Audio Engineering Society

Convention Paper 10700

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.

Towards Perceptual Modeling in Room Acoustics - Connecting Acoustic Parameters with Auditory Perception

Lukas Treybig[1], Ulrike Sloma[2], Stephan Werner[1], Juliane Höbel-Müller[3], and Andreas Nürnberger[3]

[1]*Technische Universität Ilmenau, Electronic Media Technology Group, Ilmenau, Germany*
[2]*Brandenburg Labs GmbH, Ilmenau, Germany*
[3]*Otto von Guericke Universität Magdeburg, Data and Knowledge Engineering Group, Magdeburg, Germany*

Correspondence should be addressed to Lukas Treybig (lukas.treybig@tu-ilmenau.de)

ABSTRACT

The prediction of spatial auditory perception from acoustic differences between the real listening environment and virtual audio is of vital importance to make audio augmented reality applications valid. Therefore we investigate whether perceptual mismatch can be quantitatively predicted using descriptions of parametric differences and if such predictions can be generalized across different listening environments. This contribution examines the ability of statistical models to predict auditory perceptual differences based on acoustic parameters of auditory perception, focusing on *envelopment* and *plausible fit*. Acoustic parameters are utilized from measured binaural room impulse responses of seven rooms with diverse acoustic properties. These parameters are correlated with subjective data from listening tests from three different rooms. The correlation analysis approach using a generalized linear mixed model methodically shows that auditory perception can be predicted from acoustic parameters with a certain degree of room independence. This is especially true for the feature *envelopment*, while the prediction of *plausible fit* was more strongly linked to the specific characteristics of the listening room. Additionally, the study presents limitations, such as the small sample size and the number of validated models. The paper contributes to the field of room acoustics by proposing, applying, and analyzing a methodology for predicting auditory-perceptive differences. It also highlights the challenges of achieving room-independent predictions.

1 Introduction

A key challenge in so-called auditory mixed reality is the plausible embedding of virtual objects in the real environment [1, 2]. When synthesizing and reproducing the virtual audio objects, the room acoustics of the real room must be captured and reproduced for every possible source-receiver relation in the room. The aim is the perceptual fusion of the virtual objects with the real acoustic environment [3, 4]. If the virtual acoustics differ from the real acoustics, the virtual audio objects may not be perceived as being in the room, but rather inside the listener's head. This can result in a reduced fusion of the virtual objects and the real listening room [5, 6] which can be interpreted as a lack of plausibility [7, 8]. Quantifying the subjective experience of sound in a room related to the acoustics is a challenging and complex topic [9].

Based on the work of Helmholz et al. [10] regarding the prediction of perceived acoustic similarity this paper presents an approach that aims to bridge the gap between measurable acoustic parameters and the per-

ceptual features of *envelopment* and cognitive fit between the virtual audio objects and the real listening environment, called *plausible fit*.
The research seeks to answer the questions: *Can a perceptual difference in room acoustics be modeled and predicted on the basis of parametric room acoustic differences?* and *Is this prediction independent of the room in which the listening test is performed?*
The research aims to distill the essence of acoustic perception into a set of predictive models using statistical methods, including generalized linear mixed models (GLMMs). The perceptual models are generated based on listening test data from three different listening rooms and acoustic parameters calculated from a comprehensive set of measured binaural room impulse responses (BRIRs). While acknowledging its limitations, this paper presents a path forward for predictive modeling of acoustic perception. This sets the stage for further exploration and refinement in this area of research.

2 Room Acoustic Measurements

The study utilized seven acoustically distinct rooms to auralize a five-channel surround loudspeaker setup through dynamic binaural synthesis. The setup meets the ITU-R BS.2159 [11] requirements with a 2m radius. For each loudspeaker and room, BRIRs were measured using a KEMAR dummy head on a turntable, with 5° resolution for the horizontal plane. Table 1 provides an overview of the rooms, and additional details can be found also in [12].
The acoustic parameters are calculated from the measured BRIRs. This is done for both the broadband frequency spectrum and octave bands. The calculations are performed separately for each BRIR and then averaged over the left and right ear, as well as the orientations of the dummy head and the five loudspeakers. The following nine acoustic parameters are considered: early decay time EDT, definition D_{50}, clarity index C_{80} and C_{50}, direct-to-reverberation energy ratio DRR, center time T_S, and relative strength G_{rel}([9]). For more details and the calculation of these parameters refer to [13, 14]. Instead of the reverberation time according to DIN EN ISO 3382-1/2 [13], the decay time DT_{20} according to [15] is used. The estimation of DT_{20} is less dependent on the source-receiver relation, the non omnidirectional characteristic of the dummy head, and the orientation of the dummy head compared to T_{20}. The acoustic parameters are the input variables to a GLMM for linear regression used for the auditory perception prediction (see Section 5). In order to be able to apply the later prediction also on the basis of one-channel room impulse responses, only these acoustic parameters are used.

