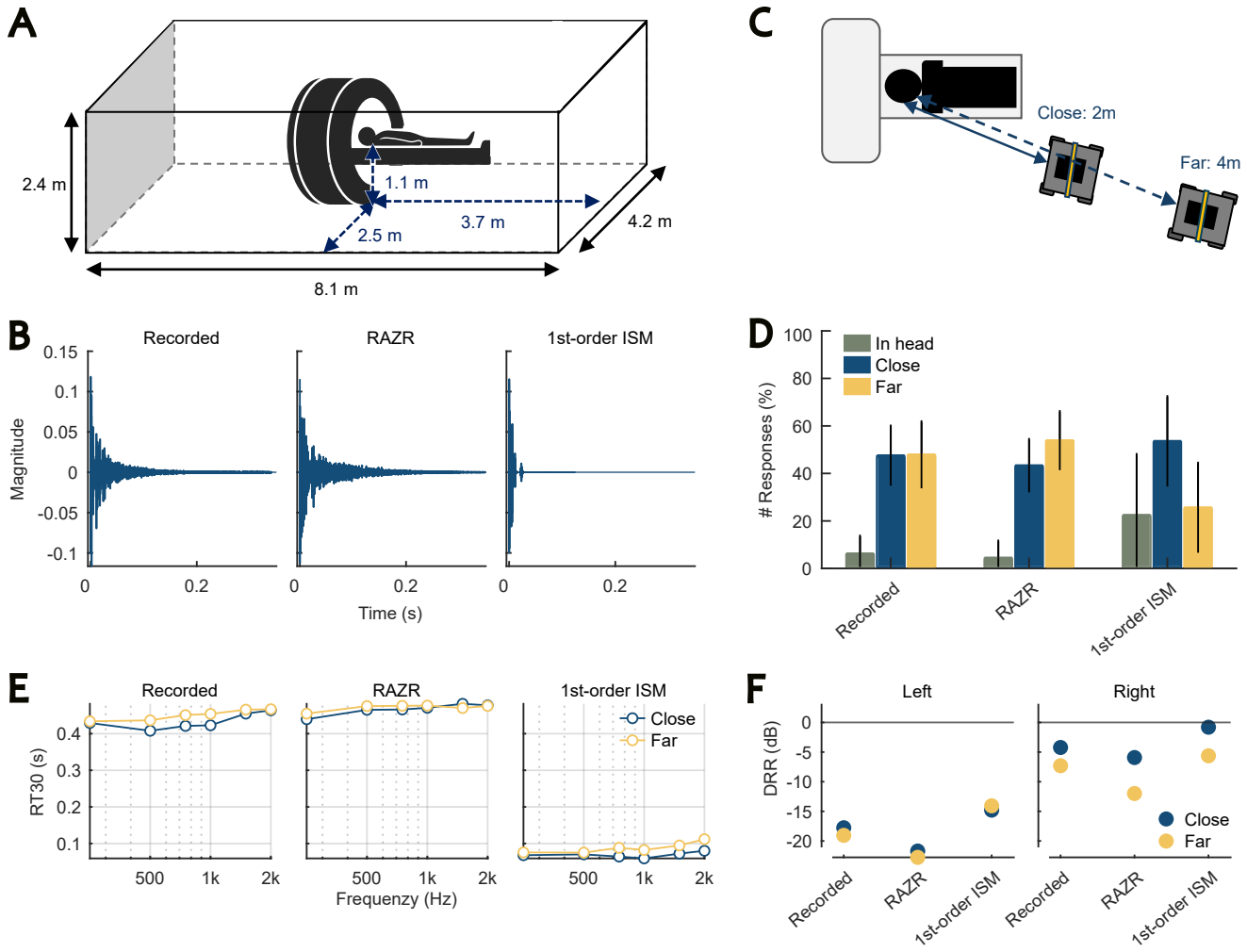

Figure 1: Three sets of BRIRs auralized the MRI room with varying degrees of accuracy. (A) MRI room dimensions. The receiver was located on the patient table in front of the scanner bore, at the meter values shown. (MRI icon: © kilroy79 |Adobe Stock) (B) Representative BRIRs for each auralization. The recorded MRI room (**Recorded**,Left), compared with two simulations generated with RAZR [77]. One simulation modeled the acoustics as accurately as possible (**RAZR**, Middle), the other simulation was a simplified 1st-order ISM without late reverberation (**1st-order ISM**, Right). BRIRs were peak normalized across auralizations. (C) BRIRs were generated at two distances for each auralization. The loudspeakers were placed on a cart to align the vertical position with the subject's head and approximately 30 degrees to the right of straight ahead. (D) Subjects performed a distance discrimination task between the two distances shown in C. The distribution of responses for each auralization is plotted here. The number of **In head**, **Close** and **Far** responses is normalized by the total number of responses per auralization, resulting in a percent value for each response. (E-F) We then performed an auditory analysis of the three room auralization. Reverberation time and DRR are plotted for the three BRIRs corresponding to one of the distances and auralization type. Labels for the auralizations as in B

the amount of **In head** responses is significantly increased in the **1st-order ISM** compared to **Recorded** and **RAZR**. Hence, subjects were less likely to assign a distance to the sound source when late reverberation was removed. This finding is in line with literature reporting an increase in internalization of sounds when reverberation is reduced [29, 78–81].

Moreover, for the **1st-order ISM** the likelihood of subjects' responding **Close** instead of **Far** is significantly increased compared to **RAZR** (Wald-$z(1.09) = 15.48, p < 0.001$) and **Recorded** (Wald-$z(0.89) = 12.52, p < 0.001$). Therefore, sounds auralized with the **1st-order ISM** are perceived closer compared to the two advanced auralizations with enhanced reverberation consistent with literature relating a reduction in reverberation to reduced distance estimates [2, 3, 18, 82].

To quantitatively assess acoustic differences in the auralizations with the three different sets of BRIRs, an auditory analysis was performed that extracted room-acoustical parameters that are known to be important for characterizing the perception of room acoustics. Among others, these parameters include RT30, and DRR (see Supplemental information for all auditory parameters

measured). The perceptible acoustic differences in the three BRIRs (**Recorded, RAZR, or 1st-order ISM**) should be related to these acoustic parameters. The auditory analysis revealed that the auralizations were highly similar, specifically between the **Recorded** and the realistic simulation **RAZR**. Primary differences were seen for the reduced simulation **1st-order ISM** in RT30, intensity and DRR (see Fig. 1e and 1f). Removing the late reverberations resulted in a reduction in RT30 (Fig. 1e). The increase in **In head** responses is consistent with previously reported increases in internalization with decreasing RT30 ([78, 81, 83]).

Although we observe a higher similarity between the response distributions of **RAZR** and **Recorded** compared to **1st-order ISM**, distance perception differs significantly between the two auralizations (fig 1d). However, note that the differences between**RAZR** and **Recorded** are much smaller than for the comparison with the **1st-order ISM**. For **RAZR** we observe more **Far** responses compared to **Close** (Wald-$z(0.21) = 3.26, p = 0.001$). In addition, we find an increase in the likelihood of subjects responding **Far** instead of **In head** (Wald-$z(0.48) = 3.08, p = 0.002$) for **RAZR** compared to **Recorded**. These results suggest that distance perception is enhanced if sounds are auralized with RAZR instead of the recorded BRIRs.

The DRR analysis predicts a similar variation in distance perception (Fig. 1f). DRR is slightly reduced for **RAZR** compared to the recorded BRIRs, which is known to increase perceived distance [3, 43]. DRR is however strongly increased for the **1st-order ISM**, which should lead to shorter perceived distances. We hypothesize that the increase in DRR for the **1st-order ISM** is related to the removal of the late reverberation while the difference between **RAZR** and **Recorded** is caused by additional early reflections from the scanner bore during the recording as well as shielding effects of the scanner (for a more detailed analysis and justification see Supplements).

Overall, the multinomial logistic regression of the behavioral data demonstrates that subjects could perceive differences between the auralizations, and that the subtle changes in acoustic cues caused by different levels of accuracy in the BRIRs are sufficient to modify auditory distance perception. A systematic psychoacoustic experiment and analysis would be necessary to clarify how the observed behavior relates to the individual cues.

Room auralizations recruit the same set of brain regions

Room auralization had a significant effect on perception, in particular the room simulation with less reverberation moved responses closer including into the head. To then investigate the brain activity during the task, we conducted a general linear model analysis and compared each room auralization to the silent condition (see Materials and Methods). We observed activity in primary and nonprimary AC as well as the thalamus and premotor areas. A visualization of the brain areas that were activated can be seen in Figure 2a.

The strong activation of the AC and thalamus demonstrates the effectiveness of the developed imaging sequence. The main challenge for acoustic experiments in the MRI machine is the noise resulting primarily from rapidly alternating readout gradients. The higher the spatial and temporal resolution, the higher the gradient amplitudes and the longer the ramp times, which typically increases the noise. Multiple approaches exist to reduce noise while measuring brain activity. We implemented an EPI sequence with ISSS imaging that alternates loud (with readout gradients) and silent periods (without readout gradients) for stimulus presentation [75]. This approach improves imaging of ascending auditory pathway [84]. We extended the method to a multi-shot 3D-EPI sequence, increasing temporal resolution to $0.57\,\mathrm{s}$ while maintaining a relatively high spatial resolution of $2.5\,\mathrm{mm}$ isotropic and, at the same time, minimizing acoustic noise further (see Materials and Methods).

When looking at the brain activity evoked by each auralization individually, neither recruited brain regions nor overall activation strength differed between the auralizations. As expected, the sounds presented during the distance discrimination task recruit the same set of auditory brain regions independent of auralization, but with a classical fMRI analysis it was not possible to separate out differences between the room auralizations.

Room auralization can be decoded from brain activity

Although we observed differences in behavior for the three room auralizations, this was not reflected in the univariate regression analysis (GLM) of brain activity. In other words, differences between the room auralizations did not change the average activation strength of a single brain area, nor did they activate additional brain areas. While a univariate approach answers questions about the size and strength of brain activity, it does not reveal changes in pattern within these regions [85, 86]. Since the auditory simulations **RAZR** and **1st-order ISM** modeled the **Recorded** room, they differ only slightly in their auditory cues, with values close to JNDs. Therefore, brain activity is most likely modified in a fine-grained fashion by these subtle difference in room auralizations.

