Audio Engineering Society

Convention Paper 10710

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Modeling the perceptibility of intra-room acoustic variations

Florian Klein[1], Lukas Treybig[1], Sebastià Amengual Garí[2], and Stephan Werner[1]

[1]*Technische Universität Ilmenau, Electronic Media Technology Group, Ilmenau, Germany*
[2]*Reality Labs Research, Meta, Redmond, USA*

Correspondence should be addressed to Florian Klein (florian.klein@tu-ilmenau.de)

ABSTRACT

Acoustic augmented realities aim to fuse real and virtual acoustic environments perceptually. The fundamental question is how accurately the virtual acoustics needs to match the real acoustics. In 6DoF scenarios, listeners can explore spaces and thus intra-room acoustic variations become of interest. This contribution presents an approach to predict the perceptibility of room acoustic variations. By utilizing an autonomous mobile robot platform dense room acoustic measurements were collected. Based on this data, listening tests were conducted with the sole purpose of matching the perceptual ratings with the room acoustic difference between positions within the same room. The perceptual ratings in the form of binary decisions (the difference is perceptible or not) and attribute ratings describing these differences are fed into a generalized mixed-effects model together with pre-calculated differences of room acoustic parameters. The results show, that we can create promising models based on the available data. However, the perceptibility in walking scenarios may differ significantly from static comparisons between positions. An approach to assess such cases is presented for discussion. Furthermore, we discuss the generalizability of the derived models, future improvements, and limitations.

1 Introduction

The exploration of room acoustics and their perceptual impact is a significant area of research within the field of audio engineering and psychoacoustics. This is all the more relevant when it involves acoustically augmented environments in which virtual audio objects are to be seamlessly integrated into the real room acoustics. This requires a position-dynamic binaural synthesis that is always adapted to the real acoustic environment to create plausible Auditory Augmented Reality (AAR) [1, 2].

Acoustic environments can vary in their local room acoustic properties. Especially in small rooms, significant acoustic variations are expected depending on position and regarding room modes, patterns of first reflections, and early to late energy ratios. Small, but perceivable differences could influence localization and distance perception and bigger differences could cause externalization issues [3, 4].

It is therefore desirable to quantify these differences in terms of their perceptual relevance to optimize the computational effort of adapting the position-dynamic binaural synthesis to the local acoustic environment. The prediction of when a perceptual relevance exists should ideally be based on acoustic parameters that can be estimated in real time.

This paper will present an approach to model the perceptibility of acoustic changes in an AAR scenario.

Listening tests are conducted based on acoustic measurements and the results are used for prediction. The choices about the measurements and the listening test design are discussed. A specific modeling approach is explained and preliminary results are presented.

2 Related Work

The International Standard ISO 3382-2009 defines several room acoustical parameters and links them to corresponding perceptual features with associated Just Noticeable Differences (JND) [5, 6]. However, several studies have shown that such JNDs are not absolute, but can in fact depend on several factors, such as signal content, signal level, or room conditions, among others [7, 8]. Additionally, even JNDs for the most common acoustical parameter, Reverberation Time (RT60), have been found to differ in various studies using different experimental setups [9, 10, 11, 12]. Recent work implementing a two-dimensional threshold test explored whether JNDs of seemingly perceptually orthogonal perceptual aspects i.e. Direct-to-Reverberant Ratio (DRR) and RT60 is in fact orthogonal or there is in fact an influence of their absolute values on each others' JND [13].

While a granular characterization of the JND of each acoustic parameter could contribute to a deeper understanding of room acoustical perception, the advent of audio augmented reality (AAR) technologies requires a practical approach to produce transfer-plausible acoustics, in which virtual sources can be seamlessly inserted into real soundscapes [14, 15, 2, 16]. For instance, discrimination between stimuli with small acoustical differences becomes more complicated when signals with different content are used [17]. While listeners are consistently able to identify whether recordings are made in the same room [18], even when clearly audible acoustical differences between different positions in a room are present, listeners are generally not able to correctly identify the actual positions of the recordings [13]. This suggests that AAR systems can indeed significantly relax the rendering constraints in certain situations.