Table 1: Rooms used for auralization; marked bold rooms are also used as listening rooms; T_{30} = reverberation time.

Abbreviation	Room	T_{30} (s)
HL	**Listening Lab**	**0.21**
H1539b	Test Lab	0.28
LR001	Living Room	0.46
ML2-102	**Conference Room**	**0.52**
HU103	Lecture Hall Audimax	0.65
H2505	**Seminar Room**	**1.20**
H1562	Foyer Helmholtz-Building	3.47

3 Listening Test

This section describes the technical setup, the design and the procedure of the listening.

3.1 Technical Setup

The measurements of the BRIRs for each rooms named in Table 1 are used for the head-pose dynamic auralization of the virtual loudspeakers using binaural synthesis via headphones. Listening tests were conducted in the rooms HL, H2505 and ML2-102. The loudspeaker setup used for the measurements is also used for the listening test as the hidden and real reference. To auralize the virtual sources without affecting the perception of the real loudspeakers the extra-aural headphones BK211 were used [16]. For each listening test, the signal chain was calibrated once by the investigators so that the real loudspeaker and its virtualized auralization over headphones had the same perceived loudness. Horizontal head movement was allowed during the listening test and the head was tracked by a HTC Vive tracker attached to the headphones. The participant's rating was captured via a graphical user interface presented via a monitor in front of the person using a mouse and keyboard. The height was chosen so that no vertical head movement was required.
The **audio signals** used were two single channel signals, a saxophone played once on the left and once on the right speaker. Also a complex music mix for all

five speakers simultaneously was used, to evaluate the use case of playing music mixes over headphones. In this study, we only look at the saxophone signal.

3.2 Design

The test follows a multiple stimulus approach, with adaptation to the ITU-R BS.2132-0 [17]. All stimulus conditions are presented simultaneously so that the participant has the opportunity to compare and rate the stimuli directly in terms of the asked attribute. The stimuli in the test are the seven auralized sources over headphones based on the measured BRIRs, and the hidden reference over the loudspeakers. One presentation of all stimuli conditions for one audio signal is called trial. All trials are presented twice to each participant per attribute. The trials to be evaluated are individually randomized, and also the stimuli in each trial.
The listening test comprises the evaluation of five **attributes**. These are *Overall Audio Quality, Envelopment, Coloration*, and *Plausible Fit*. A distinct continuous scale was used for each attribute, each marked with seven discrete points. Further the attribute *Externalization* was assessed questioning the listener if the audio signal is perceived outside the head and localizable. Since the paper focuses on *Envelopment* and *Plausible Fit* only these and their scales are described in more detail.

- **Envelopment:** Impression of being surrounded by the reproduced sound field or audio signal. *"Not surrounded at all", "Less surrounded", "Slightly less surrounded", "Medium surrounded", "Slightly more surrounded", "More surrounded", "Completely surrounded"*
- **Plausible Fit:** Impression of how well the heard audio signal fits into your current listening environment, considering room acoustic characteristics. *"Extremely bad fit", "Bad fit", "Poor fit", "Fair fit", "Good fit", "Excellent fit", "Ideal fit"*

3.3 Procedure

The listening test consists of two parts of attribute evaluation, where the order of the first one is fixed and the order of the second part is randomized per listener. The listening test procedure is shown in Fig. 1. The two parts took place on different days.
To give the listeners a first insight about the audio, to achieve an approximately similar starting point for all, and to let them adapt to the current listening room a familiarization was performed. For that, three test audio signals were presented in randomized order - once played back over the loudspeakers and once auralized via headphones based on the measured BRIRs of the current listening room. In addition, a short training phase was conducted right before the evaluation of each attribute. Each attribute was evaluated once with each auralized audio signal, allowing the listener to become familiar with the graphical user interface and the question being asked, and to ask questions as they arise.

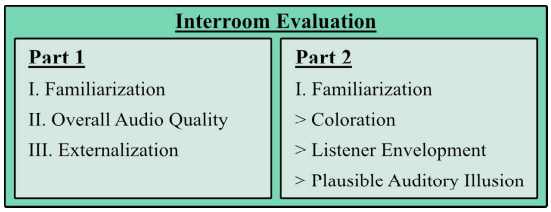


Fig. 1: Procedure of the listening test.