We hypothesized that auralizations alter the pattern of activation rather than the average activation strength of entire brain regions. In contrast to the GLM approach, multi-voxel pattern analysis (MVPA), or decoding, allows us to identify whether individual conditions evoke a consistent nonlinear change in the brain's activity pattern [85, 86]. We trained three linear support vector

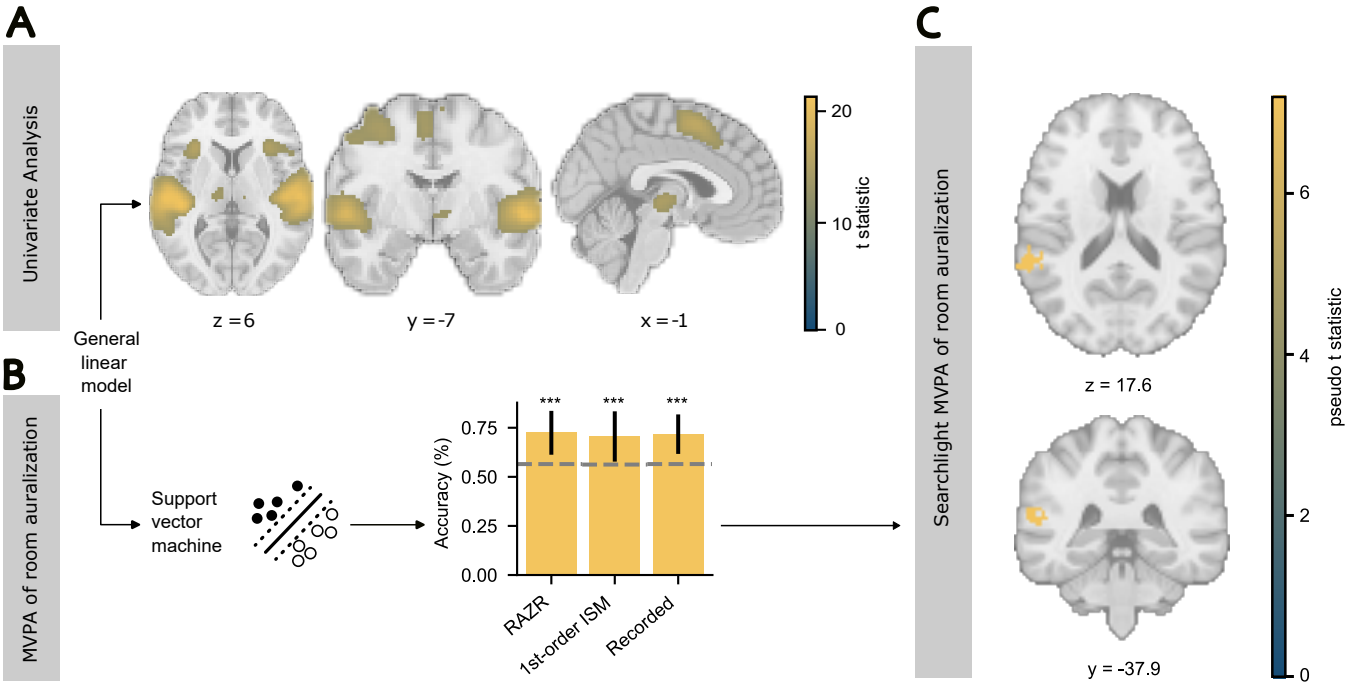

Figure 2: Univariate fMRI analyses showed similarities between room auralizations, whereas multivariate analyses revealed distinct patterns of activation for each reverberation. (A) Univariate regression analysis (GLM). We observed activity in the auditory regions including AC, thalamus, Heschl's gyrus and planum temporale. Smaller clusters were found in frontal and motor areas including the insula and the cingulate cortex (see Tab. S1 for details) . (B) Multivoxel pattern analysis (MVPA). Decoding accuracy of room auralizations from brain activity. Dashed lines indicate chance accuracy. (C) Localization of MVPA. Statistical map of the searchlight based on an SVM classifier for each BRIR auralization compared to the other two. A significant cluster was observed in the left superior temporal lobe. Maps were threshold at $p = 0.05$ FWE and only clusters with more than 50 voxels were considered significant. Brain activity patterns are overlaid on the MNI152 single subject template brain and coordinates are given in MNI space. For further details regarding cluster volume and peak statistic refer to tab. S2.

machine classifiers on the activity of the entire brain to test whether each room auralizations can be decoded from the other two based on the imaging data.

We collected eight runs per subject (see Materials and Methods). Based on these runs we conducted a leave-one-run-out crossvalidation to train the classifier. Note: We were not interested in finding a classifier that can generalize across all subjects but rather wanted to obtain an estimate of the classification performance for each subject individually. Hence, we decided to not tune the hyperparameters on an extra validation set to minimize the effect of noise but instead averaged the crossvalidation scores to obtain an estimate of the accuracy with which the room auralization can be decoded for each subject. Subsequently, we averaged the accuracies across subjects.

The final mean accuracy with which each trained classifier can discriminate one auralization from the other two can be seen in fig. 2b. Chance accuracy was determined with permutation testing. For each auralization the classification accuracy is significantly above chance (**Recorded**: t(0.72)=8.29, $p < 0.001$; **RAZR**: t(0.72)=7.80, $p < 0.001$; **1st-order ISM**: $t(0.71) = 6.00$, $p < 0.001$;). We did not find a significant difference between the mean classification accuracies ($F(2, 27) = 0.21$, $p = 0.81$). Hence, the classifiers performed equally well for all three auralization types. These results confirm our hypothesis that the small differences between the room auralizations can still modify the pattern of brain activation. It is worth noting, that we observe this effect although the acoustic parameters variedin other dimensions unrelated to the room auralization. The distances as well as the different spoken words introduced additional variations in the acoustic parameters as large as, or even larger than the variations introduced by the different auralization types (see Supplements for the acoustic analysis across distances and [87] regarding acoustic differences in speech).

The posterior superior temporal lobe encodes room auralization

After determining that the rooms auralizations could be decoded based on the pattern of activity across the entire brain, we wished to identify what specific brain regions contribute the most to this classification. To achieve this we conducted a searchlight analysis based on a classifier that utilized linear support vector machines with a one-versus-rest multiclass strategy. To assess

significance at the group level, we performed a permutation test and determined significance with a variance-smoothed t statistic to mitigate noise (for more information see [88]). Additionally, clusters with less than 100 contiguous voxels were considered noise. A cluster with significant decoding accuracy was observed in the left superior temporal lobe including the superior temporal gyrus, the planum temporal and the Heschl's gyrus (Fig.2c.

This observation suggests that the superior temporal lobe is sensitive to acoustic parameters that change between room auralizations. Alternatively, the region could encode abstract representations of auditory space (for a review see [46]). It is known that the posterior AC processes spatial information (see introduction) but it is still unclear how the information is encoded. Individual neurons might represent specific locations in space or respond to abstract spatial characteristics such as egocentric distance, azimuth or elevation instead of being sensitive to basic acoustic cues.

Recent literature has proposed the existence of neuron populations responsive to DRR in the posterior part of Heschl's gyri, the planum temporale and the superior temporal gyrus [42, 43]. As the room auralizations here differ primarily in their reverberation (see Supplements), our findings are consistent with this proposal, suggesting a higher sensitivity in the superior temporal lobe to changes in reverberation. Nevertheless, because we were interested in using realistic stimuli, the three room auralizations differed in other acoustic parameters. The question remained whether the superior temporal lobe processes reverberation-related cues in isolation or an integration of several acoustic features to form an abstract representation of auditory space. To explore this question, we conducted further analyses that study the relationship between the brain activity and the changes in acoustic parameters.

Acoustic parameters for the room auralizations show similarity with brain activity

The classification analysis revealed that changes auditory reproduction of the MRI room shape the neural response in a unique and consistent way. To understand the relation between the observed changes in the neural response and the underlying acoustic parameters, we conducted a representational similarity analysis (RSA) based on behavior, categorization and the acoustic parameters of the sounds. While MVPA tells us whether discriminant information about individual stimulus conditions exists in the multivariate pattern of activity, RSA allows us to relate this activity pattern with specific computational models of the stimuli and resulting behavior [68, 89]. RSA compares the similarity between activity patterns for different conditions with the similarity predicted by the models. Thus, the method enables us to characterize what information is represented in the brain and to correlate it quantitatively with the model predictions.

We defined a set of acoustic parameters known to be relevant for perception of spatial acoustics, in particular during a distance perception task including DRR, RT30, MFCC, ILD, perceived loudness, the Mel-Frequency-Cepstral Coefficients (MFCC) and interaural cross-correlation (IACC) [2, 3, 90]. For each condition this set of parameters was computed. Using basic measures such as pearson correlation or euclidean distance we determined the similarity between conditions for each parameter. This procedure results in a similarity matrix for each parameter. At the same time, we compared the activity maps obtained from the GLM for each condition resulting in one similarity matrix per subject. Subsequently, the matrices based on the imaging data were averaged across subjects.

We ended up with one similarity matrix characterizing the information in the activity patterns and a set of matrices describing the acoustics of the presented stimuli. By comparing the matrix based on the data and the matrices based on the acoustic parameters, it is possible to investigate to what degree the observed brain activity contains information related to the various parameters without drawing assumptions about how the parameters might be computed by the brain (see Materials and Methods and [68]). As a result we obtained a correlation coefficient for each acoustic parameter. This coefficient can be interpreted as a prediction accuracy that indicates to what degree each parameter is represented in the brain.

In addition, we determined several similarity matrices that captured the behavioral response to the stimuli per condition and a set of matrices characterizing the category of the condition i.e. two conditions were defined to have a high similarity of 1 if they belonged to the same category and 0 otherwise. The different categories were based on the auralization type or the distance. Neither the behavioral nor the categorical matrices were correlated with the imaging data. Hence, we focused the subsequent analysis on the acoustic models i.e the matrices based on the acoustic parameters.

As seen in Fig. 3a, for the acoustic parameters the highest correlation i.e. accuracy was observed for the parameters based on DRR, demonstrating the importance of this cue. However, the accuracy of the DRR models is not significantly higher than the other individual models. Testing against zero yields p-values above 0.1 for all single acoustic models. A single acoustic parameter is unlikely to significantly predict activity across the entire brain, hence these results are not surprising.

However, with RSA it is also possible to fit a linear weighted combination of the individual acoustic parameters to evaluate the contribution of each model to the optimal model. Although, the exact combination is presumably highly nonlinear, this approach allows to investigate the importance of each parameter given the brain activity in the current experiment [68, 89]. We therefore fit a weighted model with all of the parameters to the data using crossvalidation. As expected, the full weighted model performed better than the individual models (Fig. 3b, $p = 0.06$) and the weights obtained for the individual acoustic parameters

	mean DRR	Δ DRR	RT30	MFCC	IACC	Loudness	ILD
Weighted model	0.66	0.40	0.06	0.11	0.02	0.17	0.04
Weighted model DRR only	0.71	0.52	-	-	-	-	-

Table 1: The weights for the weighted model involving all studied acoustic parameters and the weighted model based on DRR.

are depicted in fig. 3b. The largest weights were observed for the two DRR models (Tab. 1). The absolute values of the weights do not have a direct interpretation. However, if compared with each other they provide measure to quantify how well one model explains the data compared to another model. The next highest weight was for loudness, but all of the weights of the remaining parameters ranged from zero to 0.2.