In order to apply these learnings systematically, AAR systems can leverage the use of perceptual models, which could predict whether the acoustics of two positions or two rooms are perceptually identical, and thus lead to the reduction of computational demands of a renderer. In [18], a room acoustical similarity model was developed from experimental tests based on paired comparisons between measured rooms. The model was based on a Gaussian Process Regression (GPR) which mapped the values of a Factor Analysis (FA) of several monaural room acoustical parameters with the results of the perceptual test.

In the present work, we introduce a model that aims at predicting the perceived acoustical similarity between different positions in the same room.

3 Data Acquisition

To create a model that describes the perceptibility of acoustic changes, measurement data as well as the corresponding perceptual ratings are needed. The following sections briefly describe the utilized measurements, how they were collected as well as the design of the listening test based on this data.

3.1 Acoustic Data

Acoustic measurements from an open-source database were used [19, 20]. The measurements were conducted by an autonomous mobile robot in an acoustically dry room that meets the acoustic requirements of the ITU-R BS.1116-3. The room was modified by sequentially adding highly reflective walls to change the patterns of the first reflections and the reverberation time. The full dataset includes nine wall configurations, five loudspeakers of type Geithain RL906 as sound sources at several positions in the room and orientation. A measurement grid with a uniform spacing of 0.25 m was choosen to achieve a high spatial density with a manageable measurement effort. For the measurements of seven-channel spatial impulse responses (SRIRs), a custom-built, circular and open microphone array was chosen, which is suitable for the application of the spatial decomposition method (SDM) [21]. The array consists of one high-quality microphone in the center of the array and six satellite microphones for the estimation of the direction of arrival (DOA). For this publication, a subset of three room configurations were selected from the dataset. Fig. 1 shows the selected configurations. The data from the room in the middle is used in the listening test. The other rooms are only used in the prediction. The mean broadband reverberation time T_{20} with standard deviation over all measured source-receiver positions of the room configurations are 0.25 s $\pm$ 0.01 s (one wall), 0.60 s $\pm$ 0.11 s (three walls), and 1.05 s $\pm$ 0.04 s (five walls).

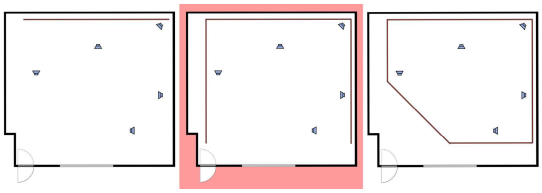


Fig. 1: Data from the room in the middle (marked red) is used for the listening test. Data from all shown rooms are used for prediction analysis. One (left), three (middle), and five (right) reflective walls (red lines) are placed inside an acoustically dry room. Figure from [20].

In order to create audio samples, listening positions have to be defined. The choice of positions influences the features the model will utilize. All positions have a distance of 1.5 m to the source to exclude all distance-related DRR (Direct-to-Reverberation energy Ratio) changes and to keep the level of direct sound constant. Loudness and DRR changes are strong distance cues that might distract from the acoustic properties of interest. Because the recordings were conducted with directional speakers, we decided to choose listening positions with different relative angles. By changing relative position, reflection energy will arrive from different spatial angles, which possibly is a strong cue in small rooms. All chosen positions are shown in Fig. 2. For each speaker, we select a frontal position (at $0°$, colored in orange), side position (at $30°$, colored in blue), far-side position (at $90°$, colored in yellow), and back position ($180°$, colored in green). In the listening test, only positions of the same color are compared to each other. This is because changes in the frequency spectrum resulting from the speaker directivity could potentially mask room acoustic properties.