The listening test was done in a total of three different rooms, which are also part of the BRIR measures. In the beginning, two acoustically different rooms were selected as listening rooms: the listening lab (HL) and a reverberant seminar room (H2505). The order of the rooms differed for the participants. The evaluation of both parts should be completed in one room before the listener performs the test in the other. After first analysis of the two rooms the same test was conducted in a third room, the conference room (ML2-102).
Nine **participants** took part in the listening test in the first two rooms, five female and four male. The average age was 33.6 years in the range from 25 to 40. In the third room ten participants took part, three female and seven male. The average age was 33.1 years in the range from 25 to 41. Seven of the persons took part in all three evaluations. All participants have experience with such listening tests. One participant was not considered for the *envelopment* results, due to a different understanding of the attribute. Within a listening test, each participant submitted eight ratings within one trial, with each trial presented twice, as noted above. For instance, if there are eight participants and the signal is coming from two different directions, with seven rooms and one hidden reference (real loudspeaker), and one repetition, this will result in 256 ratings for *envelopment*.

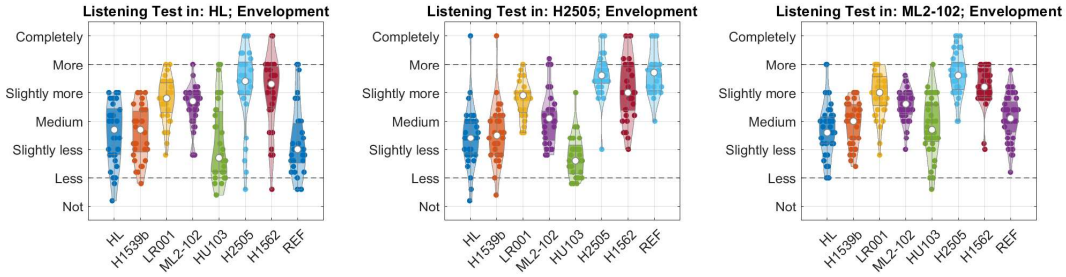


Fig. 2: Envelopment rating distribution for listening test in: **left:** HL, **center:** H2505, **right:** ML2-102; x-axis: Room name corresponds to the auralized rooms via headphones; REF corresponds to the real loudspeakers.

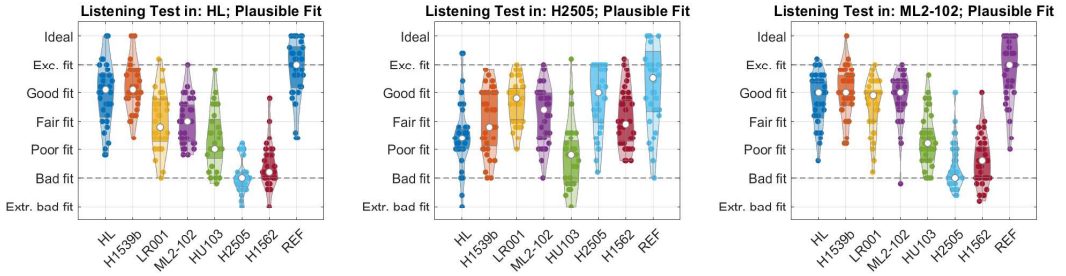


Fig. 3: Plausible fit rating distribution for listening test in: **left:** HL, **center:** H2505, **right:** ML2-102; x-axis: Room name corresponds to the auralized rooms via headphones; REF corresponds to the real loudspeakers.

4 Listening Test Analysis

In the following, the results are evaluated using the example of the single channel saxophone signals and the attributes *envelopment* and *plausible fit*. Based on the results, the modeling is performed in Section 5.
The ratings can initially be displayed directly using violin plots. Fig. 2 and 3 show these for the *envelopment* and *plausible fit* ratings for each listening room.
The spread of the individual ratings is very wide in some cases. This indicates, among other things, that the scale ranges were used to varying degrees by the test subjects or that their personal preferences are very different. The evaluation of the *envelopment* shows that the real loudspeaker reference (designated as REF in Fig. 2) and the virtual version have similar results. The evaluation of the virtual sources is also similar in the three different listening rooms. From this it can be concluded that the evaluation of the *envelopment* seems to be independent of the listening room. The assessment of the *plausible fit* differs. The real speakers (designated as REF in Fig. 3) were predominantly rated with a very high fit. However, there is also a high level of variation here. The virtualized version in the individual rooms is slightly worse on average than the reference. The other signals presented are more or less plausible depending on the listening room.