Due to the low weights of many of the acoustic parameters in the full model, it was likely that a model with a reduced number of acoustic parameters fits the brain activity better than the full model. If a reduced model fits the data better, it would be stronger evidence that only certain acoustic parameters were relevant for the task performed. To identify which acoustic parameters improved the fit, we systematically removed individual acoustic parameters from the weighted model and recalculated the accuracy for the remaining model. This approach yielded that the best fitting model is a combination of the two DRR-related parameters ($p = 0.02$). Hence, we conclude that brain activity is best described by a model that captures variations in DRR. The observation is in line with the differences in behavior found, suggesting that reverberation is the main cue used by subjects throughout the experiment.

These findings reveal that subtle variations in auditory cues are nonetheless represented in brain activity, and that a mutlivariate approach relating an acoustic analysis of the stimuli to the brain activity can disentangle what parameter are relevant to a task. As the stimuli differed primarily in their reverberation, and the task focussed subjects'attention on DRR, we cannot definitely determine the degree to which these effects are dependent on one or the other.

Interestingly, although loudness is a strong cue for auditory distance perception, and although it had the third highest weight in the full weighted model, leaving it out of the model improved the fit, suggesting that it was not relevant for this task. We believe this is primarily a result of the small differences between the two distances, and the lack of an absolute intensity reference, rendering intensity differences less interpretable for distance in this task.

DRR is represented in the posterior auditory and association cortices

In the previous RSA, the neural response of the entire brain was used for the analysis, but it did not allow us to determine where in the brain the pattern of activity reflected DRR. We assume, however, that the observed correlation between the brain activity and DRR, can primarily be attributed to a set of regions in the AC. To identify whether the intrinsic representation of a certain region resembles the representation predicted by DRR, we conducted a searchlight analysis using the weighted-DRR model. This resulted in a statistical map displaying, for each voxel, the similarity between the model and the activity in the proximity of that voxel [68, 91].

We observed a high correlation with the DRR model in the posterior AC and the auditory association area including the superior temporal lobe contralateral to the stimulus, the medial temporal lobe and the angular gyrus (Fig. 3c). Although the remaining auditory cues, such as ILD or RT30, were not significantly represented in overall brain activity, localized activation may still correlate significantly with individual parameters. Therefore, we performed a searchlight RSA for each of the parameters studied in the previous chapter. No significant correlation was found between any of the parameters and a region of the brain. We hypothesize that subjects focused their attention on DRR rather than on the remaining parameters thereby enhancing the sensitivity to changes in DRR [92–94]. These results suggest that neuronal populations in these areas are sensitive to variations in DRR resulting from distance and BRIR.

The result is consistent with previous work showing that dorsal auditory pathways process auditory location information. Neurons in the posterior non-primary AC encode spatial information about the sound such as horizontal direction changes. [34–38, 38]. Furthermore, the superior temporal gyrus has been associated with reverberation [42, 43]. Kopčo et al. conducted two adaption fMRI experiments where they varied distance or intensity independently while subjects performed an unrelated task. They found adaptation to distance, independent of intensity and binaural cues, in the posterior AC. Since DRR is the most important remaining cue, they suggested a relationship between reverberation based cues and the posterior AC. However, they could not disentangle whether the observed activity was due to the region encoding an abstract representation of auditory distance or whether it can be attributed to neuronal populations sensitive to reverberation-related cues. Our data provide evidence that these areas are specifically sensitive to DRR. Additional evidence that the posterior superior temporal lobe processes reverberation-related information comes from recent evidence that non-primary auditory cortices including the posterior superior temporal cortex show invariance in their responses to sounds with or without background noise (CITE McDermott). This invariance in a background signal, such as reverberation, would be necessary to be able to extract distance relevant informations.

Taken together, these findings support the hypothesis that neural populations in the AC that respond specifically to changes in DRR.

The correlation we found with DRR in the superior temporal lobe was contralateral to the presented stimuli, consistent with the results by Kopčo et al.. However, as sounds were presented only on one side, additional studies are required to verify whether DRR is encoded stronger by the contralateral hemisphere. However, the observed relationship between DRR and the pattern of activity in the AC could also be caused by other cues that correlate with DRR. A number of spectral and temporal cues have been suggested as relevant for the perception of DRR that covary monotonically with DRR such as IACC, spectral variance or envelope, as well as buildup and decay times [2,17,21,95,96]. We included those cues known to be related to distance perception allowing us to identify the type of distance-relevant acoustic cue represented in brain activity. A next step could be to specifically study auditory cues that correlate with DRR. Additionally, semantic perception is known to activate regions in posterior AC as well. Although we used speech in all our conditions, it was irrelevant for the task. Also the semantic content was the same across the room auralizations. Therefore, it is unlikely that our observations are purely a result of semantic processing. Nevertheless, to completely test this possibility further studies are required that use non-verbal sounds.

We were surprised that the binaural cues did not have a strong representation in the brain activity during our task. The sounds came from the right side, and differences in binaural cues for the differen distances and BRIRs were seen in the auditory analysis (See Supplements). However, the processing of binaural auditory cues such as ILD has been shown to already occur at the level of the brainstem (for a review see [47,97]). In contrast, the cortex has been associated with higher order processing [34,35,37,98,99]. Although the 3D-EPI sequence that was modified for ISSS imaging (see Materials and Methods) improves signal detection in subcortical areas, in contrast to most fMRI studies, we collected the fMRI data sagitally in order to employ completely silent, non-selective volume excitation. Nevertheless, a typical phase encoding direction approximately along the anterior-posterior commissure line was used (AP). Therefore, as usual, susceptibility-induced distortions and smearing of the signal occurred in AP direction, particularly in the brainstem. However, due to the narrow-banded water-selective excitation used, more off-resonance-induced signal drop-outs may have occurred in addition to typical intra-voxel dephasing. While geometric distortion correction can mitigate the observed distortions and signal smearing to some extent, it cannot recover signal that may have been lost in brainstem regions due to the latter. Hence, the current dataset does not provide conclusive evidence concerning the representation of auditory cues in subcortical areas. Additional studies with simpler tasks and analyses should be performed to improve the sequence and adapt the preprocessing to consistently measure the activation in the brainstem.

Conclusion

We investigated the impact of subtle variations in auditory cues based on different BRIRs of the same physical room on the neural processing of spatial acoustics during a distance discrimination task. We used the same room and subjects were in and had them perform an explicit distance task to improve ecological validity of our results. Although acoustic differences between distances and BRIR auralizations were small, they influenced the behavior of subjects and modified the multivariate neural activity pattern. BRIR auralization of the same room can be decoded from brain activity. Changes in the brain activity patterns for distance and auralization were correlated with resulting variations in DRR. These variations in DRR are represented in posterior non-primary auditory cortices responsible for higher-level cognitive processing such as identification of sound location. This provides further evidence that populations of neurons located in the posterior AC are sensitive to DRR. It is possible that these regions directly extract DRR from auditory signals, how they may be extracted elsewhere and only integrated and combined with other cues in these posterior auditory cortical regions. DRR may be extracted somewhere in the brainstem and is only integrated at this stage with abstract information about room acoustics. As the posterior AC and the association cortex are known to be involved in higher-level cognition, the observed correlation with DRR in these regions supports the relevance of DRR for extracting abstract information about the sound source [34–38]. We also demonstrate the use of RSA in determining the relevant acoustic properties in a complex auditory task. We are unaware of an auditory fMRI study that has used RSA in this manner before. We also developed MRI-compatible recording techniques and modified a 3D-EPI sequence to have quiescent periods for auditory imaging that can be used to measure brain activity ever few hundred milliseconds with a high isotropic resolution. We believe this sequence and recording equipment can benefit the larger auditory imaging community.

Materials and Methods

Subjects

A total of 63 subjects participated in the study. Out of them 33 had to be excluded due to poor data quality, initial problems with the setup or because subjects had to abort the experiment. Seventeen female and 13 male subjects with normal hearing (27 right-handed, 3 left-handed, age range: 19 to 34 years) were included in this study. All persons gave their informed consent to participate in the study and were monetarily compensated for their time at 12 € per hour. The experimental protocols were conducted in accordance with the Declaration of Helsinki and was approved by the independent ethics committee of the

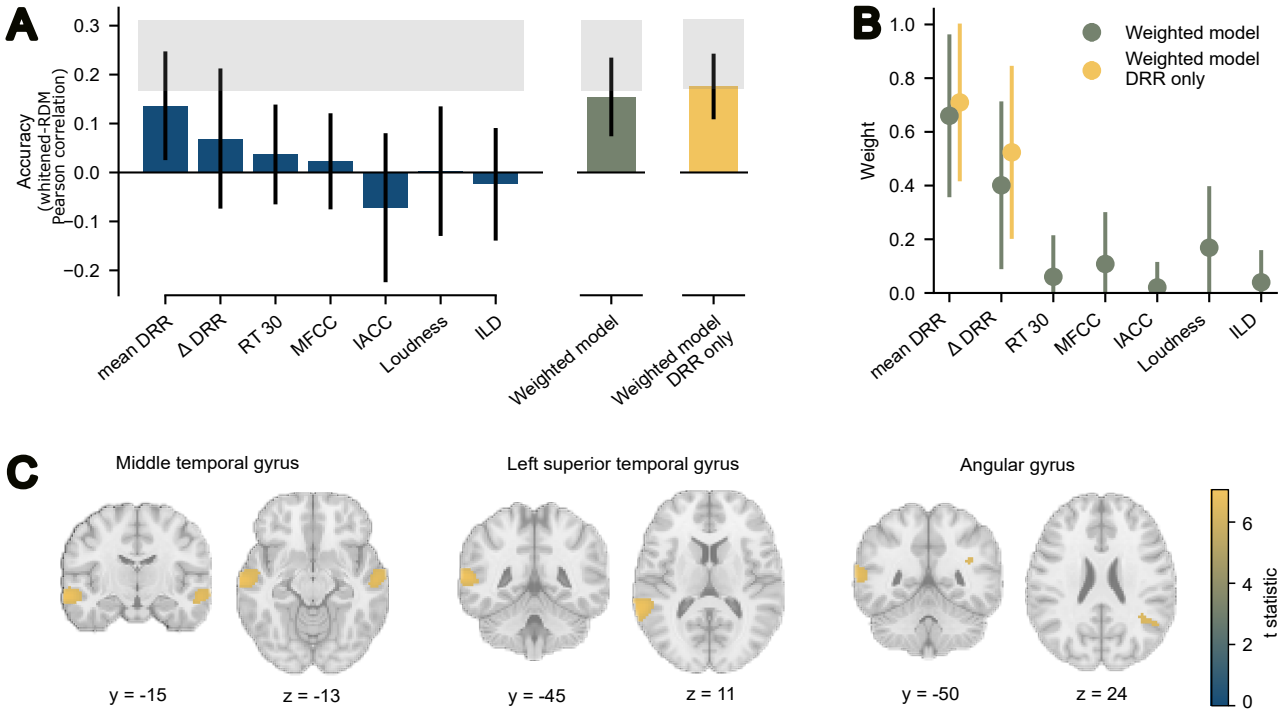

Figure 3: RSA revealed that changes in DRR are represented in the posterior AC. (A) Prediction accuracy of models based on variations in auditory cues. Error bars were obtained using bootstrapping with $N = 1000$. The gray bar indicates the noise ceiling. It provides an upper and lower bound on the performance of an optimal model given the noise in the data. Blue: Accuracy of models based on individual acoustic parameters. The two DRR parameters differ in the way the left and right channel were combined. While **mean DRR** was based in the average of both channels, **ΔDRR** was based on the difference. Green and yellow: Models based on a linear weighted combination of acoustic parameters. **Weighted model** combined all parameters shown on the left. **Weighted model DRR only** combined the two DRR models. (B) The weighted models were fit using crossvalidation and bootstrapping ($N = 1000$). The plot shows the obtained weights for the weighted model that combined all examined parameters (green) and for the best fitting model containing only the DRR parameters (yellow). (C) Results of searchlight RSA for the **Weighted model DRR only**. A cluster radius of 5 voxels has been used. Images are thresholded at 0.05 FWE. Brain activity patterns are overlaid on the MNI152 single subject template brain and coordinates are given in MNI space. For details regarding cluster location, size and peak statistic refer to tab. S3.