3.2 Listening Test

For the auralization a publicly available SDM toolbox [22] and the preprocessed KEMAR HRTFs from [23] are used to calculate binaural room impulse responses (BRIRs) from the seven-channel SRIR recordings. For each position, the BRIR was chosen which places the virtual sound source in front of the listener. That means, that independent of the room configuration, speaker orientation, or listening position the direct sound always arrives from the same direction. Again, this was necessary to avoid localization shifts of the direct sound

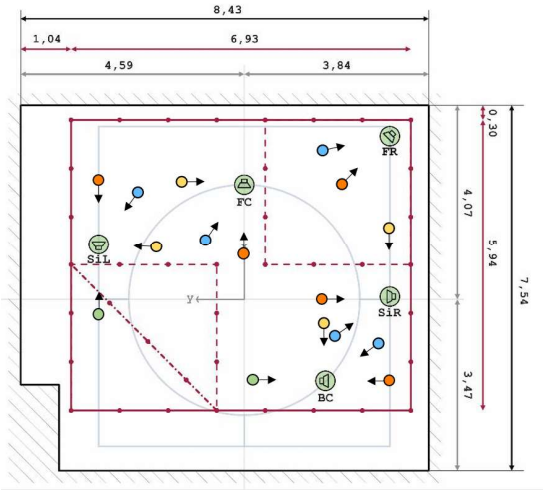


Fig. 2: Floor plan, room configurations, speaker positions and listening positions used in the listening test. For each speaker (SiL, FC, FR, SiR, BC), listening positions in the same relative direction and distance are defined. Frontal positions in orange, Side positions in blue, Far-side positions in yellow, and back positions in green. Figure modified from [20].

masking room acoustic differences. BRIRs were convoluted with anechoic recordings of a voice and a guitar signal with a length of 6 s.

For data collection, we decided on an ABX test design. This test procedure features an open reference and two samples, where one of these samples is equivalent to the open reference. Usually, the task is to identify which of the samples equals the open reference. ABX-tests are typically used for assessing the detection ability of subtle differences between two stimuli [24].

An example ABX trial in the current listening test may be constructed in the following way: The open reference is the yellow position facing speaker SiL and the available choices A and B are playing the yellow position facing speaker BC and again the yellow position facing speaker SiL. Participants had the task to identify the stimulus (A or B) that sounds *different* to the open reference.

All positions within a listening test are selected from the room with three walls, shown in the middle of Fig. 1.

For this room configuration 92 ABX trials need to be rated (46 unique combinations and one repetition each).

For each trial, there was an additional task: Participants should select attributes explaining the perceived difference (if applicable) and rate them on a 6-point Likert scale. The attributes were: *Coloration*, *Loudness*, *Distance*, *Direction*, *Reverberance*, *Externalization*, *Apparent source width*, *Ease of localization*. All attributes, a group or just one attribute can be selected according to their relevance. For example, if only a small difference was perceivable and this could only be accounted to coloration, participants should only rate this attribute with a 1. If they, for example, perceive a strong difference in direction and reverberance, they rate both attributes with a 4.

The test is conducted in the same room as for the SRIR measurements but without reflective walls. A head-pose dynamic binaural headphone auralization was available and each participant conducted 92 trials with either the guitar or the voice signal. This test design is expected to be valid in terms of externalization and localization as well as the detection of small differences.

4 Modelling Approach

The central idea of our approach is shown in Fig. 3. For each room combination, pre-selected source-receiver combinations with their corresponding measurements are available.

First, differences between each of these combinations are calculated concerning their room acoustic parameters or frequency spectra. Depending on the number of parameters or frequency bands, dimensionality reduction approaches (such as PCA) may be needed.

Second, perceptual differences for the same differences of source-receiver combinations are collected by means of listening tests. From the ABX test we collect binary data whether a difference was perceivable or not. This information can be used for the estimation of just noticeable differences (JNDs). The attribute ratings deliver information about the type of differences on a category scale.

Third, acoustic and perceptual differences are combined to build a prediction model. The model can be used to predict the attributes or to derive a discrimination threshold. Based on this, iso-perceptive areas can be derived. These are areas within a room that are perpetually indistinguishable concerning the room acoustics.

