Article Identifier
DOI: 10.22032/dbt.58929
URN: urn:nbn:de:gbv:ilm1-2023isc-136:3

60th ILMENAU SCIENTIFIC COLLOQUIUM
Technische Universität Ilmenau, 4 – 8 September 2023
URN: urn:nbn:de:gbv:ilm1-2023isc:1

Acoustic inter- and intra-room similarity based on room acoustic parameters

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

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

ABSTRACT

This paper shows various approaches for determining acoustic (dis-)similarity based on room acoustic parameter values derived from real measurements. The similarity is calculated across different room configurations and/or between different microphone-loudspeaker positions within the same room configuration. We compare supervised (LDA, Random Forrest) and unsupervised techniques (PCA, SPPA) and pre-selected visualizations in terms of their ability to exhibit inter- and intra-room (dis-)similarities. The data set generated comprises spatially high-resolution room impulse responses obtained from multiple source-receiver positions within a room configuration. The room acoustics are varied by introducing active walls and geometries accounting for specific room configurations. The results show that the separation of room configurations primarily relies on specific acoustic parameters, with the reverberation time playing an important role. Within a given room configuration, the acoustic parameters excluding the reverberation time mainly capture the orientation and distance between the source and receiver.

Index Terms – room acoustics, acoustic similarity, dimensionality reduction, classification, feature selection, spatial audio

1. INTRODUCTION

One application of spatial audio technology is the enrichment of a real physical room with additional virtual audio objects [1, 2]. The objects should fit seamlessly into the listener's spatial auditory perception of the room or environment. Therefore, there is a need to know the room acoustics of the real room to the extent that these can be reproduced for the virtual sound sources [3]. The calculation, evaluation, and comparison of acoustic parameters between different room configurations (inter-room similarity) and within a room configuration (intra-room similarity) is consequently purposeful as a basis for the

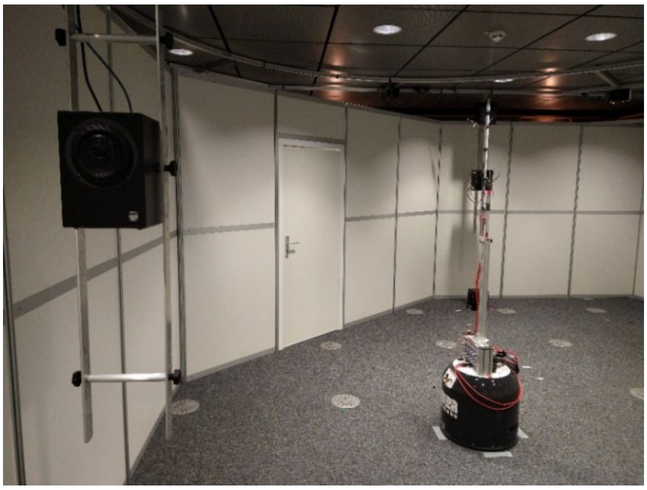

Figure 1: Arrangement of movable walls in the listening lab of the TU Ilmenau with two loudspeakers and a robot with microphone array showing one configuration of a room labeled with HL05W

so-called auditory augmented reality (AAR) [4]. In addition to room acoustic simulation, the measurement of room impulse responses is an adequate approach to determine accurate room acoustic properties and parameters [5, 6]. With high-resolution spatial measurements in a room and variations of the room geometry, a data set can be created that represents to a high degree the real physical variance of the room acoustics within a room configuration and between different room configurations [7] (see Figure 1). To quantify the acoustic properties of a real or an auditory augmented room, a bunch of different parameters could be analyzed, each trying to represent different aspects of sound propagation and perception. However, the amount of data requires a time-consuming (dis-)similarity analysis. Therefore, we propose to use different machine learning techniques and methods, as well as data compressive visualizations of room measurements, to get a parsimonious data view to understand how room acoustic parameter combinations describe the (dis-)similarity between room configurations and within a room configuration.

First, unsupervised and supervised data compression as well as classification and feature selection are applied to the measurements. We evaluate each approach with respect to its ability to separate the data set into classes. Furthermore, we determine which acoustic parameters and parameter combinations are useful for classification and with what importance.

The paper provides insight into room acoustics data at various source-receiver combinations in Section 2. Section 3 briefly describes the room acoustic parameters used. Sections 4 and 5 provide insight into the analysis methods used and a detailed analysis of the data set with respect to inter- and intra-room acoustic similarities. Section 6 concludes the paper with a summary and critical considerations for further work

2. DATASET DESCRIPTION AND CREATION

In this Section, we describe the data set used and its creation. First, the basic room and its subsequent modifications within the individual measurement series and the arrangement of the loudspeaker setup is explained. Subsequently, the microphone array used is discussed. Finally, automated measurement with a robot is explained in more detail.

2.1 Room Variation and Speaker Setup

The base room in which the measurements were made is the listening lab of the “Technische Universität Ilmenau”. The room measures 8.4m x 7.6m x 2.8m and is acoustically treated according to ITU-R BS.1116-3 [8].

For each series of measurements, movable, highly reflective walls were added to the room to change its acoustic properties. The walls have a height of 2.48 m, so there is an air gap of 0.3 m between the walls and the ceiling. A total of eight room configurations are considered, as shown in Figure 2. A room configuration is defined by a specific wall arrangement. Five R906 ME Geithain loudspeakers were used as sources. The acoustic center of all loudspeakers was at the same height of 1.73 m above the floor. The arrangement of the loudspeakers is also shown in Figure 2. On the one hand, measurements should be included where the source is not pointed directly at the receiver. On the other hand, sources excite the room differently, which can lead to higher variance between data within a room configuration. In addition, we want to have a high variance between source-receiver-room combinations.