University of Munich (21-0035). Participants were not recruited if they were pregnant or had any contraindication for entry into the MRI bore.

Stimuli and Setup

Stimuli were originally anechoic sounds consisting of semantically neutral words with a fixed duration of $550\,\mathrm{ms}$ that were designed specifically for acoustic experiments in the MRI scanner [100]. Sounds were subsequently convolved with one of three sets of BRIRs at one of two fixed positions within the environment. Three sets of BRIRs were created by capturing the acoustics of the magnet room (Fig. 4a) with varying acoustic complexity. For each auralization of the MRI room, BRIRs were generated at two distances (see Fig. 1c). Both distances were rightly lateralized with respect to the listener, as externalization is reported to be reduced for frontal sources compared to lateral ones [67]. The farthest sound was placed in the right corner of the room to maximize the distance between the source and the receiver.

The first set of BRIRs was obtained by recording the acoustics of the magnet room. Since the magnetic fields of the scanner prohibit the usage of traditional dummy heads, we built an MRI compatible head and torso simulator. We combined a three-dimensional print of a KEMAR head and torso model by Braren and Fels [69] with pinnae from an existing Brüel&Kjær dummy head. Inside the simulator we positioned the microphone MO 2000 CU (Sennheiser, Germany) (Fig. 4b). Headphones were connected to a Babyface Pro (RME, Germany) audio interface. Using the aforementioned MRI compatible head and torso, we recorded BRIRs inside the magnet room. The MRI cooling system and cold head were temporarily turned off during recording.

The second and third auralization were created using the offline version of the room acoustic simulator RAZR [77]. The simulator combines an image source model for the specular reflections with a feedback delay network for modeling the diffuse

late reverberation. The simulation (referred to as **RAZR**) was adjusted to match the acoustical properties of the recorded room by aligning RT30, geometry as well as source and receiver position while neglecting objects inside the room. For the third auralization (referred to as **1st-order ISM**), we simplified the above RAZR model from a third-order to a first-order image source model and disabled the feedback delay network.

This procedure resulted in six different BRIRs: two distances x three room auralizations. BRIRs were peak normalized. Additionally, stimuli were equalized to account for the headphone transfer function and in the case of the recorded BRIRs, the transfer function of the loudspeaker and microphone. All stimuli were pre-generated at a sampling rate of $44.1\,\mathrm{Hz}$.

Experimental procedure

The experiment started with a training phase with feedback, which was conducted in the same position as the final experiment. During training as well as the experiment, the task was to identify the perceived distance to the presented sound, classifying it as originating from the corner of the room (**Far**), at the level of their feet (**Close**), or very close to or inside their head (**In head**).

Stimuli were presented via MRI compatible electronic in-ear headphones (Model S14, Sennheiser, Germany). Words were presented in blocks of four, in which the same word and room simulation was maintained within each block. Subjects were correctly informed that the first sound of each block was located at the level of their feet (**Close**). For the remaining sounds the distance were randomly varied for a fixed set of eight randomizations the counterbalanced distance and room auralization.. Subjects completed eight runs, with 9 trials for a total of 36 stimuli for each run. All runs were acquired in one session with short breaks between the runs. A large number of shorter runs was chosen to have a greater number of independent datasets for classification.

fMRI acquisition

The experiment was performed on a Siemens Prisma $3\,\mathrm{T}$ MRI scanner using a 64-channel head coil (Siemens Healthineers, Erlangen, Germany). We implemented a custom multi-shot 3D-EPI sequence with interleaved silent steady state imaging (ISSS, [75]). ISSS is a sparse imaging method that allows for multiple image acquisitions after a silent period used for stimulus presentation. Within the silent period, the imaging gradients are not performed to remove the most dominant, high-pitched sequence noise. At the same time, the steady-state longitudinal magnetization is maintained by continued excitation at the same rate as in the loud period, thus avoiding T1-related signal changes that would otherwise occur.

The original ISSS method used a 2D single band EPI sequence. We extended it to a multi-shot 3D-EPI sequence with blipped-CAIPI parallel imaging [76] and time-efficient water excitation [101]. Thus, we achieve whole brain coverage at a high temporal resolution of $0.57\,\mathrm{s}$ volume repetition time (TR) along with $2.5\,\mathrm{mm}$ isotropic voxel size. With 3D-EPI, the silent period can be defined by any integer multiple of the TR per shot instead of the volume TR. Here, we used $88 \times 51.81\,\mathrm{ms}$. Furthermore, 3D-EPI allowed us to use spatially non-selective water excitation (no slice-selective gradient pulses) to substantially reduce non-pitched noise contributions. Finally, extremely quite spoiler gradients were implemented in the silent periods by maximizing their duration between successive excitations (minimzing their amplitudes) while maintaining the same effective spoiling moments as in the loud periods. The final fMRI sequence parameters were: $225\,\mathrm{mm}\times225\,\mathrm{mm}\times150\,\mathrm{mm}$ field-of-view, sagittal slice orientation (phase encoding approximately along the anterior-posterior commissure line), 'WE-rect 1-2-1' water excitation [102], 16° flip angle, $1\times6_{z1}$ blipped-CAIPI parallel imaging, 7/8 partial Fourier sampling (skipping only most decayed echoes $>$ TE), 11 shots per volume including one initial "dummy shot" used for phase navigator acquisition, TE=$30\,\mathrm{ms}$.

The stimulus was presented within the silent period of $4.559\,\mathrm{s}$, which was followed by 12 fMRI image volumes in a period of $6.84\,\mathrm{s}$. In total, 542 image volumes were acquired per run, resulting in TA=8:43 min total acquisition time.
Between runs, two additional 3D-EPI images were acquired for subsequent geometric distortion correction, one with identical and one with inverted phase encoding, as well as a high-resolution T1-weighted anatomical MPRAGE image at $0.75\,\mathrm{mm}$ isotropic resolution ($240\,\mathrm{mm}\times240\,\mathrm{mm}\times192\,\mathrm{mm}$ field-of-view, sagittal slice orientation, 12° flip angle, 2-fold GRAPPA parallel imaging, TE=$2.17\,\mathrm{ms}$, TR=$2060\,\mathrm{ms}$, TI=$1040\,\mathrm{ms}$, TA=6:35 min).

Data Analysis

Acoustic Parameters
In order to quantify how the six auralizations (2 distances in 3 room auralizations) differed from one another, we analysed the acoustic parameters of the sounds presented. All acoustic parameters were computed using MATLAB 2021b [103]. RT30 was determined based on the schroeder integration method [104]. For the determination of the DRR, the direct and reverberant part of the sound were extracted based on a time-window procedure. The exact separation between direct sound and reverberation was known for the simulations. Hence, we could compute the DRR for the simulation exactly. The computation of the DRR for the recorded BRIRs required an estimate of the duration of the direct part i.e. the start of the reverberant part. The duration of the

window for the direct part was determined by aligning the DRR of the simulations with the DRR of the recorded BRIRs i.e. we varied the estimate for the start of the reverberation of the recorded BRIRs until the DRR of the recorded BRIR matched the DRR of the simulations. The duration of the direct sound was defined as $0.65\,\mathrm{ms}$ starting from the first peak of the BRIR that reached a magnitude of ten times the noise floor. Everything after $0.65\,\mathrm{ms}$ was considered reverberation. To avoid artifacts, the filters were designed to start and end with with the decreasing and increasing part of a Hann-filter, respectively. Time to peak of the Hann-filter was $2\,\mathrm{ms}$. The perceived loudness of the convolved sounds was computed according to ISO 532-2 (Moore-Glasberg) [105].

Behavioral data
The percentage of the different possible responses was examined for each room auralization. Correct responses were not calculated, as the additional possible "in head" response made an exact calculation of correct performance infeasible. Since responses were categorical, a multinomial logistic regression was used to model the response behavior. Logistic models predict the log-odds for the dependent variable based on a linear combination of the independent variables. The estimated coefficients describe how the independent variables influence the probability of the dependent variable to take a certain level, in our case **Close**, **Far** or **In head**. The regression analysis was carried out using R [106] with the R package "mclogit" [107].

Preprocessing of imaging data
Preprocessing of the imaging data was done using fmriprep. Results included in this manuscript come from preprocessing performed using fMRIPrep 23.1.3 ([108, 109] RRID:SCR_016216), based on Nipype 1.8.6 ([110, 111]; RRID:SCR_002502). The T1-weighted (T1w) images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection ([112]), distributed with ANTs ([113], RRID:SCR_004757). The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL, RRID:SCR_002823, [114]). An anatomical T1w-reference map was then computed. Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs), using brain-extracted versions of both the T1w reference and the T1w template.

The following preprocessing steps were performed for each of the 8 BOLD functional image series. First, a reference volume and its skull-stripped version were generated. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering were performed. The BOLD time-series were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as preprocessed BOLD. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration ([115]). Co-registration was configured with six degrees of freedom. Several confounding time-series measures were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS. FD was computed in two ways ([116], [117]). FD and DVARS were calculated for each functional run. Frames that exceeded a threshold of 0.9 mm FD or 1.5 standardized DVARS were annotated as motion outliers, and added as single nuisance regressor.

The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. All resamplings were performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels ([118]). Many internal operations of fMRIPrep use Nilearn 0.10.1 ([119], RRID:SCR_001362), mostly within the functional processing workflow. For further details refer to [120]. Subsequently, the images were smoothed with a kernel of $8\,\mathrm{mm}$ full-width at half maximum (FWHM) using the SPM12 toolbox from MATLAB 2021b [103].