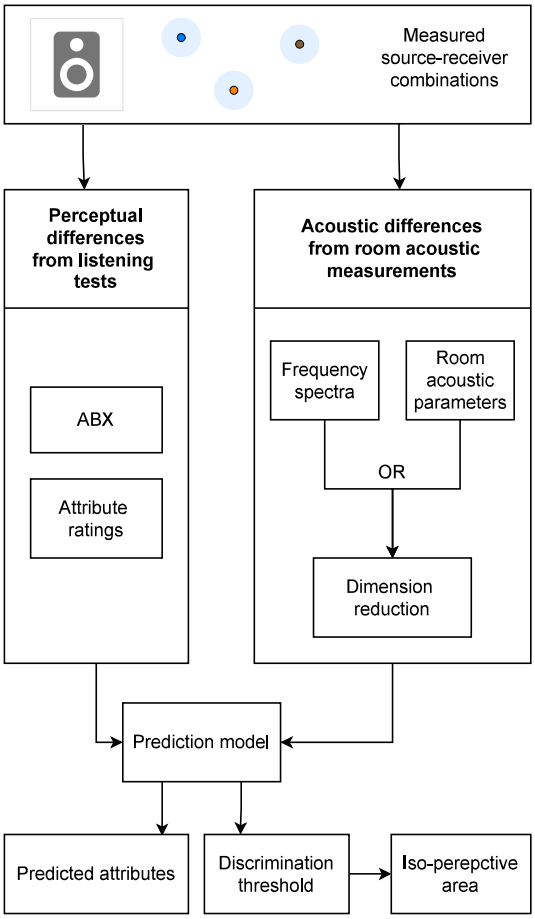


Fig. 3: General workflow showing the steps towards the prediction of "iso-perceptive areas".

4.1 Exemplary Results

The listening test was conducted by 10 participants for the voice signal an 7 participants for the guitar signal. This results in a total of 1564 ratings for each attribute: 920 for the voice signal and 644 for the guitar signal. With the exception of one woman, all participants were male and between the age of 24 to 42. All participants have experience with such listening tests and can be described as expert listeners.

The approach to be described can be used to create prediction models for various use cases, such as predicting the hit accuracy, the probability or strength of

perceiving a difference of a certain attribute, or the probability of perceiving any difference describable by attributes. No model could be generated for hit accuracy based on the ratings, as only 23 of the 1564 ratings were incorrect. This results in a detection rate of 98.5%. This shows that most differences were above the just noticeable thresholds. In order to be able to predict the general perceived similarity between two positions, the approach was chosen in which the attribute ratings are considered in combination.

Attribute selection: In this example, we aim to create a model based on combined attribute ratings. For this purpose, we include the ratings for the attributes of coloration, loudness, distance, direction, reverberance, and apparent source width. These attributes were selected because they are directly or indirectly related to the room acoustic parameters. This way, a model is created which predicts the general perceived acoustical similarity.

The ratings of the attributes are summarized by the root mean square according to the Eq. 1, so that a score value is obtained for each set of attribute ratings.

$$\text{score}_k = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{rating}_k(i)^2)} \tag{1}$$

where k is the index of a participant's rating, i is the number of the rated attribute of n attributes. The root mean square was selected to better distinguish between many small ratings and one large rating.

Acoustic feature selection: Absolute difference of the broadband room acoustic parameters early decay time EDT, definition D_{50}, clarity index C_{80} and C_{50}, direct-to-reverberation energy ratio DRR and center time T_S are calculated based on the room impulse response (RIR) derived from the omnidirectional measurement microphone of the microphone array. For more details and the calculation of these parameters refer to [5, 25]. Instead of the reverberation time according to DIN EN ISO 3382-1/2 [5], the decay time DT_{20} according to [26] is used.

Prediction model: The prediction model is built using a generalized linear mixed-effects model (GLMM) for linear regression. For further details about GLMM we refer to [27, 28]. The absolute parameter differences are independently included in the model as fixed effects. The signal type is also considered as a fixed effect. The participant ID is used as a random effect. This should compensate for the different scale usage of the participants. The following model function[1] can be obtained:

$$\begin{aligned}\text{score} \sim \text{signal} + EDT_{|a-b|} + DT20_{|a-b|} + \\ D50_{|a-b|} + C50_{|a-b|} + C80_{|a-b|} + \\ DRR_{|a-b|} + T_{S|a-b|} + (1|\text{participant})\end{aligned} \tag{2}$$

where $|a-b|$ is the absolute difference of the corresponding acoustic parameter related to the comparison of positions. Score is the predicted attribute, which in this case is the attribute combination according to Eq. (1).