2.2 Microphone Array

For the recording of the spatial room impulse responses (SRIRs), a custom microphone array suitable for the spatial decomposition method (SDM) was used [9, 10, 11]. It allows the estimation of the direction of arrival (DoA) of direct sound and reflections as well as the

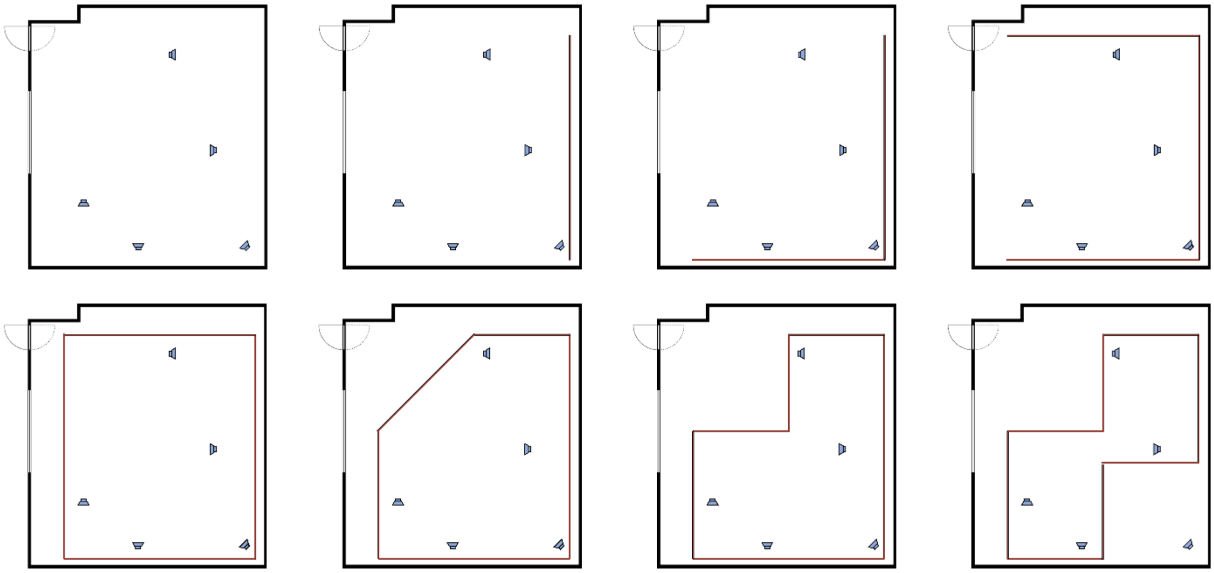

Figure 2: Room configurations for the individual measurement series with marked added walls in red and speakers in blue. The room labels in the upper row (left to right) are HL00W, HL01W, HL02W and HL03W. The room labels in the lower row (left to right) are HL04W, HL05W, HL06W and HL08W

synthesis of binaural room impulse responses (BRIRs). The microphone array consists of an Earthworks Audio M30 measurement microphone and six EM258 omnidirectional electret condenser satellite microphones [12]. The height of the measurement microphone during the measurements was 1.73 m.

2.3 Autonomous Measurements

The calculation of the SRIRs was based on sweep measurements. A logarithmic sweep from 50 Hz to 22 kHz with a length of six seconds was used. The following two seconds after the sweep was also recorded. For the purpose of redundancy, three measurements were taken for each source-receiver combination. In addition to the sweep measurement, five seconds of noise floor was recorded for each position. For HL00W to HL04W, the measurement area is set to 5 m x 6 m. For room configurations HL05W, HL06W and HL08W, the measurement area is reduced according to the room geometry. The spatial sampling resolution was set to 0.25 m along the x- and y-axis. This resulted in 188 - 466 positions to be addressed, depending on the measurement series. Due to the high manual measurement effort, a measurement robot, a TORY v2 from MetraLabs, was used. The microphone array, the microphone preamplifiers and a USB audio interface were mounted on this platform. The measurements could thus be performed autonomously. Details on the design and operation of the robot are described in [13]. A total of 3,196 measurement positions were covered in all measurement series. For five loudspeakers, this results in a total of 15,980 measurements, which are included in the analysis in Section 5. Our data set provides a balanced number of room-specific measurements. Room configurations HL00W and HL01W provide 2330 measurements, each. Room configurations HL02W and HL03W provide 2225 measurements, respectively. Room configuration HL04W is associated with 2335, room configuration HL05W with 2060, room configuration HL06W with 1535 and room configuration HL08W with 940 measurements.

3 ROOM ACOUSTIC PARAMETERS

For the room acoustics analysis in this study, several room acoustics parameters are extracted from the room impulse response recorded by a measurement microphone. The parameters used are explained below.
The early decay time (EDT) is the time it takes for the energy decay curve (EDC) to drop from 0 dB to -10 dB. It is related to the perceived reverberation time [14]. According to [14], the reverberation time is the time it takes for the sound pressure level in a closed room to decrease by 60 dB after the sound source is turned off. In this paper, the reverberation time DT_{20}, according to [15] is used. When calculating the DT_{20}, the direct sound is cut off and the reverberation time is averaged from several iterations by sampling the EDC over 20 dB decays. This minimizes the influence of the sound source, especially when the distance between the source and receiver is short and the source and receiver are not omnidirectional [15]. The definition index D_{50} and the clarity index C_{50} are related to the perceived transparency of sound and the syllable intelligibility of speech. The clarity index C_{80} is the equivalent for music reproduction. In addition to the definition and clarity index, the third parameter related to the perceived transparency of the sound or speech intelligibility is the center time T_S. It describes the time in the RIR at which half of the signal energy is reached. Other parameters are the direct-to-reverberant-energy-ratio DRR, as well as the individual components of the direct energy ED and reverberant energy ER of the DRR. In addition, the inverse measure G_{rel} is used according to [16], which is a loudness-related measure. A more detailed consideration of the general room acoustic parameters can be found in [17, 14]. The parameters are calculated for the single octave bands from 125 Hz to 16 kHz.

4 ANALYSIS METHODOLOGY

To gain insight into how combinations of room acoustic parameters describe the (dis-)similarity between different room configurations and within a room configuration, we use a data-driven approach to analyze the measurements more efficiently than the traditional approach. Specifically, we explore a data compression, classification and feature selection approach. In the data compression approach, we use Principal Components Analysis (PCA), Sparse Projection Pursuit Analysis (SPPA), and Linear Discriminant Analysis (LDA). LDA and Random Forests (RF) models are employed as classifiers, using the room acoustic parameters mentioned in Section 3. By examining the confusion matrices and the class-specific and macro-average precision, recall and F1 scores, we gain insights into the (dis-)similarities between the room configurations from the perspective of both classifiers. To identify important features, we use a permutation feature importance wrapper heuristic. In this Section, we describe the methods used in our study.