Univariate analysis
A univariate analysis was conducted using the SPM12 (7771) toolbox with MATLAB 2021b [103]. A general linear model (GLM) was created with regressors for each of the six BRIRs. These were modeled as events time-locked to the sound onset. The Canonical Hemodynamic Response Function was used to model the hemodynamic effects. The regressors of interest were first created as if the fMRI data were collected with continuous sampling and then modified to reflect the silent periods without image data collection. Additionally, Ssx motion regressors, three for translation and three for rotation, modeled the head-movement related changes in the MRI signal. A within-subject ANOVA with multiple levels was carried out to assess the main effects at the group level.

Decoding
Three linear support vector classifiers with L1-penalty were trained on the activity maps obtained from the univariate analysis to decode each room auralization from brain activity data. We applied the decoding scheme 'FREMclassifier' of the python package nilearn [121]. As the noise within one run is autocorrelated, the activity maps were grouped according to the runs for each subject. Crossvalidation was performed, and the mean classification accuracy was calculated for each subject across

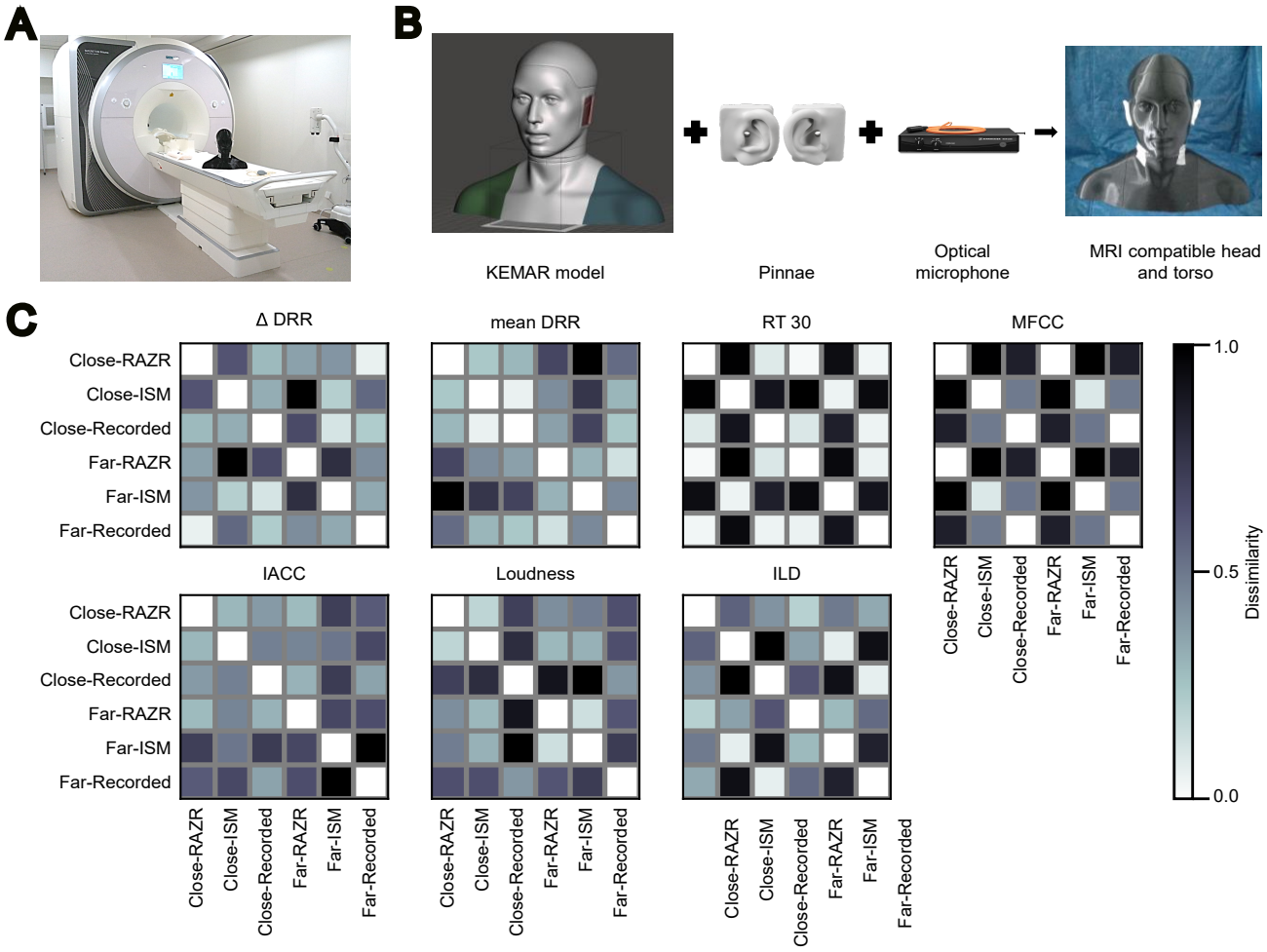

Figure 4: We conducted an RSA to examine the brain response to auralizations of the MRI scanner room that describe the acoustics with different degrees of accuracy. (A) The magnet room with the scanner and the MRI compatible head and torso simulator. (B) To record the magnet room we built an MRI compatible head and torso simulator that consisted of a KEMAR head and torso model [69], pinnae from a Brüel&Kjær dummy head [122] and the microphone MO 2000 CU from Sennheiser [123]. (C) Model RDMs based on the acoustic parameters of the stimuli.

folds. After testing for normality via a Lilliefors test (**Recorded**: $D = 0.09$, $p = 0.72$; **RAZR**: $D = 0.11$, $p = 0.53$; **1st-order ISM**: $D = 0.11$, $p = 0.49$;), we conducted a t-test to assess whether the average of these mean classification accuracies was significantly larger than chance.

Chance accuracy was defined as the average across chance levels per subject. A permutation test with $N = 1000$ was run for each subject to determine the chance accuracy at the subject level.

The searchlight analysis was based on three one-versus-rest support vector machines with L1-penalty. The support vector machines were combined in a one-versus-rest strategy for multiclass classification. The analysis was performed using the software package nilearn from python [121]. The searchlight radius was set to $5.6\,\mathrm{mm}$ and the performance was evaluated based on the classification accuracy. To determine significance at the group level, we conducted a permutation test with 5000 permutations utilizing the MATLAB toolbox SnPM [88].

Representational similarity analysis
The representational similarity analysis was conducted with the python package rsatoolbox [124]. RSA relates neuroscientific models and/or behavior to brain activity. For more details refer to [68]. In a first step, we conducted a GLM on the subject level to determine the activity map for each combination of room and distance. Subsequently, we computed the crossvalidated squared Mahalanobis distance for each pair of activity maps [125] resulting in a dissimilarity matrix (RDM, as in fig. S1c) for each subject. The matrices were averaged over all subjects to obtain one RDM characterizing the measured activity pattern.

To identify how acoustic cues relate to brain activity, the acoustic parameters that are relevant for auditory distance perception [2,3] were calculated (see *Methods Section Acoustic Parameters*). For each parameter, dissimilarity matrices were created.

The matrices based on DRR, RT30, MFCC and IACC were generated by directly computing the respective parameter for each BRIR per condition. In contrast, perceived loudness and ILD were calculated for each of the convolved sounds and these values were then averaged per condition.

The parameters RT30, IACC and perceived loudness were computed for the frequencies $250\,\mathrm{Hz}$, $500\,\mathrm{Hz}$, $750\,\mathrm{Hz}$, $1\,\mathrm{kHz}$, $1.5\,\mathrm{kHz}$ and $2\,\mathrm{kHz}$. Hence, for these parameters we obtained a vector with six values for each of the six conditions. The dissimilarity metric between conditions based on these acoustic cues was defined as one minus the Pearson correlation of the respective vectors.

For ILD, we obtained one value per condition. The dissimilarity measure was based on the Euclidean distance between the ILD values of two conditions. Regarding MFCC, dissimilarity was defined as one minus the Pearson correlation of the mel-coefficients for each BRIR. For DRR, we defined two dissimilarity metrics. First, DRR was computed for each channel resulting in two values per condition. Second, we calculated either the average DRR of the two channels (**mean DRR**) or the difference between the right and the left channel (Δ**DRR**). Finally, for both metrics we determined the dissimilarity between conditions by computing the euclidean distance between the values for each condition.

The resulting model RDMs were normalized between 0 and 1 so that the weights reflected relative contributions and not an inverse of the absolute values of the parameters themselves. Figure 4c depicts the model RDMs for each parameter.

To investigate how acoustic parameters may be related to the brain activity we computed the whitened-RDM Pearson correlation between each model RDM and the RDM based on the measured brain activity. For details regarding the correlation parameter, see [125]. The weighted models predict the data RDM as a linear weighted sum of the individual model RDMs. An optimization algorithm based on the Broyden–Fletcher–Goldfarb–Shanno procedure was applied to fit the weighted models to the data. Fit accuracy was assessed using crossvalidation. To obtain an estimate of the variance, we conducted bootstrapping with 1000 samples.

Acknowledgements

We thank Stefan Fichna, Thomas Eggert and Thomas Stephan for their comments on the analysis, Hans Hintermaier for his support with the MRI compatible head and torso and Siegfried Gündert for his assistance and the provided software concerning the measurement of the IRs. The work was supported by the Deutsche Forschungsgemeinschaft (DFG) SPP Audictive - Project ID 444827755, and by the Bundes Ministerium für Bildung und Forschung (BMBF) through the Bernstein Center Munich (Grant No. 01GQ1004A).

References

[1] A. Bergman, "Auditory scene analysis," *The perceptual organisation of sound*, 1990.

[2] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, and Psychophysics*, vol. 78, pp. 373–395, 2016.

[3] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *ACTA Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.

[4] N. Kopčo and B. G. Shinn-Cunningham, "Effect of stimulus spectrum on distance perception for nearby sources," *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1530–1541, 2011.

[5] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources. ii. localization of a broadband source," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1956–1968, 1999.

[6] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *The Journal of the Acoustical Society of America*, vol. 104, pp. 3048–3058, 11 1998.

[7] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, pp. 1465–1479, 09 1999.

[8] D. C. Greene, "Comments on "Perception of the Range of a Sound Source of Unknown Strength" [H. R. Hirsch, J. Acoust. Soc. Am. 43, 373–374 (1968)]," *The Journal of the Acoustical Society of America*, vol. 44, pp. 634–634, 08 1968.

[9] P. Cochran, J. Throop, and W. E. Simpson, "Estimation of distance of a source of sound," *The American Journal of Psychology*, vol. 81, no. 2, pp. 198–206, 1968.

[10] G. Von Békésy, "Über die entstehung der entfernungsempfindung beim hören," *Akust. Z*, vol. 3, pp. 21–31, 1938.

[11] E. A. Gamble, "Minor studies from the psychological laboratory of wellesley college: Intensity as a criterion in estimating the distance of sounds.," *Psychological Review*, vol. 16, no. 6, p. 416, 1909.