4.1 Generalized Linear Mixed Model

To provide some more statistical evidence we looked at mixed statistical models. We use generalized linear mixed models (GLMMs). In a mixed model, fixed and random effects can be considered in one model. Fixed effects have a fixed coefficient and are constant for the observations. Random effects, on the other hand, allow differences between individuals to be taken into account. For further details on that refer to [18, 19].
In this case, the fixed effects are the labels of the rooms reproduced. The random effects are the individual assessments of the participants. This should compensate for the different scale usage of the participants. Fig. 4 shows the predictions of the models based on the listening room HL for *envelopment* and *plausible fit*. The prediction scores for the reference speaker is given, which corresponds to the actual evaluation scale (0 to

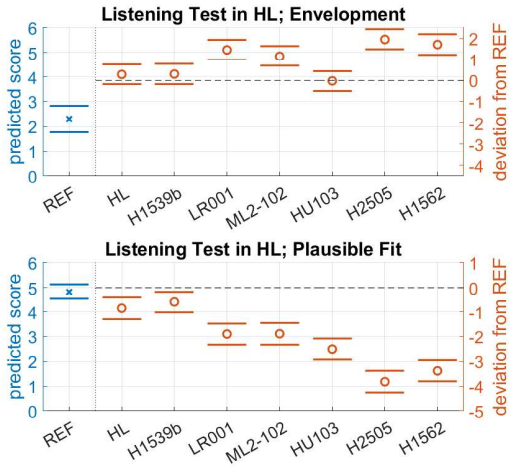


Fig. 4: Predictions of the model based on the listening room HL for **upper:** Envelopment and **lower:** Plausible fit; **blue:** perceptual prediction of the loudspeaker reference with 95% confidence interval; **red:** Predicted deviation from the reference with 95% confidence interval.

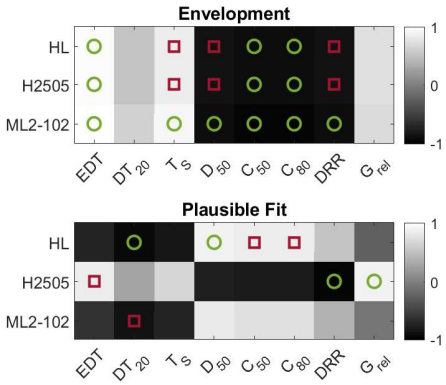


Fig. 5: Correlations of the broadband room acoustic parameters of the reproduced rooms with the medians of the corresponding model prediction based on the three listening tests; **green circle:** $P_{adj} \leq 0.1$; **red square:** $0.1 < P_{adj} \leq 0.2$.

6). The deviation to the reference is given for the other rooms. The confidence intervals are shown next to the mean value. The mean of the deviation from the reference is the basis for the next observations.

4.2 Correlation

In this step, it is verified whether the ratings predicted on the basis of the GLMM correlate with the acoustic parameters or other distance measures of the rooms heard. As the data is ordinal, the Spearman's rank correlation is used. In addition, within a listening test, the p-value is adjusted with the Bonferroni correction and denoted as P_{adj}. The implementation is based on the following idea: *If an acoustic parameter or a defined distance measure correlates at a similarly high level and significant in all three listening tests, this value can be used to draw conclusions about perception.*

As only one fixed value per parameter and listening room can be assumed in this model, these were averaged over all measurements in one room. These values were then correlated with the mean prediction values of the rooms heard[1].

Fig. 5 shows the correlation results for the broadband parameters in the individual listening rooms for *envelopment* and *plausible fit*. In addition, the significance level for the adjusted p-values are shown. For the correlation with the *envelopment* evaluation, it can be seen that the acoustic parameters, with the exception of the reverberation time, correlate highly or inversely highly and significant. In addition, the correlation is similar in all three listening rooms. From this it can be concluded that, on the basis of the correlating broadband room acoustic parameters, the generation of models which can predict the perception of the *envelopment* independently of the listening room is possible. This is discussed in more detail in Section 5.

The correlation results are different in terms of *plausible fit*. It can be seen that the broadband parameters correlate similarly with the listening test data for the HL and ML2-102 and are inverted for the H2505. This means that a listening room independent model based on the individual broadband parameters is not possible in this case.

Based on these results, other similarity measures based on room acoustic parameters were explored to find possible combinations for a *plausible fit* prediction independent of the listening room. For this purpose, principal component analysis (PCA) as unsupervised and linear discriminant analysis (LDA) as supervised method could be used. These are two methods of multivariate statistics for transforming high-dimensional data into low-dimensional spaces. A description of the

[1] compared to Fig. 4 deviation from REF

methods can be found in [20, 21]. In terms of acoustic parameters, this can be used to compress the most important information from a large number of parameters into fewer components. In this paper, the PCA and LDA were applied to the standardized octave band parameters in the range of 500Hz to 4kHz, similar to [12, 22]. The first three components of the PCA and the LDA are considered below, which together explain 94.36%[2] of the variance in the PCA and 98.34%[3] in the LDA.