4.1 Data Compression

In this section, we frame the estimation of the inter-room (dis-)similarity as a dimensionality reduction problem. We apply Principal Component Analysis (PCA), Sparse Projection Pursuit Analysis (SPPA) and Linear Discriminant Analysis (LDA) to standardized data. Each projection describes combinations of the room acoustic parameters, indirectly providing insights into the importance of these parameters in inter- and intra-room similarity. To visually identify patterns, we map the observations to exemplary room grids. The calculations of these methods were performed using MATLAB.

4.1.1 Principal Component Analysis
Principal Component Analysis (PCA) is a statistical method used for dimensionality reduction in multi-variate data. It is an unsupervised technique that aims to identify the most important features in a data set and project them into a lower-dimensional space. PCA calculates principal components, which are orthogonal linear combinations of the original features and captures the maximum variance present in the data set. PCA can compress data, reduce noise, and thereby identify patterns. For a more detailed look at how PCA is calculated, please refer to [18, 4].

4.1.2 Sparse Projection Pursuit Analysis
Sparse Projection Pursuit Analysis (SPPA) is a statistical method used for dimension reduction and pattern recognition in multi-variate data. Unlike PCA, SPPA focuses on finding a linear combination of a few features. It aims to identify a small number of features that are most important in describing the data while disregarding the other features. SPPA is based on the concept of projection pursuit, which aims to find interesting projections of the data that emphasize underlying structures or patterns. For a more detailed look at how SPPA works, please refer to [19, 20]. In this study, we use the MATLAB implementation of SPPA as described in [20].

4.1.3 Linear Discriminant Analysis
Linear Discriminant Analysis (LDA) is a supervised statistical method for finding a linear combination of features that separate two or more classes. Unlike PCA, LDA considers the group membership or class labels in the data. LDA defines an objective function by considering the ratio of the between-class variance to the within-class variance. The goal is to find a projection direction that maximizes this ratio. In other words, LDA aims to maximize the separation between different classes while minimizing the variance within each class. For a more comprehensive understanding of how LDA is calculated, please refer to [21, 4].

4.2 Classification
We approach the estimation of inter-room (dis-)similarity as a multi-class classification problem with eight classes, where each class corresponds to a specific room as depicted in Figure 2. To analyze the tendencies of the room similarities, we employ two classifiers. The first classifier we consider is Linear Discriminant Analysis (LDA) (see Section 4.1.3), which is suitable because the classes are partially linearly separable and have relatively balanced class-specific feature distributions. The standardized values of the room acoustic parameters are used as input for the LDA classifier. For modeling complex decision boundaries, we opt for the Random Forests (RF) classifier as our second model. As the RF classifier is not sensitive to the scale of the features, we train it on the original data. To assess the classifiers' performance, we employ a stratified 5-fold cross-validation procedure [22]. Our classification report includes the confusion matrices and the macro-averages of precision, recall, and F1 score [23].

Random Forests (RF) is a supervised machine learning algorithm widely used for classification [24, 25]. It incorporates four key concepts: ensemble learning, bootstrap aggregating or bagging, feature randomness, and voting. The RF classifier is an ensemble of decision trees designed to mitigate the overfitting issue associated with a single decision tree [26]. Each decision tree in the ensemble is constructed by recursively partitioning a subset of the training data, based on selected features. To obtain homogeneous child nodes, each node split minimizes a specific impurity criterion, such as the Gini index [27]. The RF algorithm employs bootstrap aggregating, which involves training each decision tree on random samples with replacement from the original training set [28]. Additionally, at each node, a subset of features is randomly selected to determine the best split [28]. A trained RF model makes predictions by aggregating

the individual tree predictions through majority voting, where the predicted class is determined by the most frequently predicted class among the trees.
We employ the RF classifier from the Scikit-learn library [29] with its default hyperparameter configuration. We use 100 decision trees in the ensemble. The node splits are determined based on the Gini index [25]. The tree nodes continue to split until all leaves become pure, and the minimum number of samples required at a leaf node is set to one. When searching for the best split, the RF classifier considers $\sqrt{\#\,\text{features}}$ features [30]. Each decision tree is built using bootstrapped samples. To evaluate the performance of the classifier, we calculate class-specific precision, recall, and F1 scores, as well as the macro-average of these metrics, based on the confusion matrices. Precision is computed as the ratio of true positives to the sum of true positives and false positives. Recall is the ratio of true positives to the sum of true positives and false negatives. The F1 score is the harmonic mean of precision and recall.

4.3 Feature Selection

In order to capture the importance of room acoustic parameters in automated room (configuration) discrimination, we employ wrapper feature selection with cross-validation [22]. Given the high dimensionality of the data and the presence of highly correlated features, we adapt a computationally efficient permutation feature importance wrapper heuristic [31]. This heuristic evaluates the importance of a feature by measuring the change in a performance metric when a particular feature is randomly permuted while keeping the other features unchanged [31]. For this purpose, we leverage Scikit-learn's [29] implementation of RF classification and permutation feature importance.
In our study, the RF classifier is trained on non-standardized data. The wrapper heuristic employed is based on the mean decrease in F1 score, considering both type I and II errors in room (configuration) discrimination. To obtain reliable F1 scores with low variance, we perform a stratified 5-fold cross-validation. For each feature, we repeat the feature permutation 500 times within each fold [31]. We obtain a total of 5 x 500 importance scores per feature. A feature is considered important if it consistently yields positive scores throughout the iterations [31]. Note that there may be features that have no "own" effect on the classification, but are associated with the prediction due to their correlation with important features [31].

5 RESULTS

In this Section, we investigate the (dis-)similarity between room configurations and within a room configuration using the machine learning methods mentioned in Section 4. We first evaluate the data compression approach, followed by classification and feature selection.

5.1 Data Compression

In the subsequent analysis, we focus on the condensed approaches of Principal Component Analysis (PCA), Sparse Projection Pursuit Analysis (SPPA), and Linear Discriminant Analysis (LDA) in relation to room configurations.

5.1.1 Principal Component Analysis

Figure 3 shows the PCA plot, providing a 2D visualization of the samples based on their projections onto the first two principal components. The first component accounts for 61.07% of the variance, while the second component explains 11.78% of the total variance encompassed by all room-acoustic parameters. In the Figure 3, we mark each room configuration with a distinct color and shape. There is an order in the arrangement of observations based on room configurations, although overlap exists between the clusters.