The generated model contains a total of 1564 observations with 9 fixed effects coefficients, 17 random effects coefficients and a covariance parameter of 2. A normal distribution of the data was assumed and the *restricted maximum pseudo likelihood* method was used as a fit method.

According to the model coefficients, there is no significant difference in this model in terms of signal type. The estimated intercept coefficient is given as 1.06 and a standard deviation of 0.1. For the prediction, this means that if the same stimuli were compared, the parameter differences would all be 0 and the prediction result would be approximately 1. If we assume that the model outputs the value 0 for the prediction of an identical stimulus, the intercept must be subtracted from the predicted value.

Finally, a brief examination of the quality of the model is presented. The adjusted R-squared (R^2_{adj}) is used as a quality measure to evaluate the models or the regression function. It indicates the proportion of variance that is explained and also takes into account the sample size [29]. A value of R^2_{adj} between 0.35 and 0.6 seems to indicates a good prediction. Higher values may suggest overfitting or other issues [30]. For the generated model, the R^2_{adj} is 0.37.

Iso-perceptive areas: The next step is to create iso-perceptive areas using the prediction model. For this purpose, threshold values are defined to create the categories *no/very small*, *small*, *medium* or *strong differences* (see Table 1). This categorization only serves for

[1] Formula notation according to Matlab function fitglme.

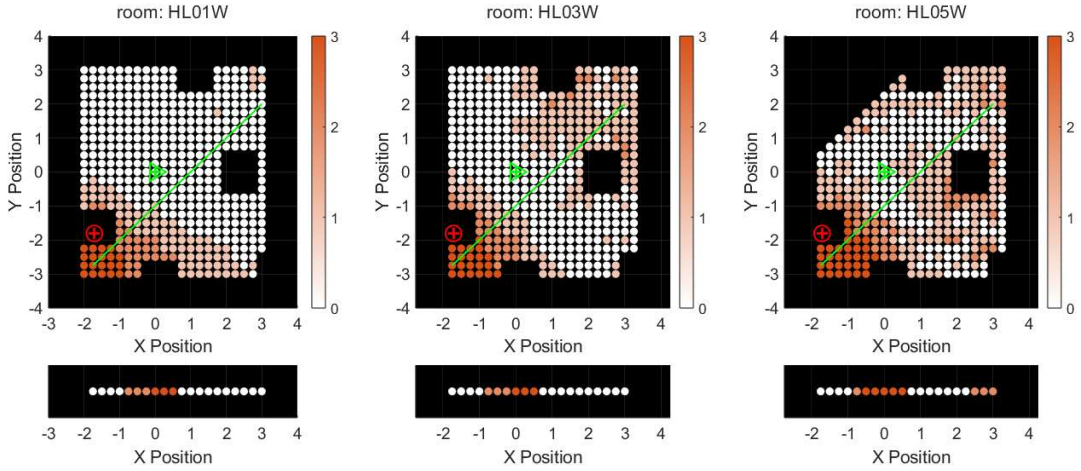


Fig. 4: Grid: Prediction of similar perceptual areas in relation to the reference listening position $P(0,0)$ (green triangle); **Line:** prediction of perceived differences along the green line in the grid (from $P(-1.75,-2.75)$ to $P(3.00,2.00)$); Considerations for loudspeaker BC (red circle with red cross) and voice signal; Categories of Table 1 are color-coded.

Table 1: Categories (y) related to thresholds (x) with associated descriptions

y	Thresholds	Perceived difference
0	$x \leq 0.5$	no/very small
1	$0.5 < x \leq 1$	small
2	$1 < x \leq 2$	medium
3	$x > 2$	strong

an easier interpretation of the data and is not directly related to any just noticeable differences.

Fig. 4 displays predictions with thresholds based on the room configurations depicted in Fig.1 and for the voice signal. Two different ways of predicting perceptual differences are presented in the graphics. On the one hand the grid with the green triangle, on the other hand the green line in relation to the lower graphic.