The following similarity measures were calculated based on the mean values of the components. Thereby, only the distance of the spatial mean values within the first three components and the cosine similarity between the first and second as well as the first and third component are considered here as similarity measures. The similarity measures are always calculated between the listening room and the reproduced rooms.

The distance within the first three components is calculated according to:

$$d_C = \sqrt{\sum_{n=1}^{3} \left(\overline{C}_{n,R_{ref}} - \overline{C}_{n,R_i}\right)^2}, \tag{1}$$

where $\overline{C}_{n,R_{ref}}$ is the mean of the nth component of the listening room and $\overline{C}_{n,R_i}$ is the mean of the nth component of the room being reproduced. With the same definitions, the cosine similarity is calculated according to:

$$CS_{C_{1k}} = \frac{\sum_{n=1}^{k} \overline{C}_{n,R_{ref}} \overline{C}_{n,R_i}}{\sqrt{\sum_{n=1}^{k} \overline{C}^2_{n,R_{ref}}} \sqrt{\sum_{n=1}^{k} \overline{C}^2_{n,R_i}}}, \tag{2}$$

where k is only two or three in the following. The components of the PCA are labeled C and those of the LDA LD. Fig. 6 shows the correlation results for the described similarity measures for the PCA over the individual listening rooms for *envelopment* and *plausible fit*. For *envelopment*, it is highly correlated in HL and H2505, but inverted. For the listening test in ML2-102, there is a tendency as in HL, but very low and not significant. For *plausible fit*, the same significant correlation values are shown for HL and H2505. However, there is no correlation with ML2-102. It follows that the similarity features based on PCA cannot be used as prediction parameters independent of the listening room in this case. In comparison to Fig. 6 and Fig. 7 show the correlation results based on the LDA.

[2] C_1: 72.67%; C_2: 16.7%, C_3: 4.99%

[3] LD_1: 92.23%; LD_2: 4.55%, LD_3: 1.59%

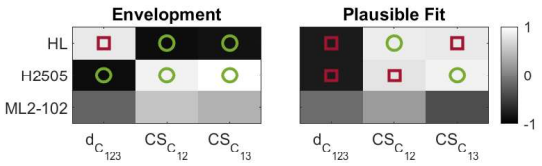


Fig. 6: Correlations of the used similarity measures based on PCA of the reproduced rooms with the medians of the corresponding model prediction based on the three listening tests; **green circle:** $P_{adj} \leq 0.1$; **red square:** $0.1 < P_{adj} \leq 0.3$.

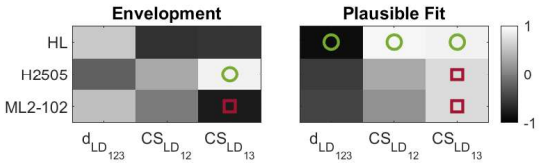


Fig. 7: Correlations of the used similarity measures based on LDA of the reproduced rooms with the medians of the corresponding model prediction based on the three listening tests; **green circle:** $P_{adj} \leq 0.1$; **red square:** $0.1 < P_{adj} \leq 0.3$.

For *envelopment*, the cosine similarity for component one and three correlates with H2505 and inversely with ML2-102, but not with HL. The other distances do not correlate either.

For *plausible fit*, all similarity measures correlate for the HL listening room. For the other listening rooms, only a similar trend as for HL can be recognized, but it is clearly lower. Only the cosine similarity for components one and three correlates higher in all three rooms. From this it is concluded that $CS_{LD_{1,3}}$ is suitable for predicting the evaluation of the *plausible fit*.

5 GLMM for Perception Prediction

Based on the results from Section 4.2, models for predicting *envelopment* and *plausible fit* based on the same listening test in the three different rooms are now to be generated and analyzed.

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 [20]. A value of R^2_{adj} between 0.35 and 0.6 indicate a good prediction. Higher values may suggest overfitting or other

issues [23]. Other measures would be the F-statistic or the standard error, but these are not considered separately at this point.