Figure 4 shows the positive and negative weights of the room acoustic parameters in the first three principal components. In the first component, EDT, DT_{20} and TS as well as D_{50}, C_{50} and C_{80}, exhibit similar absolute weights. The remaining parameters show relatively lower (absolute) weights in the first component, but their (absolute) weights increase in the second and third components. These findings align with those reported in [4], validating the consistency of our results.
Figure 5 depicts the projections of the samples onto each of the first three principal components for a specific loudspeaker position, considering their corresponding measurement positions in the room grid (rows one to three). Notably, clear spatial patterns are discernible in components one and two. These patterns reflect the orientation of the directed sound source, which varies across different room configurations, thereby manifesting in distinct value ranges. The first and second components capture the characteristics of both the source and the room, respectively. The representation of the third component on the room grid does not exhibit noticeable patterns.

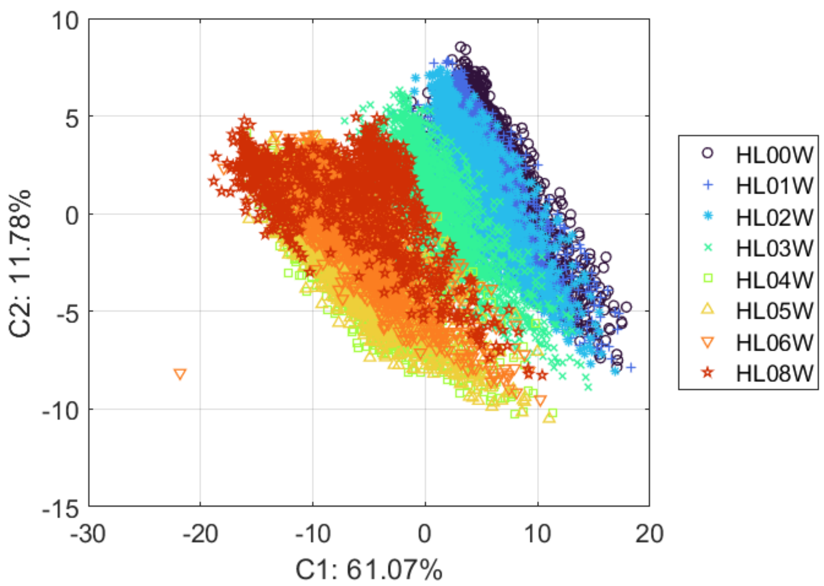

Figure 3: PCA plot showing the two-dimensional representation of the standardized room acoustic measurements using their projections onto the first two principal components C1 and C2 with explained variances. The room labels are marked by color and shape

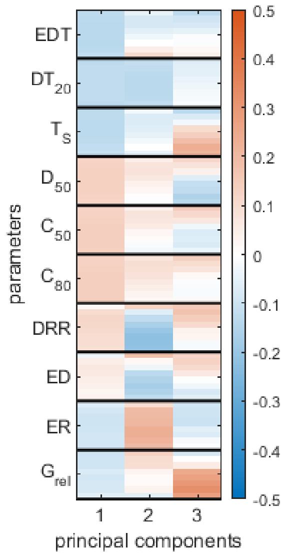

Figure 4: PCA plot showing the weightings of the standardized room acoustic parameters on the first three principal components. The weightings are based on all source-receiver combinations

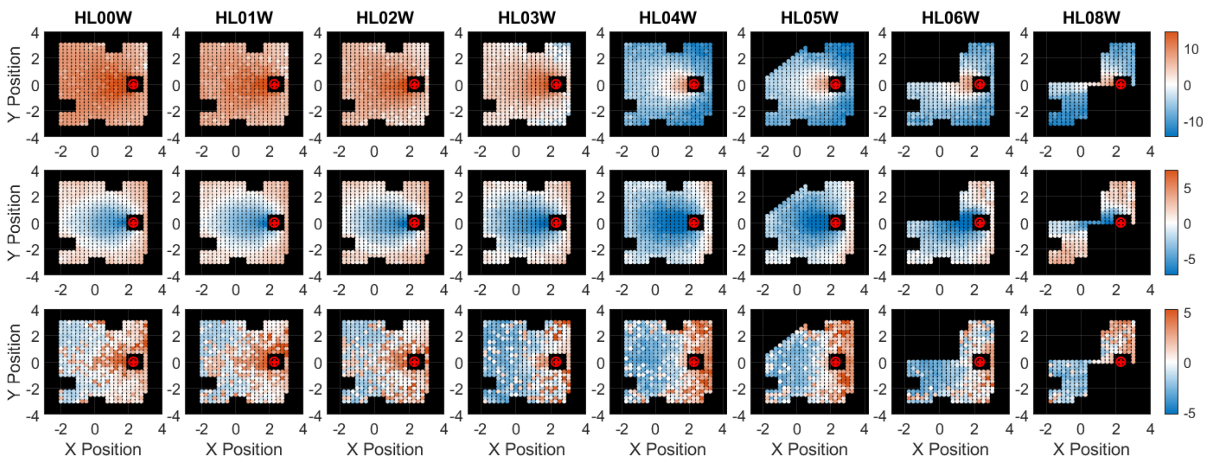

Figure 5: PCA grid plots showing the standardized room acoustic parameters projected onto the first three principal components (row one to three) for a specific loudspeaker placement across the room configurations. The red filled circle indicates the location of the loudspeaker in the room grid

The absence of patterns may be attributed to potential estimation errors. Similar patterns can be observed for the remaining loudspeaker positions.

In Figure 6, we compare the absolute weightings of the acoustic parameters in the first component between the inter-room and intra-room conditions. The first column corresponds to the absolute values extracted from the first column of Figure 4. Notably, DT_{20} contributes only marginally to the component in the inter-room condition ("all rooms"), in contrast to its variation in the intra-room conditions. This observation indicates that the variance of DT_{20} within a room configuration is small and does not contribute to the characterization of a room's properties. Instead, it primarily captures the variations between the room configurations, emphasizing its meaning in discriminating between different room configurations. The other room acoustic parameters also capture the variation between room configurations, but are more dependent on the source-receiver combination, so they show more variation within a room configuration than the DT_{20}.