[12] D. H. Mershon and L. E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception and Psychophysics*, vol. 18, pp. 409–415, 1975.

[13] J. W. Philbeck and D. H. Mershon, "Knowledge about typical source output influences perceived auditory distance," *The Journal of the Acoustical Society of America*, vol. 111, pp. 1980–1983, 05 2002.

[14] Ö. Demirkaplan and H. Hacıhabiboğlu, "Effects of interpersonal familiarity on the auditory distance perception of level-equalized reverberant speech," *Acta Acustica*, vol. 4, no. 6, p. 26, 2020.

[15] M. G. Wisniewski, E. Mercado III, K. Gramann, and S. Makeig, "Familiarity with speech affects cortical processing of auditory distance cues and increases acuity," *PLoS One*, vol. 7, no. 7, p. e41025, 2012.

[16] W. M. Hartmann, "Localization of sound in rooms," *The Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1380–1391, 1983.

[17] A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, no. 6719, pp. 517–520, 1999.

[18] D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, "Effects of room reflectance and background noise on perceived auditory distance," *Perception*, vol. 18, no. 3, pp. 403–416, 1989.

[19] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1832–1846, 2002.

[20] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *The Journal of the Acoustical Society of America*, vol. 112, pp. 2110–2117, 10 2002.

[21] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *The Journal of the Acoustical Society of America*, vol. 124, pp. 450–461, 07 2008.

[22] A. Kolarik, S. Cirstea, and S. Pardhan, "Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues," *The Journal of the Acoustical Society of America*, vol. 134, pp. 3395–3398, 11 2013.

[23] M. Schutte, "Aspects of room acoustics, vision and motion in the human auditory perception of space," Juli 2021.

[24] B. Shinn-Cunningham, "Learning reverberation: Considerations for spatial auditory displays," *Proceedings of the 2000 International Conference on Auditory Display*, 2000.

[25] N. Kopčo, M. Schoolmaster, and B. Shinn-Cunningham, "Learning to judge distance of nearby sounds in reverberant and anechoic environments," in *Proceedings of the Joint Congress CFA/DAGA*, vol. 4, 2004.

[26] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2016.

[27] F. Klein, S. Werner, and T. Mayenfels, "Influences of training on externalization of binaural synthesis in situations of room divergence," *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 178–187, 2017.

[28] K. Brandenburg, F. Klein, A. Neidhardt, U. Sloma, and S. Werner, *Creating Auditory Illusions with Binaural Technology*, pp. 623–663. Cham: Springer International Publishing, 2020.

[29] V. Best, R. Baumgartner, M. Lavandier, P. Majdak, and N. Kopco, "Sound externalization: A review of recent research," *Trends in hearing*, vol. 24, p. 2331216520948390, 09 2020.

[30] S. Sutojo, J. Thiemann, A. Kohlrausch, and S. van de Par, *Auditory Gestalt Rules and Their Application*, pp. 33–59. Cham: Springer International Publishing, 2020.

[31] N. Kanwisher, "Functional specificity in the human brain: a window into the functional architecture of the mind," *Proceedings of the national academy of sciences*, vol. 107, no. 25, pp. 11163–11170, 2010.

[32] G. C. Thompson and A. M. Cortez, "The inability of squirrel monkeys to localize sound after unilateral ablation of auditory cortex," *Behavioural brain research*, vol. 8, no. 2, pp. 211–216, 1983.

[33] R. J. Zatorre and V. B. Penhune, "Spatial localization after excision of human auditory cortex," *Journal of Neuroscience*, vol. 21, no. 16, pp. 6321–6328, 2001.

[34] J. P. Rauschecker, B. Tian, and M. Hauser, "Processing of complex sounds in the macaque nonprimary auditory cortex," *Science*, vol. 268, no. 5207, pp. 111–114, 1995.

[35] J. Rauschecker, "Processing of complex sounds in the auditory cortex of cat, monkey, and man," *Acta Oto-Laryngologica*, vol. 117, no. sup532, pp. 34–38, 1997.

[36] J. P. Rauschecker and B. Tian, "Mechanisms and streams for processing of "what" and "where" in auditory cortex," *Proceedings of the National Academy of Sciences*, vol. 97, no. 22, pp. 11800–11806, 2000.

[37] J. P. Rauschecker, "Cortical processing of complex sounds," *Current opinion in neurobiology*, vol. 8, no. 4, pp. 516–521, 1998.

[38] J. Ahveninen, I. P. Jääskeläinen, T. Raij, G. Bonmassar, S. Devore, M. Hämäläinen, S. Levänen, F.-H. Lin, M. Sams, B. G. Shinn-Cunningham, *et al.*, "Task-modulated "what" and "where" pathways in human auditory cortex," *Proceedings of the National Academy of Sciences*, vol. 103, no. 39, pp. 14608–14613, 2006.

[39] M. Brunetti, P. Belardinelli, M. Caulo, C. DelGratta, S. Della Penna, A. Ferretti, G. Lucci, A. Moretti, V. Pizzella, A. Tartaro, K. Torquati, M. Belardinelli, and G.-L. Romani, "Human brain activation during passive listening to sounds from different locations: An fmri and meg study," *Human brain mapping*, vol. 26, pp. 251–61, 12 2005.

[40] L. Y. Deouell, A. S. Heller, R. Malach, M. D'Esposito, and R. T. Knight, "Cerebral responses to change in spatial location of unattended sounds," *Neuron*, vol. 55, no. 6, pp. 985–996, 2007.

[41] M. S. Tata and L. M. Ward, "Early phase of spatial mismatch negativity is localized to a posterior "where" auditory pathway," *Experimental Brain Research*, vol. 167, pp. 481–486, 2005.

[42] N. Kopčo, K. K. Doreswamy, S. Huang, S. Rossi, and J. Ahveninen, "Cortical auditory distance representation based on direct-to-reverberant energy ratio," *NeuroImage*, vol. 208, p. 116436, 2020.

[43] N. Kopčo, S. Huang, J. W. Belliveau, T. Raij, C. Tengshe, and J. Ahveninen, "Neuronal representations of distance in human auditory cortex," *Proceedings of the National Academy of Sciences*, vol. 109, no. 27, pp. 11019–11024, 2012.

[44] K. Krumbholz, M. Schönwiesner, D. Y. von Cramon, R. Rübsamen, N. J. Shah, K. Zilles, and G. R. Fink, "Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe," *Cerebral Cortex*, vol. 15, no. 3, pp. 317–324, 2005.

[45] J. D. Warren, B. A. Zielinski, G. G. Green, J. P. Rauschecker, and T. D. Griffiths, "Perception of sound-source motion by the human brain," *Neuron*, vol. 34, no. 1, pp. 139–148, 2002.

[46] J. Ahveninen, N. Kopčo, and I. P. Jääskeläinen, "Psychophysics and neuronal bases of sound localization in humans," *Hearing Research*, vol. 307, pp. 86–97, 2014. Human Auditory NeuroImaging.

[47] B. Grothe, M. Pecka, and D. McAlpine, "Mechanisms of sound localization in mammals," *Physiological reviews*, vol. 90, no. 3, pp. 983–1012, 2010.

[48] L. A. Jeffress, "A place theory of sound localization.," *Journal of comparative and physiological psychology*, vol. 41, no. 1, p. 35, 1948.

[49] D. McAlpine, "Creating a sense of auditory space," *The Journal of Physiology*, vol. 566, no. 1, pp. 21–28, 2005.

[50] S. A. McLaughlin, N. C. Higgins, and G. C. Stecker, "Tuning to binaural cues in human auditory cortex," *Journal of the Association for Research in Otolaryngology*, vol. 17, pp. 37–53, 2016.

[51] D. Phillips and D. Irvine, "Responses of single neurons in physiologically defined area ai of cat cerebral cortex: sensitivity to interaural intensity differences," *Hearing research*, vol. 4, no. 3-4, pp. 299–307, 1981.

[52] J. Zhang, K. T. Nakamoto, and L. M. Kitzes, "Binaural interaction revisited in the cat primary auditory cortex," *Journal of neurophysiology*, vol. 91, no. 1, pp. 101–117, 2004.

[53] P. Ungan, S. Yagcioglu, and C. Goksoy, "Differences between the n1 waves of the responses to interaural time and intensity disparities: scalp topography and dipole sources," *Clinical Neurophysiology*, vol. 112, no. 3, pp. 485–498, 2001.

[54] G. C. Stecker, S. A. McLaughlin, and N. C. Higgins, "Monaural and binaural contributions to interaural-level-difference sensitivity in human auditory cortex," *Neuroimage*, vol. 120, pp. 456–466, 2015.

[55] K. J. Palomäki, H. Tiitinen, V. Mäkinen, P. J. May, and P. Alku, "Spatial processing in human auditory cortex: the effects of 3d, itd, and ild stimulation techniques," *Cognitive brain research*, vol. 24, no. 3, pp. 364–379, 2005.

[56] I. Rapin, H. Schimmel, L. M. Tourk, N. A. Krasnegor, and C. Pollak, "Evoked responses to clicks and tones of varying intensity in waking adults," *Electroencephalography and clinical neurophysiology*, vol. 21, no. 4, pp. 335–344, 1966.

[57] U. Hegerl, J. Gallinat, and D. Mrowinski, "Intensity dependence of auditory evoked dipole source activity," *International Journal of Psychophysiology*, vol. 17, no. 1, pp. 1–13, 1994.

[58] M. Neukirch, U. Hegerl, R. Kötitz, H. Dorn, U. Gallinat, and W. Herrmann, "Comparison of the amplitude/intensity function of the auditory evoked n1m and n1 components," *Neuropsychobiology*, vol. 45, no. 1, pp. 41–48, 2002.

[59] L. Jäncke, N. Shah, S. Posse, M. Grosse-Ryuken, and H.-W. Müller-Gärtner, "Intensity coding of auditory stimuli: an fmri study," *Neuropsychologia*, vol. 36, no. 9, pp. 875–883, 1998.

[60] M. Röhl and S. Uppenkamp, "Neural coding of sound intensity and loudness in the human auditory system," *Journal of the Association for Research in Otolaryngology*, vol. 13, pp. 369–379, 2012.

[61] A. Brechmann, F. Baumgart, and H. Scheich, "Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fmri study," *Journal of Neurophysiology*, vol. 87, no. 1, pp. 423–433, 2002.

[62] H. C. Hart, D. A. Hall, and A. R. Palmer, "The sound-level-dependent growth in the extent of fmri activation in heschl's gyrus is different for low-and high-frequency tones," *Hearing research*, vol. 179, no. 1-2, pp. 104–112, 2003.

[63] H. C. Hart, A. R. Palmer, and D. A. Hall, "Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex," *Hearing research*, vol. 171, no. 1-2, pp. 177–190, 2002.

[64] C. Mulert, L. Jäger, S. Propp, S. Karch, S. Störmann, O. Pogarell, H.-J. Möller, G. Juckel, and U. Hegerl, "Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel eeg and fmri," *Neuroimage*, vol. 28, no. 1, pp. 49–58, 2005.