Envelopment: In the Fig. 5 it can be seen that the room acoustic parameters correlate with the mean prediction values of the GLMM based on the reproduced rooms. Using this knowledge, a GLMM is generated with the fixed effect being the broadband differences of the room acoustic parameters between the listening room (LR) and the reproduced room (RR), and the participant set as the random effect. The model function[4] used as an example is:

$$\begin{aligned} Envelopment \sim -1 + EDT_{LR-RR} + T_{S_{LR-RR}} + \\ C_{50_{LR-RR}} + DRR_{LR-RR} + (1|participant) \end{aligned} \quad (3)$$

where -1 indicates that there is no intercept. The choice of parameters is based on the correlation shown in Fig.5. However, other different types of parameters should be included in the model as well. Using the model function according to Eq. (3), a GLMM was generated for each listening test. For the prediction of the *envelopment*, the parameter differences between the rooms were calculated. Each room was defined as a listening room and a reproduced room.

Fig. 8 shows the prediction results of the three models, based on the listening tests in the rooms HL, H2505 and ML2-102, in the upper part. First the diagonals of the predictions are considered. If the listening room is the same as the room heard, then the perceived *envelopment* should also be the same, and the model predicts a value of 0. When comparing with other predictions, a value greater than 0 indicates a stronger envelopment, while a value less than 0 indicates a weaker *envelopment* compared to the listening room. The magnitude of the difference depends on the deviation from 0.

For example, suppose the participant is in room LR001 (LR) and a signal is reproduced over headphones with the acoustics of H2505 (RR). In this case, the models predict that the participant would feel more enveloped by the signal from RR relative to LR. On the other hand, if HL (RR) is used in LR001 (LR), the models predict that the participant would feel less enveloped by the signal from RR relative to LR.

Comparing the predictions of the models, it can be seen that all three models give similar predictions. This is also shown by R^2_{adj} (see Fig. 8), which are all in the same range. The lower row matrices show the corresponding confidence intervals. These are smaller in the model based on the ML2-102 listening test, as it contains the ratings of more participants.

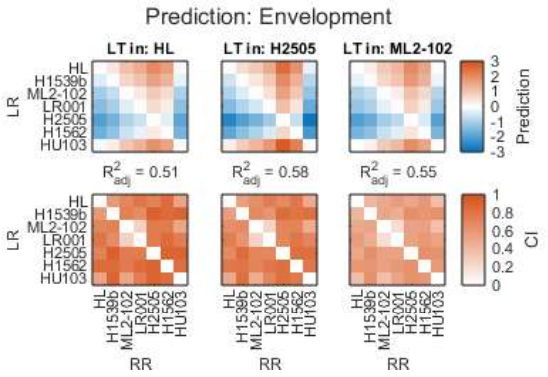


Fig. 8: Top row: Envelopment prediction based on listening tests (LT) in HL, H2505 and ML2-102 and with model function according to Eq. (3); **y-axis:** *Listening room (LR)* denotes the room in which a person is located; **x-axis:** *Reproduced rooms* (RR) denotes the rooms reproduced over headphones; **bottom row:** associated 95% confidence intervals (CI); **between:** adjusted R-squared.

Another observation that can be made is how the model quality in terms of R^2_{adj} changes depending on the parameter configurations used. All six parameters are included, which are significantly correlated across all three listening tests in Fig. 5. The results of this analysis are partially shown in Table 2. The parameter configurations are sorted in descending order according to the mean values of R^2_{adj} over the listening tests. There are 63 combinations in total, and the selected combination is ranked 27th. The selected combination may not have the highest R^2_{adj} value, but the difference is only slight in the upper 50%.

With respect to the initial questions, it was found that the *envelopment* can be predicted based on room acoustic parameters. Furthermore, it was shown that the prediction seems to be independent of the listening room. However, this requires that the listening room is included as a reference in the form of a difference measure between listening room and reproduced room.

Plausible Fit: Fig. 7 shows that the parameter $CS_{LD_{13}}$ correlates with the models in all three listening tests.

[4] Formula notation according to Matlab function fitglme

Table 2: R^2_{adj} values for GLMM based on the three listening test and different parameter combinations sorted by the mean R^2_{adj} over all three tests, for envelopment prediction. **bold:** selected parameter combination

Parameter set	HL	H2505	ML2-102
$D_{50}, C_{50}, C_{80}, DRR$	0.522	0.583	0.551
EDT, C_{50}, C_{80}, DRR	0.522	0.581	0.550
$EDT, D_{50}, C_{50}, C_{80}, DRR$	0.519	0.583	0.549
...	...	...	...
EDT, T_s, C_{50}, DRR	**0.513**	**0.575**	**0.550**
...	...	...	...
C_{50}	0.491	0.456	0.512
DRR	0.469	0.475	0.465
C_{80}	0.434	0.346	0.445

Instead of the names of the listened rooms, this similarity measure is now used as a fixed effect. The following model function[5] can be obtained:

$$Plausible fit \sim 1 + CS_{LD_{13}} + (1|participant) \quad (4)$$

Using the model function according to Eq. (4), a GLMM was generated for each listening test. For the following prediction $CS_{LD_{13}}$ is first calculated for all room combinations according to the equation 2. This way each room is considered as a listening room and a reproduced room in the prediction. The predictions within the models are then normalized to 1 with respect to the diagonal in order to obtain a prediction value between 0 and 1, where 1 corresponds to an *"Ideal fit"* and 0 to a *"Extremely Bad fit"*.