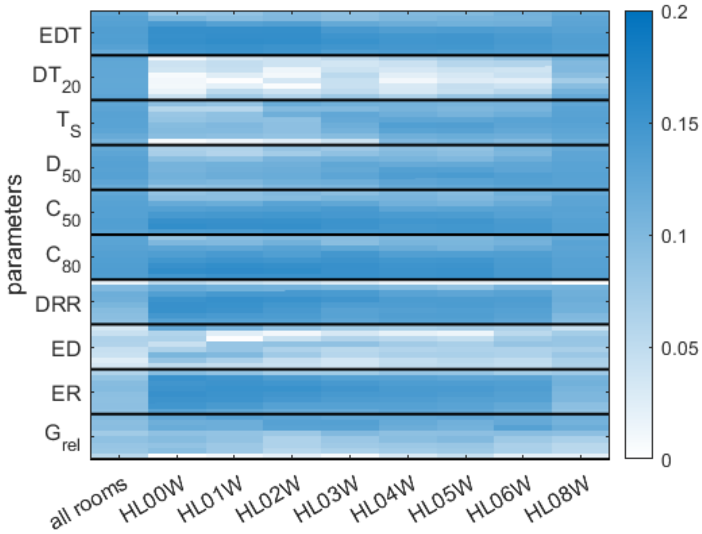

Figure 6: PCA plot showing the weightings of the standardized room acoustic parameters on the first principal component for both the inter-room and intra-room conditions

An interesting exception in Figure 6 is the case of HL008W, where DT_{20} plays a significant role in the first principal component. This suggest that there is considerable scatter in DT_{20} values within this particular room, contributing to the characterization of the source-receiver combinations. This scatter is likely attributed to the connection between two small room sections. Figure 7 provides further insight by displaying the samples projected onto the first component for the intra-room condition HL008W, considering each loudspeaker, in relation to their measurement positions in the room. With regard to the source-receiver combinations, two different areas can be distinguished in this room configuration: source and receiver within the same area or in different areas. An acoustically special case can be seen in the second subplot

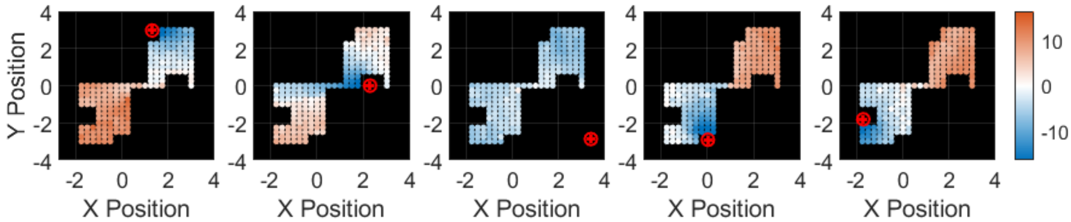

Figure 7: PCA plot showing the standardized room acoustic parameters projected onto the first principal component for the intra-room condition HL008W with different loudspeaker locations in the room grid. The red filled circle indicates the location of the loudspeaker

with a loudspeaker radiating into both room areas simultaneously. Also, the loudspeaker outside the arrangement (shown in the third subplot in Figure 7) creates an acoustically special case.

5.1.2 Sparse Projection Pursuit Analysis

In the case of Sparse Projection Pursuit Analysis (SPPA), we perform dimensionality reduction to reduce the parameter space to five dimensions. In the subsequent analysis, we focus on the projection of the samples onto a single component, since we have seen that a multidimensional view does not yield additional insights. As described in [19], in SPPA, the starting point for each new run is chosen randomly, leading to different combinations of parameters. To account for this variability, we conduct 100 runs and examine the parameters that were chosen most frequently, as described in [19]. In Figure 8, we observe that DT_{20} emerges as the parameter that primarily characterizes the underlying structure in the data. To provide a more detailed illustration, Figure 9 displays the distributions of the samples of the individual room configurations over the component, representing one run. It can be seen that the data is separated into two distinct clusters: one comprising HL00W, HL01W and HL02W, and the other comprising HL04W, HL05W and HL06W. Furthermore, the HL03W and HL08W room configurations exhibit unique structures that differentiate them from the other room configurations. The distribution observed for HL08W shows two Gaussian-like components. This can be attributed to the positioning of the speaker outside the designated room configuration. As this loudspeaker is not placed within the room like the other loudspeakers, it introduces a different set of reverberation times. Consequently, there is a coupling effect between an additional acoustic environment (outside the room configuration) and the actual room configuration. In contrast to the other room configurations, the HL03W room configuration exhibits a wide and flat distribution. The variation in reverberation times depends on the specific source-receiver arrangement in relation to the room configuration. This is illustrated in Figure 10 for a particular loudspeaker arrangement for each room configuration. According to the SPPA, the main information of the data consists of the variations between the room configurations, which can be described mainly by the reverberation time DT_{20}.

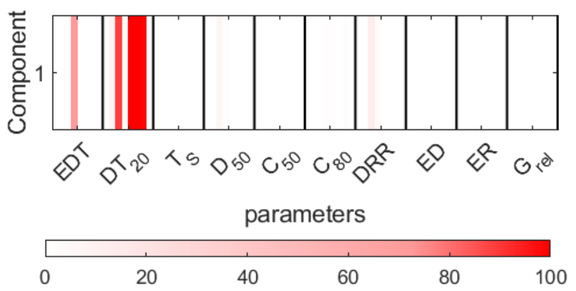

Figure 8: SPPA plot showing the number of parameter selections for 100 SPPA runs. The number of selections is marked by the intensity level of red

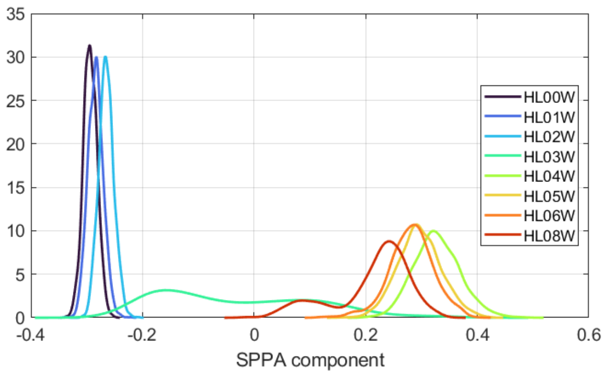

Figure 9: SPPA plot showing distributions of the standardized room acoustic measurements of the SPPA component across all rooms with all loudspeaker-receiver locations