[65] A. Gutschalk, R. D. Patterson, A. Rupp, S. Uppenkamp, and M. Scherg, "Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex," *Neuroimage*, vol. 15, no. 1, pp. 207–216, 2002.

[66] J. M. Arend, M. Ramírez, H. R. Liesefeld, and C. Pörschmann, "Do near-field cues enhance the plausibility of non-individual binaural rendering in a dynamic multimodal virtual acoustic scene?," *Acta Acustica*, vol. 5, p. 55, 2021.

[67] H. G. Hassager, F. Gran, and T. Dau, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2992–3000, 2016.

[68] N. Kriegeskorte, M. Mur, and P. Bandettini, "Representational similarity analysis - connecting the branches of systems neuroscience," *Frontiers in Systems Neuroscience*, vol. 2, 2008.

[69] H. S. Braren and J. Fels, "A High-Resolution Head-Related Transfer Function Data Set and 3D-Scan of KEMAR," 2020.

[70] P. Rämä and S. M. Courtney, "Functional topography of working memory for face or voice identity," *Neuroimage*, vol. 24, no. 1, pp. 224–234, 2005.

[71] C. L. Grady, J. W. Van Meter, J. M. Maisog, P. Pietrini, J. Krasuski, and J. P. Rauschecker, "Attention-related modulation of activity in primary and secondary auditory cortex," *Neuroreport*, vol. 8, no. 11, pp. 2511–2516, 1997.

[72] L. Jäncke, S. Mirzazade, and N. J. Shah, "Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects," *Neuroscience letters*, vol. 266, no. 2, pp. 125–128, 1999.

[73] L. Jancke and N. Shah, "Does dichotic listening probe temporal lobe functions?," *Neurology*, vol. 58, no. 5, pp. 736–743, 2002.

[74] C. I. Petkov, X. Kang, K. Alho, O. Bertrand, E. W. Yund, and D. L. Woods, "Attentional modulation of human auditory cortex," *Nature neuroscience*, vol. 7, no. 6, pp. 658–663, 2004.

[75] C. Schwarzbauer, M. H. Davis, J. M. Rodd, and I. Johnsrude, "Interleaved silent steady state (isss) imaging: A new sparse imaging method applied to auditory fmri," *NeuroImage*, vol. 29, no. 3, pp. 774–782, 2006.

[76] R. Stirnberg and T. Stöcker, "Segmented k-space blipped-controlled aliasing in parallel imaging for high spatiotemporal resolution epi," *Magnetic resonance in medicine*, vol. 85, no. 3, pp. 1540–1551, 2021.

[77] T. Wendt, S. van de Par, and S. D. Ewert, "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *J. Audio Eng. Soc*, vol. 62, no. 11, pp. 748–766, 2014.

[78] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[79] J. Catic, S. Santurette, and T. Dau, "The role of reverberation-related binaural cues in the externalization of speech," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 1154–1167, 2015.

[80] T. Leclère, M. Lavandier, and F. Perrin, "On the externalization of sound sources with headphones without reference to a real source," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2309–2320, 2019.

[81] F. E. Toole, "In-head localization of acoustic images," *The Journal of the Acoustical Society of America*, vol. 48, no. 4B, pp. 943–949, 1970.

[82] R. A. Butler, E. T. Levy, and W. D. Neff, "Apparent distance of sounds recorded in echoic and anechoic chambers.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 6, no. 4, p. 745, 1980.

[83] J. Catic, S. Santurette, J. M. Buchholz, F. Gran, and T. Dau, "The effect of interaural-level-difference fluctuations on the externalization of sound," *The Journal of the Acoustical Society of America*, vol. 134, pp. 1232–1241, 2013.

[84] R. S. Dewey, D. A. Hall, C. J. Plack, and S. T. Francis, "Comparison of continuous sampling with active noise cancelation and sparse sampling for cortical and subcortical auditory functional mri," *Magnetic resonance in medicine*, vol. 86, no. 5, pp. 2577–2588, 2021.

[85] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in Cognitive Sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[86] J.-D. Haynes, "Decoding mental states from brain activity in humans," *Nature Reviews Neuroscience*, vol. 7, no. 7, 2006.

[87] K. N. Stevens, *Sources of Inter- and Intra-Speaker Variability in the Acoustic Properties of Speech Sounds*, pp. 206–232. Berlin, Boston: De Gruyter Mouton, 1972.

[88] "Chapter 46 - nonparametric permutation tests for functional neuroimaging," in *Human Brain Function (Second Edition)* (R. S. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, J. T. Ashburner, and W. D. Penny, eds.), pp. 887–910, Burlington: Academic Press, second edition ed., 2004.

[89] J. Xu, H. Dong, N. Li, Z. Wang, F. Guo, J. Wei, and J. Dang, "Weighted rsa: An improved framework on the perception of audio-visual affective speech in left insula and superior temporal gyrus," *Neuroscience*, vol. 469, pp. 46–58, 2021.

[90] P. D. Coleman, "An analysis of cues to auditory depth perception in free space.," *Psychological bulletin*, vol. 60, no. 3, p. 302, 1963.

[91] N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-based functional brain mapping," *Proceedings of the National Academy of Sciences*, vol. 103, no. 10, pp. 3863–3868, 2006.

[92] S. P. Gandhi, D. J. Heeger, and G. M. Boynton, "Spatial attention affects brain activity in human primary visual cortex," *Proceedings of the National Academy of Sciences*, vol. 96, no. 6, pp. 3314–3319, 1999.

[93] D. C. Somers, A. M. Dale, A. E. Seiffert, and R. B. Tootell, "Functional mri reveals spatially specific attentional modulation in human primary visual cortex," *Proceedings of the National Academy of Sciences*, vol. 96, no. 4, pp. 1663–1668, 1999.

[94] J. J. Foster and S. Ling, "Feature-based attention multiplicatively scales the fmri-bold contrast-response function," *Journal of Neuroscience*, vol. 42, no. 36, pp. 6894–6906, 2022.

[95] A. Bronkhorst, "Effect of stimulus properties on auditory distance perception in rooms," *Physiological and Psychological Bases of Auditory Function*, pp. 184–191, 2001.

[96] H.-Y. Kim, Y. Suzuki, S. Takane, and T. Sone, "Control of auditory distance perception based on the auditory parallax model," *Applied Acoustics*, vol. 62, no. 3, pp. 245–270, 2001.

[97] T. C. T. Yin, P. H. Smith, and P. X. Joris, "Neural mechanisms of binaural processing in the auditory brainstem," Sept. 2019.

[98] R. J. Zatorre and P. Belin, "Spectral and temporal processing in human auditory cortex," *Cerebral cortex*, vol. 11, no. 10, pp. 946–953, 2001.

[99] C. M. Wessinger, J. VanMeter, B. Tian, J. Van Lare, J. Pekar, and J. P. Rauschecker, "Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging," *Journal of cognitive neuroscience*, vol. 13, no. 1, pp. 1–7, 2001.

[100] S. Quadflieg, A. Mohr, H.-J. Mentzel, W. H. Miltner, and T. Straube, "Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia," *Biological Psychology*, vol. 78, no. 2, pp. 129–137, 2008.

[101] R. Stirnberg, W. Huijbers, D. Brenner, B. A. Poser, M. Breteler, and T. Stöcker, "Rapid whole-brain resting-state fMRI at 3 Tesla: Efficiency-optimized three-dimensional EPI versus repetition time-matched simultaneous-multi-slice EPI," *NeuroImage*, vol. 163, no. August, pp. 81–92, 2017.

[102] O. Norbeck, T. Sprenger, E. Avventi, H. Rydén, A. Kits, J. Berglund, and S. Skare, "Optimizing 3D EPI for rapid T 1 -weighted imaging," *Magnetic Resonance in Medicine*, pp. mrm.28222–mrm.28222, Feb. 2020.

[103] T. M. Inc., "Matlab version: 9.13.0 (r2021b)," 2017.

[104] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 6_Supplement, pp. 1187–1188, 1965.

[105] B. C. Moore and B. R. Glasberg, "A revision of zwicker's loudness model," *Acta Acustica united with Acustica*, vol. 82, no. 2, pp. 335–345, 1996.

[106] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.

[107] M. Elff, *mclogit: Multinomial Logit Models, with or without Random Effects or Overdispersion*, 2022. R package version 0.9.6.

[108] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, M. Snyder, *et al.*, "fmriprep: a robust preprocessing pipeline for functional mri," *Nature methods*, vol. 16, no. 1, pp. 111–116, 2019.

[109] O. Esteban, R. Blair, C. J. Markiewicz, S. L. Berleant, C. Moodie, F. Ma, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, *et al.*, "fmriprep," *Software*, 2018.

[110] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh, "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python," *Frontiers in neuroinformatics*, vol. 5, p. 12318, 2011.

[111] K. Gorgolewski, O. Esteban, C. Markiewicz, E. Ziegler, D. Ellis, M. Notter, D. Jarecka, *et al.*, "Nipype," *URL https://doi.org/10.5281/ZENODO*, vol. 6834519, 2018.

[112] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4itk: improved n3 bias correction," *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[113] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[114] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE transactions on medical imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[115] D. N. Greve and B. Fischl, "Accurate and robust brain image alignment using boundary-based registration," *Neuroimage*, vol. 48, no. 1, pp. 63–72, 2009.

[116] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fmri," *Neuroimage*, vol. 84, pp. 320–341, 2014.

[117] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.

[118] C. Lanczos, "Evaluation of noisy data," *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, vol. 1, no. 1, pp. 76–85, 1964.

[119] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Frontiers in neuroinformatics*, vol. 8, p. 71792, 2014.

[120] O. Esteban, R. Blair, C. J. Markiewicz, S. L. Berleant, C. Moodie, F. Ma, A. I. Isik, A. Erramuzpe, M. Goncalves, R. A. Poldrack, and K. J. Gorgolewski, "poldracklab/fmriprep: 1.0.0-rc5," Sept. 2017.

[121] N. contributors, "nilearn."

[122] "Symmetrical pinnea |dz-9770 + dz-9769 |brüel and kjær." Accessed: 2024-05-07.

[123] T. Ford, "Sennheiser mo 2000 industrial optical analog microphone." Accessed: 2010-05-07.

[124] Z. Lu and Y. Ku, "Neurora: A python toolbox of representational analysis from multi-modal neural data," *Frontiers in neuroinformatics*, vol. 14, p. 563669, 2020.

[125] J. Diedrichsen and N. Kriegeskorte, "Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis," *PLoS computational biology*, vol. 13, no. 4, p. e1005508, 2017.

[126] P. Reinhart, P. E. Souza, N. K. Srinivasan, and F. J. Gallun, "Effects of reverberation and compression on consonant identification in individuals with hearing impairment," *Ear and Amp; Hearing*, vol. 37, pp. 144–152, 2016.