First, consider the grid in relation to the green triangle located at point $P(0,0)$. The point $P(0,0)$ is compared to every other point in the grid with respect to the loudspeaker BC. This can be equated to a virtual jump from position $P(0,0)$ to any other arbitrary position. This results in areas in which the model predicts *no/very small*, *small*, *medium* or *strong differences* in perception in relation to $P(0,0)$.

In all three room configurations, it can be seen that the model predicts stronger perceptual differences when $P(0,0)$ is compared to positions within the direct sound range of the loudspeaker (see also Fig. 2). In HL01W (one reflective wall), the model predicts slight differences in the area around $P(2.5,2.5)$. Compared to HL01W, the changes in this area become more severe for HL03W (three reflective walls) and HL05W (five reflective walls).

Please note that the displayed grid is dependent on the room, source position, and reference listener position. If a different reference position is selected, the grid predictions will change accordingly. Similarly, if a different source is considered while keeping the listener's position constant, the predictions will also change. This specifically covers the effects that would also occur in a real listening environment.

The lower plots in Fig. 4 refer to the green line in the grid. Here, the model's prediction is based on the difference to the next point along this line. This corresponds to a movement or small jumps along the green line in the direction from $P(-1.75,-2.75)$ to $P(3.00,2.00)$. A gradual change better represents the experience in

position-dynamic applications than a jump between positions. The range of P_X from -1.75 to 0.75 shows a similar progression across all room configurations. This is due to the dominant orientation of the loudspeaker. The predicted differences between the points in this range become stronger up to $P_X = 0.75$. After this point, the model does not predict any further differences. Only in room HL05W noticeable differences are predicted at the end of the line when reaching the corner of the room. Room acoustic variations in the middle of the room are likely not perceivable when walking along the line. The predictions of perceptual differences along these lines need verification with a specially tailored listening test.

5 Discussion and Summary

Data acquisition: Data for modeling was collected via acoustic measurements and listening tests. On the upside, the measurement environment and their modifications were strongly controlled. The acoustic changes by inserting walls led to clearly audible changes which offered good variations for our modeling approach. On the downside, acoustic properties of the floor and the ceiling were always the same.

For the listening test, we had to pick certain positions to create a feasible test. We believe that our approach of keeping distance and relative angle constant was essential to create a model that focuses on room acoustic properties instead of loudness and coloration of the direct sound.

ABX tests only make sense when the difference between the stimuli is rather small, and to build a model from binary data, the acoustic differences must be close to the barely perceptible difference. This criterion is not met in our listening test and the three-wall configuration (detection rate over 98%), and therefore modeling was only possible on the attribute ratings.

Parameter selection: The acoustic features currently used can not fully explain the perceptual attributes of the listening test. For example, strong first reflections can alter the perceived direction, but none of the parameters address this directly. However, the effects of those reflections may be indirectly included in the energy ratios in some constellations (such as C_{50}). In the case of externalization, this is even more obvious, because no acoustic parameter is known that exclusively correlates with it. Coloration is also not covered directly because currently only broadband parameters are used. The simple answer may be to include more parameters and their band-wise values, but with GLMM collinearity can become an issue. However, dropping strongly correlating parameters can not be easily generalized, since the correlation of parameters depends also on the room of interest. Dimension reduction may be a solution to this problem. However, it is important to consider a few factors: Firstly, differences can no longer be attributed directly to differences in room acoustic parameters, but only to defined distances in the components. Secondly, if the parameters are standardized beforehand, they must be comprehensive enough. Additionally, it is important to note that the dimensions retrieved by such reduction methods are based on statistics and not on their perceptual relevance.

Modeling: It is obvious that the example given cannot provide a universally valid prediction at this stage due to the limited number of participants and the need to reconsider the choice of parameters. However, the authors suggest that it is worthwhile to explore this modeling approach further.

The proposed approach can be used to model the perceptibility of intra-room acoustic changes based on classic room acoustic parameters. Perceptual models can be utilized to validate 6DoF AAR renders. Depending on the requirements of the AAR renders, such models can help to find the optimal balance between computational efficiency and perceptual accuracy.