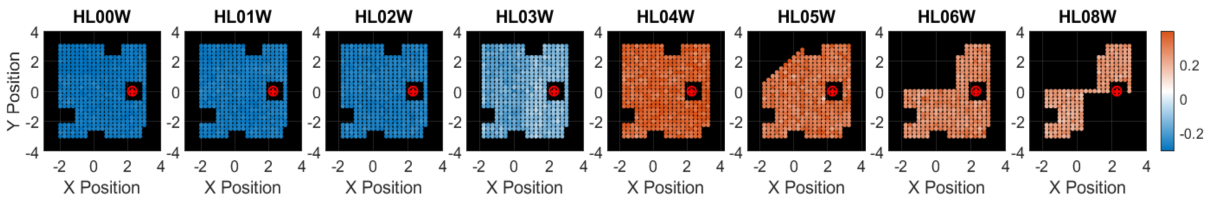

Figure 10: SPPA plot showing standardized room acoustic measurements projected onto the SPPA component for a specific loudspeaker location in the room grid. The red filled circle indicates the location of the loudspeaker

5.1.3 Linear Discriminant Analysis

In this part, we evaluate the dimensionality reduction capabilities of LDA in uncovering room (dis-)similarity. Figure 11 illustrates the samples projected onto the first and second linear discriminant (LD). Unlike PCA, which identifies directions with the largest variances in the data (as shown in Figure 3), LDA projects the data onto a low-dimensional space while maximizing the between-class and minimizing the within-class variance (as depicted in Figure 11). The arrangement of individual room configurations on the first LD component resembles that of SPPA (refer to Figure 9). Figure 12 demonstrates that the first LD component is particularly weighted by the individual octave bands of the reverberation time. The first LD already explains 92.86% of the variance of the data. These findings align with the results presented in [4]. Figure 13 shows the samples projected onto LD one and LD two for a specific loudspeaker, highlighting their relationship to the measurement positions within the different room configurations.

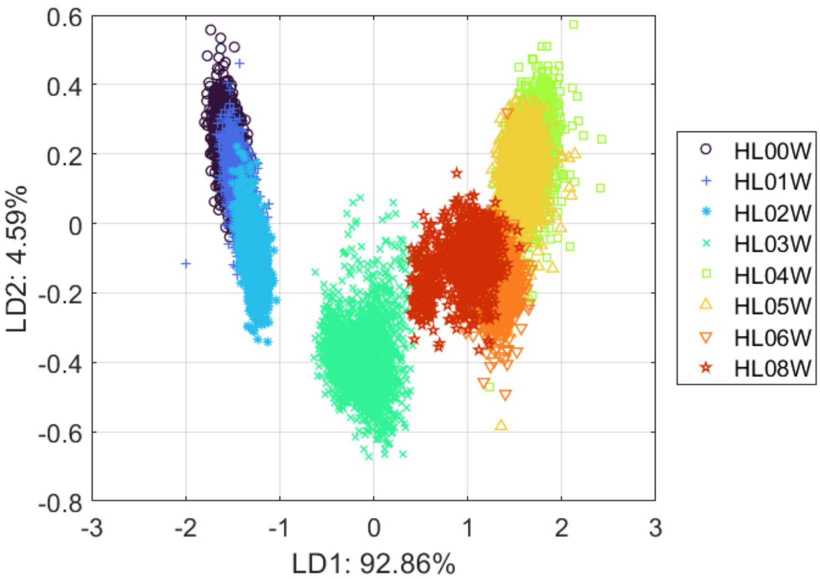

Figure 11: LDA plot of the first two linear discriminant functions LD1 and LD2 showing the room separation based on the standardized room acoustic parameters. The room labels are marked by color and shape

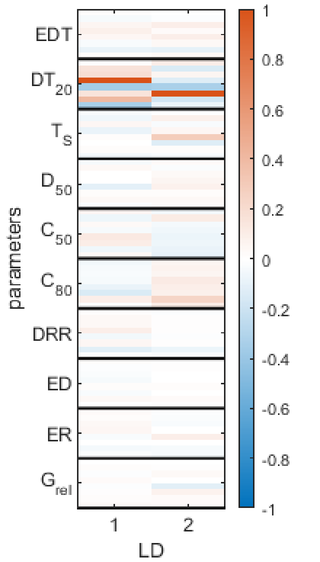

Figure 12: LDA plot showing the weightings of the standardized room acoustic measurements in the first two linear discriminant functions LD1 and LD1

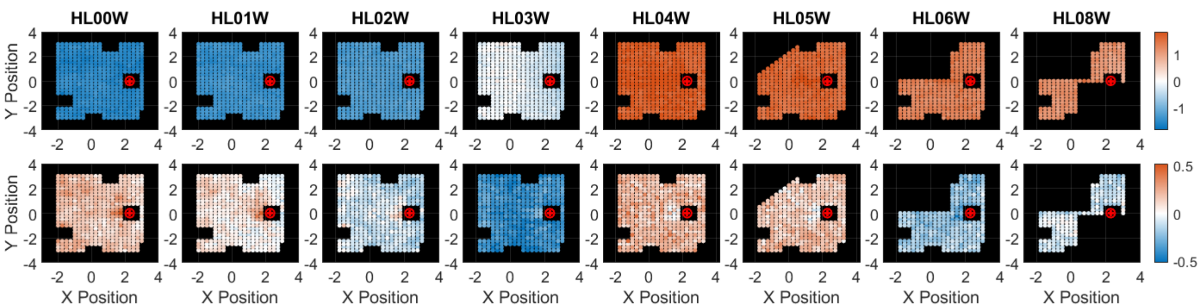

Figure 13: LDA plots showing the room acoustic measurements projected onto the first two linear discriminant functions LD1 (upper row) and LD2 (lower row) for a specific loudspeaker location in the room grid. The red filled circle indicates the location of the loudspeaker

5.2 Classification

We formulate the estimation of the inter-room (dis-)similarity as a multi-class classification problem with eight classes. Each class corresponds to a room shown in Figure 2. We use the LDA and RF classifiers with the room acoustic parameters mentioned in Section 3.