[127] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low-and high-frequency sound levels as measures of acoustical quality in concert halls," *The Journal of the Acoustical Society of America*, vol. 98, no. 2, pp. 988–1007, 1995.

[128] J. You and J. Y. Jeon, "Just noticeable differences in sound quality metrics for refrigerator noise," *Noise Control Engineering Journal - NOISE CONTR ENG J*, vol. 56, 11 2008.

Supplementary Material

Acoustic Analysis of Stimuli

This supplement provides more detailed information on how the acoustic analysis of the stimuli used in this study was performed. In addition, the analysis will show how acoustic cues vary across the different distances used in the experiments, as well as across the three different auralizations that were used (**Recorded, RAZR, 1st-order ISM**). Modifying the complexity of the room auralization changed the acoustic cues in the studied dimensions distance and room auralization.

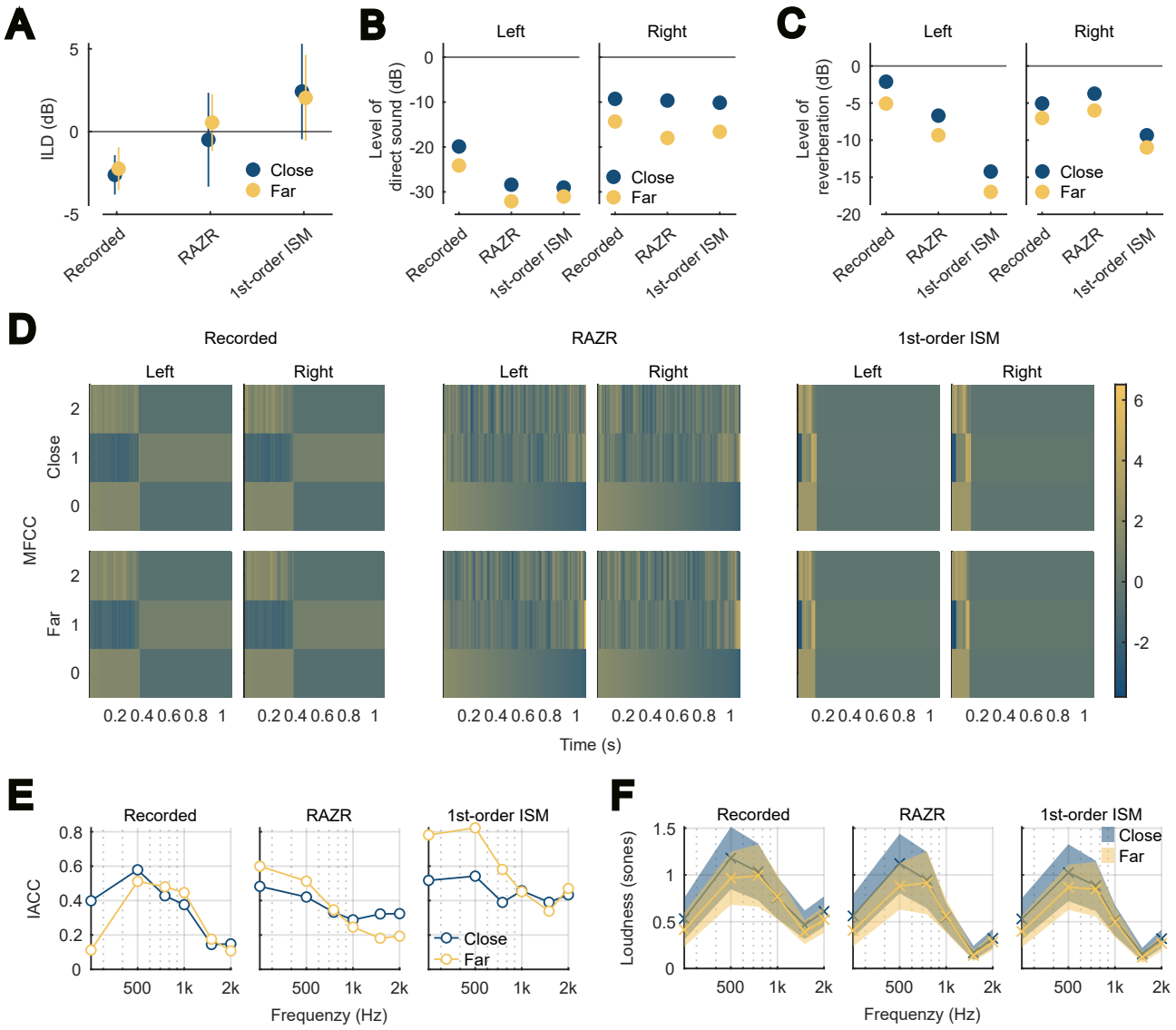

Figure S1: Acoustic analysis of the stimuli in the two dimensions auralization and distance. Stimuli differ mainly between auralizations and less between distances. (A) The ILD averaged over sounds per condition. Errorbars display standard deviation from the mean. (B-D) The level of the direct sound and the reverberation as well as the MFCC were calculated for the left and the right channel of each BRIR per condition. (E) The IACC between the left and the right channel of the BRIRs was determined for each condition. (F) The predicted loudness averaged over sounds per condition. Errors represent standard deviation from the mean.

Reverberation time, RT30

The reverberation time, RT30, of the recorded room ranges between $0.4\,\mathrm{s}$ and $0.45\,\mathrm{s}$ for audible frequencies and matches the RT30 of a typical office or residential building [126].

While RT30 was in a similar range for the recorded BRIRs and the full **RAZR** simulation, we observed a decrease for the **1st-order ISM**. This decrease can be attributed to the missing late reverberation in the **1st-order ISM**.
Between the distances RT30 was found to be highly similar.

Direct-to-reverberant ratio, DRR
In general, the direct-to-reverberant ratio, DRR is larger for the right channel than the left channel as sound sources are simulated on the right side of the subject.
As expected DRR is smaller for sounds located further away from the listener. However, with an average of $2.8 \pm 2.2\,\mathrm{dB}$ DRR differences between distances are below JNDs [20, 21].
Between auralizations differences in DRR are larger with an average of $4.5 \pm 2.1\,\mathrm{dB}$ and a maximum differences of $9\,\mathrm{dB}$ between the **1st-order ISM** and **RAZR**. This increase in DRR for the **1st-order ISM** is expected due to omitting part of the reverberant tail.
Additionally, DRR of the recorded BRIRs is slightly larger than for the full RAZR simulation. We designed the full RAZR simulation such that it models the acoustics of the recorded room as accurate as possible but neglects objects inside the room. We hypothesize that the MRI scanner introduces additional very early reflections and thereby causes the observed change in DRR. Note also that the subject was positioned such that there was a direct acoustic path from source to the right ear while the left ear was in the acoustic shadow of the head. For that reason the direct sound component was more pronounced in the right ear leading to generally higher DRR values.

Interaural Level Differences, ILDs
Figure S1a visualizes the average interaural level difference, ILD, for all sounds presented per condition. We observed an increase in ILDs with increased complexity of the auralization.
Surprisingly, we found a negative ILD for the recorded BRIRs although sounds are presented on the right side.
Plotting the level of the direct sound and the level of the early reflections and reverberation separately suggests that this observation was caused by early reflections and reverberation. As expected the level of the direct sound was larger for the right channel than the left channel (fig S1b).
As demonstrated in fig. S1c, the same applies to the reverberation level for the two simulations. However, in the case of the recorded BRIRs the level of reverberation is larger for the left than for the right channel. We hypothesize that this effect is caused by reflections and diffraction at the scanner bore which was not included in the room acoustics simulations.

MFCC
The MFCCs are a perceptually relevant representation of the spectral shape of an acoustic stimulus, it is closely related to timbre of a sound. The successive components, represent the spectral shape on a MEL-frequency scale in terms of Discrete Cosine Transform. The first coefficient represents overall level, the second, the strength of mid-frequency reduction of spectral energy. The coefficients provide an orthogonal decomposition of the MEL-frequency spectrum, with ever increasing detail with increasing coefficient number. MFCCs have been used extensively to represent the acoustic spectrum in a compact manner to allow pattern classifiers to recognize speakers identity or to recognize spoken words.
A comparison of the conditions revealed a high similarity between the MFCCs of the two distances (Fig. S1d). This observation is supported by the RDM constructed based on the MFCCs (Fig. 4c. The dissimilarity values are small for comparisons between the same room but different distances.
Between the auralizations we observed similar values for the first coefficient but considerable differences for the higher order coefficient. Hence, the data suggests that variations in level are small while the spectral composition differs between the auralizations.

IACC
The IACC is a parameter that measures the similarity of the acoustic waveforms arriving at both ears. Specifically, in reverberant conditions, the IACC will drop in value due to incoherent reflections that arrive from different directions. It has been shown that the IACC correlates to perceived source width. In the evaluation of concert-halls this parameter predicts Apparent Source Width and allows predicting preferences of concert halls [127].
As can be expected IACC is increased for the **1st-order ISM** in fig. S1e compared to Recorded and RAZR.

Perceived Loudness
With reduced auralization accuracy, we observed a slight decrease in average loudness (Fig. S1f). As expected, loudness was elevated for condition **Far** compared to **Close**. However, in both dimensions differences were below JNDs [128].

GLM

cluster ID	X (mm)	Y (mm)	Z (mm)	peak t statistic	volume (mm^3)
1	-56.5	-17.5	4.0	21.5	38687
2	61.0	-7.5	-1.0	20.1	38859
3	-6.5	5.0	54.0	12.5	17906
4	31.0	27.5	-1.0	10.9	4453
5	-31.5	25.0	4.0	10.6	3390
6	-39.0	-22.5	54.0	10.1	17218
7	48.5	2.5	49.0	9.4	3968
8	43.5	15.0	26.5	8.6	4250
9	6.0	-20.0	-1.0	8.1	6125

Table S1: Data of the significant clusters obtained with the GLM. The statistical map has been thresholded at $p = 0.05$ FWE. Only cluster with more than 100 voxels are considered significant. Coordinates are given in MNI125 space.

Searchlight MVPA

cluster ID	X (mm)	Y (mm)	Z (mm)	peak pseudo-t	volume (mm^3)
1	-64.0	-43.5	14.0	6.8	1750

Table S2: Data of the significant clusters obtained after the searchlight MVPA. The statistical map has been thresholded at $p = 0.05$ FWE and only clusters with more than 50 voxels were considered significant. Coordinates are given in MNI125 space.

Searchlight RSA

cluster ID	X (mm)	Y (mm)	Z (mm)	peak t statistic	volume (mm^3)
1	-69.0	-15.0	-13.5	7.1	3453
2	-69.0	-45.0	11.5	7.0	5312
3	61.0	-12.5	-13.5	6.2	2546
4	38.5	-50.0	24.0	5.7	625

Table S3: Data of the significant clusters obtained with the searchlight RSA. The statistical map has been thresholded at $p = 0.05$ FWE. Coordinates are given in MNI125 space.