6 Future Work and Outlook

Currently, an online version of the listening test is under development. This will allow us to increase the participant count significantly and gives us the opportunity to test multiple room configurations as well as audio signals. We expect, that the number of correct identifications will be lower, because of the non-expert listeners and because of lower intra-room acoustic variations in the other room configurations. This will enable modeling based on the binary data, too. Additionally, the modeling approach could be extended by incorporating auditory models. Currently, acoustic parameters derived from RIRs are used, but the listening test was conducted using stimuli convolved by BRIRs. Incorporating BRIR based estimates of externalization [31] and coloration [32] in the model may improve prediction performance. In addition, the predictions along

the line in Fig. 4 are to be validated by a further study. The inclusion of signal-type properties in the model is a useful extension, as it allows for content-dependent predictions to be made.

7 Acknowledgments

The presented work is funded within the projekts: *"ISOPERARE"*, funded by Reality Labs Research at Meta and *"Multiparties"* in the framework "KMU-innovativ: Interactive Technologies for Health and Quality of Life" of the Federal Ministry of Education and Research, Germany

References

[1] Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K., "A summary on acoustic room divergence and its effect on externalization of auditory events," in *8th International Conference on Quality of Multimedia Experience (QoMEX), Portugal*, 2016, doi:10.1109/QoMEX.2016.7498973.

[2] Neidhardt, A., Schneiderwind, C., and Klein, F., "Perceptual matching of room acoustics for auditory augmented reality in small rooms - literature review and theoretical framework," *Trends in Hearing*, 26, 2022, doi:10.1177/23312165221092919.

[3] Hodgson, M., Scannell, L., and Higgins, J., "Acoustics of small rooms," *The Journal of the Acoustical Society of America*, 137(3), pp. 1609–1610, 2015, ISSN 0001-4966, doi:10.1121/1.4913275.

[4] Brandenburg, K., Klein, F., Neidhardt, A., Sloma, U., and Werner, S., *Creating auditory illusions with binaural technology*, pp. 623–663, Springer International Publishing, Cham, 2020, ISBN 978-3-030-00386-9, doi:10.1007/978-3-030-00386-9_21.

[5] DIN EN ISO 3382-1:2009-10, "Akustik - Messung von Parametern der Raumakustik - Teil 1: Aufführungsräume (ISO 3382-1:2009); German version EN ISO 3382-1:2009," 2009-10.

[6] DIN EN ISO 3382-2:2008-09, "Akustik - Messung von Parametern der Raumakustik - Teil 2: Nachhallzeit in gewöhnlichen Räumen (ISO 3382-2:2008); German version EN ISO 3382-2:2008," 2008-09.

[7] Martellotta, F., "The just noticeable difference of center time and clarity index in large reverberant spaces," *The J. of the Acoust. Soc. of America*, 128(2), pp. 654–663, 2010, ISSN 0001-4966, doi:10.1121/1.3455837.

[8] Klein, F., Treybig, L., Schneiderwind, C., Werner, S., and Sporer, T., "Just noticeable reverberation difference at varying loudness levels," in *Audio Eng. Soc. Conv. 154*, 2023.

[9] Seraphim, H.-P., "Untersuchungen über die Unterschiedsschwelle exponentiellen Abklingens von Rauschbandimpulsen," *Acta Acustica united with Acustica*, 8(4), pp. 280–284, 1958.

[10] Blevins, M. G., Buck, A., Peng, Z. E., and Wang, L., "Quantifying the just noticeable difference of reverberation time with band-limited noise centered around 1000 Hz using a transformed up-down adaptive method," in *Proceedings of the International Symposium on Room Acoustics*, Toronto, Canada, 2013.

[11] Frissen, I., Katz, B. F., and Guastavino, C., "Effect of sound source stimuli on the perception of reverberation in large volumes," in *International Symposium on Computer Music Modeling and Retrieval*, pp. 358–376, Springer, 2009.

[12] Klein, F., Gari, S. V. A., Arend, J. M., and Robinson, P. W., "Towards determining thresholds for room divergence: A pilot study on detection thresholds," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–7, IEEE, 2021.

[13] Meyer-Kahlen, N., Garí, S. V. A., Ananthabhotla, I., and Calamia, P., "A two-dimensional threshold test for reverberation time and direct-to-reverberant ratio," in *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–8, 2023, doi:10.1109/I3DA57090.2023.10289365.