Fig. 9 shows the prediction results of the three models, based on the listening test in the three rooms HL, H2505 and ML2-102, in the upper part. The values along the diagonals are, as already mentioned, 1. As the prediction is an absolute deviation, the other values are mirrored along the diagonals. It can be seen that the models based on the three rooms provide similar results. The HL and ML2-102 models appear to provide a better prediction than the H2505-based model, as they make better use of the range of values. This is also confirmed by the adjusted R-squared parameter. It shows that the model based on the H2505 listening test is a poorer fit to the data. The corresponding confidence intervals are shown in the lower part of the Fig. 9. The result is consistent with the expectation regarding the

[5]Formula notation according to Matlab function fitglme

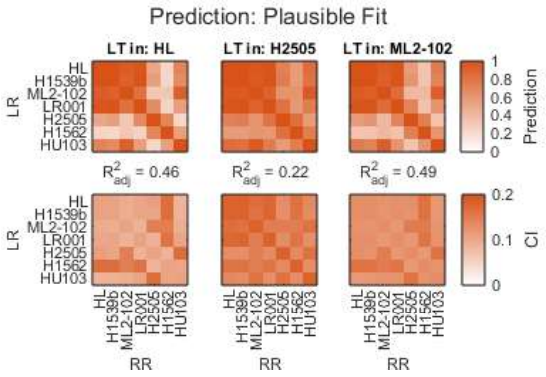


Fig. 9: Top row: Plausible fit prediction based on listening tests (LT) in HL, H2505 and ML2-102 and with model function according to Eq. (3); **y-axis:** *Listening room (LR)* denotes the room in which a person is located; **x-axis:** *Reproduced rooms* denotes the rooms reproduced over headphones; **bottom row:** associated 95% confidence intervals (CI); **between:** adjusted R-squared.

results of [6].

Regarding the initial questions, it was found that *plausible fit* can be predicted using a similarity measure based on room acoustic parameters. However, the prediction is not independent of the listening room. In addition, the prediction is much more critical, as it requires further processing of the acoustic parameters.

6 Discussion

The authors are aware that the generated models cannot be considered generally valid as there are a number of limitations. For example, only seven rooms were considered, a single stimulus signal, a fixed source-receiver arrangement, and a small number of participants. Different conclusions can be drawn between the prediction models for *envelopment* and *plausible fit*.

The prediction of *envelopment* is based on the differences in room acoustic parameters. These parameters are relatively easy to determine, making the prediction suitable for comparisons between rooms not included in this data set. However, the situation is different for *plausible fit* prediction models. These models are determined by a similarity measure based on an upstream LDA and a standardization of the parameters. The classification and resulting position of individual rooms

Table 3: Medians and interquartile ranges of all residuals from the training-validation split test for the respective models based on the listening test in the three different rooms.

	Envelopment		Plausible Fit	
	Median	IQR	Median	IQR
HL	-0.12	1.32	-0.07	1.56
H2505	-0.07	1.06	0.11	1.83
ML2-102	-0.02	1.00	-0.06	1.59

within the components are highly dependent on the dataset and cannot be assumed to be universally valid. To improve accuracy, it is recommended to generate a more diverse dataset that includes the rooms under consideration and then apply the methods.
Another critical consideration or question is how well the models can predict a subject that is not included in the modeling. In order to obtain an overview of the prediction performance, the data was split into training and validation data. Each participant was excluded first from the training data set and then the generated predictions were compared to the ratings of these participant. Table 3 shows the median and the interquartile range (IQR) for the residuals of all comparisons in the individual listening tests. The residuals are equally distributed around zero for *envelopment* and *plausible fit*. The IQR for the *plausible fit* is greater than for the envelopment. This indicates that the model for *envelopment* predicts better than that for the *plausible fit*.