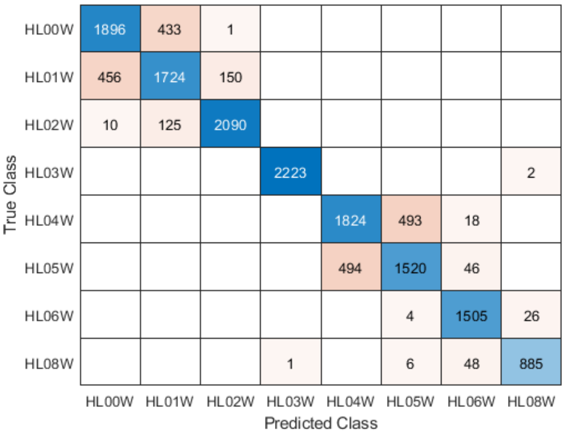

Figure 14: Confusion matrix representing the result of performing an 8-class room classification using LDA. The blue color represents a high number of correct classifications. The red color represents a high number of misclassifications

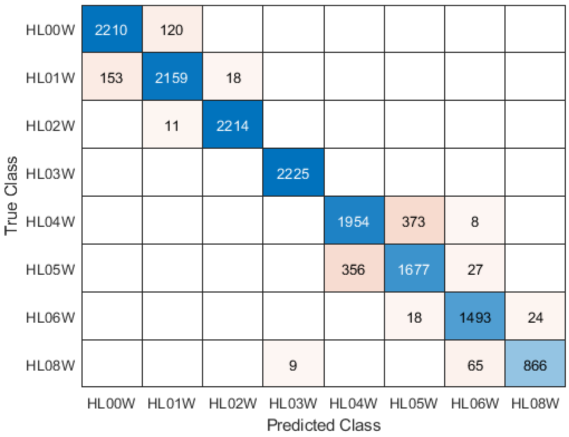

Figure 15: Confusion matrix representing the result of performing an 8-class room classification using RF. The blue color represents a high number of correct classifications. The red color represents a high number of misclassifications

Table 1: Class-specific average and the macro-average scores for precision, recall, and F1 score of an LDA classifier in a stratified 5-fold cross-validation using standardized room acoustic measurements

Class (Room)	Precision	Recall	F1-Score	Support
1 (HL00W)	80.27 %	81.37 %	80.82 %	2330
2 (HL01W)	75.55 %	73.99 %	74.76 %	2330
3 (HL02W)	93.26 %	93.93 %	93.61 %	2225
4 (HL03W)	99.96 %	99.91 %	99.93 %	2225
5 (HL04W)	78.69 %	78.12 %	78.40 %	2335
6 (HL05W)	75.14 %	73.79 %	74.46 %	2060
7 (HL06W)	93.07 %	98.05 %	95.49 %	1535
8 (HL08W)	96,93 %	94.15 %	95.52 %	940
Macro Avg.	86.61 %	86.66 %	86.62 %	15980

Table 2: Class-specific average and the macro-average scores for precision, recall, and F1 score of a RF classifier in a stratified 5-fold cross-validation using standardized room acoustic measurements

Class (Room)	Precision	Recall	F1-Score
1 (HL00W)	93.53 %	94.85 %	94.18 %
2 (HL01W)	94.28 %	92.66 %	93.46 %
3 (HL02W)	99.19 %	99.51 %	99.35 %
4 (HL03W)	99.60 %	100.00 %	99.8 %
5 (HL04W)	84.59 %	83.68 %	84.13 %
6 (HL05W)	81.09 %	81.41 %	81.25 %
7 (HL06W)	93.72 %	97.26 %	95.46 %
8 (HL08W)	97.30 %	92.26 %	94.64 %
Macro Avg.	92.91 %	92.13 %	94.64 %

By examining confusion matrices and class-specific and macro-average precision, recall and F1 scores, we gain insights into room (dis-)similarities from the perspective of both classifiers. The scores are based on the confusion matrix shown in Figure 14 for LDA and Figure 15 for RF. Tables 1 and 2 show the class-specific average and the macro-average values for the precision, recall, and F1 scores of an LDA and RF classifier.
Both the RF and LDA classifiers exhibit a similar pattern in revealing room (dis-)similarity, with the RF demonstrating slightly better performance in room discrimination. In a stratified 5-fold cross-validation, the RF classifier performs with macro-average precision, recall and F1 scores exceeding 92 % each. The analysis of the confusion matrices obtained from cross-validation highlights notable misclassifications between class 1 (HL00W) and 2 (HL01W), as well as between 5 (HL04W) and 6 (HL05W), indicating potential acoustic similarities between these room pairs. The decreasing room-specific average of precision, recall, and F1 scores shown in Table 1 and Table 2 support the increasing acoustic similarity in the mentioned order of room pairs. HL03W and HL02W emerge as the room configurations exhibiting the highest degree of acoustic distinctiveness when compared to the other room configurations, followed by HL06W and HL08W.

5.3 Feature Selection

By applying a hold-out permutation feature importance heuristic, we can determine the important and non-important room acoustic parameters for classification-based (dis-)similarities. Figure 16 shows the distribution of changes in the F1 score of a Random Forests classifier when a particular feature is permuted. The boxplots represent the aggregated changes in F1 score over five folds and 500 permutations per feature. The boxplots reveal features with importance scores of zero, indicating their non-importance for room (configuration) classification. Negative importance scores are caused by random feature variation leading to a slightly better F1 score. The corresponding features are likely irrelevant for room (configuration) discrimination. These results are consistent with the low feature weightings in the SPPA and LDA data compression approach. DT_{20} in the octave bands 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz and 16 kHz exhibit exclusively positive importance scores. These six features are associated with average decreases in F1 score of 2 %, 1.31 %, 6.25 %, 2.96 %, 1.99 % and 2.92 %, respectively. The 2 kHz octave band of DT_{20} appears to be by far the most important feature for room (configuration) discrimination. These results are consistent with the high feature weightings in the SPPA and LDA data compression approach.