[14] Wirler, S. A., Meyer-Kahlen, N., and Schlecht, S. J., "Towards transfer-plausibility for evaluating mixed reality audio in complex scenes," in *AES Conf.: 2020 AES Int. Conf. on Audio for Virtual and Augmented Reality*, 2020.

[15] Meyer-Kahlen, N., Schlecht, S. J., and Lokki, T., "Clearly audible room acoustical differences may

not reveal where you are in a room," *The J. of the Acoust. Soc. of America*, 152(2), pp. 877–887, 2022, doi:10.1121/10.0013364.

[16] Meyer-Kahlen, N., Amengual Garí, S., McKenzie, T., J. Schlecht, S., and Lokki, T., "Transfer-plausibility of binaural rendering with different real-world references," *Fortschritte der Akustik - DAGA 2022*, 2022.

[17] McKenzie, T., Meyer-Kahlen, N., and Schlecht, S. J., "The role of source signal similarity in distinguishing between different positions in a room," in *AES Conf.: 2023 AES Int. Conf. on Spatial and Immersive Audio*, Audio Engineering Society, United States, 2023.

[18] Helmholz, H., Ananthabhotla, I., Calamia, P. T., and Amengual Gari, S. V., "Towards the prediction of perceived room acoustical similarity," in *AES Conf.: 2022 AES Int. Conf. on Audio for Virtual and Augmented Reality*, 2022.

[19] Treybig, L., Klein, F., Stolz, G., Werner, S., and Amengual Garí, S., "A high spatial resolution dataset of spatial room impulse responses for different acoustic room configurations," *Fortschritte der Akustik - DAGA 2024*, 2024.

[20] Treybig, L., Klein, F., Stolz, G., and Werner, S., "A high spatial resolution dataset of spatial room impulse responses for different acoustic room configurations," 2024, doi:10.5281/zenodo.10450779.

[21] Tervo, S., Pätynen, A., J.and Kuusinen, and Lokki, T., "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, 61(1/2), pp. 16–27, 2013.

[22] Amengual Garí, S., Brimijoin, W. O., Hassager, H. G., and Robinson, P., "Flexible binaural resynthesis of room impulse responses for augmented reality research," in *EAA SASP*, 2019, doi:10.25836/sasp.2019.31.

[23] Braren, H. and Fels, J., "A high-resolution head-related transfer function data set and 3D-scan of KEMAR," 2020, doi:10.18154/RWTH-2020-11307.

[24] Clark, D. L., "Ten years of A/B/X Testing," in *91st AES Convention*, NY, USA, 1991.

[25] Kuttruff, H., *Room acoustics*, Spoon Press, 5 edition, 2009, ISBN 978-0-203-87637-4.

[26] Treybig, L., Werner, S., Klein, F., and Gari, S. V., Amengual Garí, "Robust reverberation time estimation for audio augmented reality applications," in *Audio Eng. Soc. Conv. 154*, 2023.

[27] Field, A., Miles, J., and Field, Z., *Discovering statistics using R*, SAGE Publications Ltd, 2012, ISBN 978-1-4462-0046-9.

[28] Agresti, A., *An introduction to categorical data analysis*, John Wiley & Sons, 2. edition, 2007, ISBN 978-0-471-22618-5.

[29] Backhaus, K., Erichson, B., Plinke, W., and Weiber, R., *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, Springer, Berlin, 15. edition, 2018, ISBN 978-3-662-56654-1.

[30] Ozili, P. K., "The acceptable R-square in empirical modelling for social science research," Social Research Methodology and Publishing Results, 2022, doi:http://dx.doi.org/10.2139/ssrn.4128165.

[31] Baumgartner, R. and Majdak, P., "Decision making in auditory externalization perception: model predictions for static conditions," *Acta Acust.*, 5, p. 59, 2021, doi:10.1051/aacus/2021053.

[32] McKenzie, T., Armstrong, C., Ward, L., Murphy, D. T., and Kearney, G., "Predicting the colouration between binaural signals," *Applied Sciences*, 12(5), 2022, ISSN 2076-3417, doi:10.3390/app12052441.