7 Summary and Outlook

This paper presented a methodology for developing models that can predict acoustic perception on the basis of acoustic parameters. The research questions posed at the beginning can be answered as follows.
The question: *"Can a perceptual difference in room acoustics be modeled and predicted on the basis of parametric room acoustic differences?"* can be answered in the affirmative to a limited extent. If the modeling is based on parameter differences, this is less critical. If the acoustic parameters are initially processed further using PCA or LDA methods, the consideration of the subsequent similarity measures must be viewed critically. At this point, the authors recommend using more general data sets as a starting point for future considerations, which include the rooms under consideration. In addition to the different rooms, a variance in the source and receiver differences should also be taken into account. It should also be questioned whether the parameters used can contain the desired or required information at all.
The second question: *Is this prediction independent of the room in which the listening test is performed?* can also only be answered in the affirmative with restrictions. The models created on the basis of the three listening tests in the different rooms are not completely identical. However, they tend to show the same predictions. Further efforts can be made with regard to the definition of room categories in order to create specific models.
With the presented research the authors want to demonstrate an approach to bring perception and room acoustics together. This serves as a promising foundation for future research, which aims to address the identified limitations and weaknesses, leading to robust and widely applicable prediction models.

8 Acknowledgments

The presented work is funded within the projekts *"Multiparties"* in the framework "KMU-innovativ: Interactive Technologies for Health and Quality of Life" of the Federal Ministry of Education and Research, Germany; *"ISOPERARE"*, funded by Reality Labs Research at Meta; *"Mittelstand-Digital Zentrum Magdeburg, Teilvorhaben: IT-Strategien und Sicherheit"* of the Federal Ministry for Economic Affairs and Climate Action (BMWK), AiF, Germany and is part of *"Innovation network Chemnitz-Ilmenau-Magdeburg"*

References

[1] 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.

[2] Kuhn-Rahloff, C., *Realitätstreue, Natürlichkeit, Plausibilität: Perzeptive Beurteilungen in der Elektroakustik*, Springer-Verlag, 2012.

[3] 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.

[4] Werner, S., *Über den Einfluss kontextabhängiger Qualitätsparameter auf die Wahrnehmung von Externalität und Hörereignisort*, Ph.D. thesis, Technische Universität Ilmenau, Ilmenau, 2019.

[5] Gil-Carvajal, J. C., Cubick, J., Santurette, S., and Dau, T., "Spatial hearing with incongruent visual or auditory room cues," *Nature Scientific Reports volume 6, Article number: 37342*, 55(6(37342)), 2007, doi:10.1038/srep37342.

[6] Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K., "A summary on acoustic room divergence and its effect on externalization of auditory events," in *2016 8th Int. Conf. on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2016, doi:10.1109/QoMEX.2016.7498973.

[7] Keen, R. and Freyman, R. L., "Release and rebuildup of listeners' models of auditory space," *The J. of the Acoust. Soc. of America*, 125(5), pp. 3243–3252, 2009, ISSN 0001-4966, doi:10.1121/1.3097472.

[8] Brinkmann, F., Lindau, A., and Weinzierl, S., "On the authenticity of individual dynamic binaural synthesis," *The J. of the Acoust. Soc. of America*, 142(4), pp. 1784–1795, 2017, ISSN 0001-4966, doi:10.1121/1.5005606.

[9] 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, ISSN 0001-4966, doi:10.1121/10.0013364.

[10] 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.

[11] Int. Telecommunication Union, Rec. ITU-R BS.2159-4, "Multichannel sound technology in home and broadcasting applications," 2012.

[12] Treybig, L., Saini, S., Werner, S., Sloma, U., and Peissig, J., "Room acoustic analysis and BRIR matching based on room acoustic measurements," in *AES Conf.: 2022 AES Int. Conf. on Audio for Virtual and Augmented Reality*, 2022.

[13] 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.

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

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

[16] Erbes, V., Schultz, F., Lindau, A., and Weinzierl, S., "An extraaural headphone system for optimized binaural reproduction," *Proc. of the 38th DAGA (in: Fortschritte der Akustik)*, 2012.

[17] Int. Telecommunication Union, Rec. ITU-R BS.2132-0, "Method for the subjective quality assessment of audible differences of sound systems using multiple stimuli without a given reference," 2019.

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

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

[20] 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.

[21] Jolliffe, I. T., *Principal component analysis*, Springer, New York, 2. edition, 2002, ISBN 978-0-387-95442-4.

[22] Treybig, L., Höbel-Müller, J., Werner, S., and Nürnberger, A., "Acoustic inter- and intra-room similarity based on room acoustic parameters," *Engineering for a Changing World: Proc.; 60th ISC*, Ilmenau Scientific Colloquium, p. 2023, 2023, doi:10.22032/DBT.58929.

[23] 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.