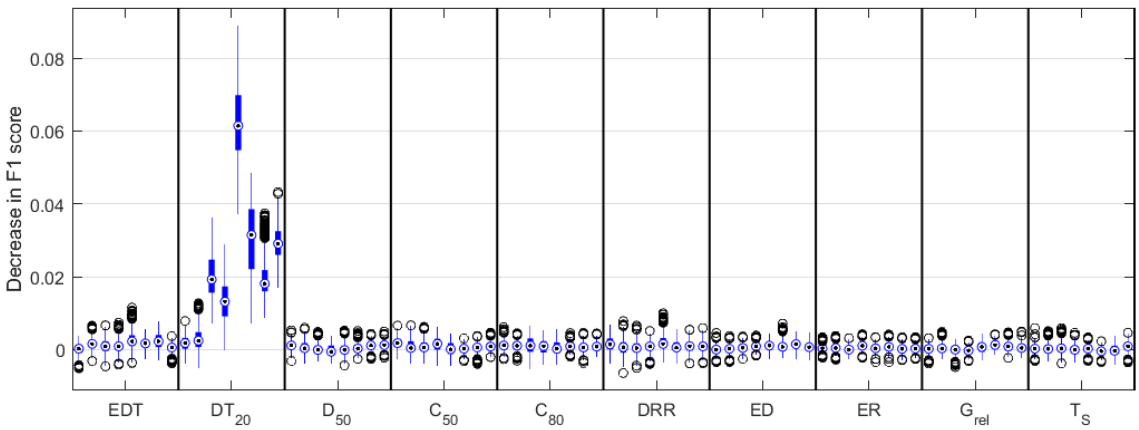

Figure 16: Hold-out permutation feature (parameter) importance scores for all room acoustic parameters based on the RF room classification of acoustic measurements. Each parameter is permuted 500 times, yielding a total of 5 x 500 scores in each boxplot. The vertical axis shows the decrease in macro F1 score. The horizontal axis shows the room acoustic parameters in octave bands 125 Hz, 250 Hz. 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz and 16 kHz

6 CONCLUSION AND OUTLOOK

This paper presents different approaches to determine the acoustic similarity between and within room configurations using room acoustic parameters extracted from real measurement data. The (dis-)similarities between the different room configurations as well as the (dis-)similarities within the configurations are investigated. For this purpose, unsupervised (PCA, SPPA) and supervised (LDA, RF) machine learning techniques are used, which can be further categorized in data compression (PCA, SPPA, LDA), classification (LDA, RF) and feature selection (wrapper heuristic) methods. These methods are applied to a room parameter set based on high spatial resolution room impulse responses recorded from numerous source-receiver combinations in a total of eight room configurations (see Figure 2).

The results of all applied methods show that the separation of room configurations, based on the selected room acoustic parameters, is largely due to the reverberation time (see Figures 6, 8, 12 and 16). These results are similar to those reported in [4]. The PCA shows that within a room configuration, the reverberation time hardly contributes to the weighting of the first component (see Figure 6). We conclude that the reverberation time does not contribute any information to the similarity consideration within a room configuration. By mapping the room acoustic measurements projected onto the first two principal components in the PCA onto the room grid, it can be seen that the principal components contain information about the orientation and distance of the source-receiver combination as well as the differences between the room configurations (see Figure 5). The SPPA shows that by reducing the parameter space to 5 parameters, the reverberation time is mainly selected (see Figure 8). Thus, the method separates reverberant and non-reverberant observations based on the used data set. For the LDA and RF classification, the different room configurations were chosen as the target attribute. The confusion matrices (see Figures 14 and 15) clearly show that the incorrect prediction of the models occurs mainly in similar reverberant room configurations. The wrapper feature selection heuristic shows that the RF classifier exhibits the reverberation time as the most important feature for room (configuration) discrimination (see Figure 16), in line with the results of the data compression analysis (see Figures 6, 8, 12).

For future considerations, several aspects can be considered, such as more measurements in other environments, more parameter extraction based on the SRIRs, or other analysis methods. A more important point, however, is linking the room acoustic parameter data with perception-related listening tests. Our previous analysis uses a set of acoustic parameters. No explicit link is made between these parameter data and the perceptual evaluation or comparison from listening tests. Although the acoustic parameters relate to the perception of e.g. reverberation or speech intelligibility, the methods used do not make any link to perceptual distinctions between individual observations. For example, instead of using the individual room configurations as classes, as in this work, the results of listening tests on perceptual differences at different measurement points could serve as classification features. Linking perceptual ratings and acoustic parameters could be used for model generation to make predictions about perception based on acoustic parameters. Especially for mixed reality applications, such models could be used to embed virtual sources into real environments.

ACKNOWLEDGMENT

The presented work is funded within the projects "Multiparties" in the framework "KMU-innovativ: Interactive Technologies for Health and Quality of Life" of the Federal Ministry of Education and Research (BMBF), Germany, and "Mittelstand-Digital Zentrum Magdeburg, Teilvorhaben: IT-Strategien und Sicherheit" of the Federal Ministry for Economic Affairs and Climate Action (BMWK), AiF, Germany.

REFERENCES

[1] G. C. Stecker, T. M. Moore, M. Folkerts, D. Zotkin and R. Duraiswami, "Toward Objective Measures of Auditory Co-Immersion in Virtual and Augmented Reality," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, 2018.

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

[3] S. Werner, F. Klein, A. Neidhardt, U. Sloma, C. Schneiderwind and K. Brandenburg, "Creation of Auditory Augmented Reality Using a Position-Dynamic Binaural Synthesis System - Technical Components, Psychoacoustic Needs, and Perceptual Evaluation," *Applied Sciences,* vol. 11, 2021.

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

[5] F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *The Journal of the Acoustical Society of America,* vol. 145, pp. 2746-2760, April 2019.

[6] S. Pelzer, M. Aretz and M. Vorländer, "Quality assessment of room acoustic simulation tools by comparing binaural measurements and simulations in an optimized test scenario," in *Proc. Forum Acusticum Aalborg*, 2011.

[7] G. Georg, S. J. Schlecht, A. M. Ornelas and V. Pulkki, "Autonomous Robot Twin System for Room Acoustic Measurements," *J. Audio Eng. Soc,* vol. 69, p. 261–272, 2021.

[8] International Telecommunication Union Rec. ITU-R BS.1116-3, "Methods for the subjective assessment of small impairments in audio systems," 2015.

[9] S. Tervo, J. Pätynen, A. Kuusinen and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the audio engineering society,* vol. 61, pp. 17-28, January 2013.

[10] S. V. Amengual Garí, J. M. Arend, P. T. Calamia and P. W. Robinson, "Optimizations of the spatial decomposition method for binaural reproduction," *Journal of the audio engineering society,* vol. 68, pp. 959-976, December 2021.

[11] N. Meyer-Kahlen, S. Amengual Garí and T. Lokki, "What the Spatial Decomposition Method can and cannot do," in *ICA 2022 proceedings*, 2022.

[12] L. Treybig, S. Werner, U. Sloma and G. Stolz, "Measure - Analyze - Auralize: From room impulse response to room classification and binaural reproduction," *Fortschritte der Akustik - DAGA 2022; Deutsche Gesellschaft für Akustik e.V. (DEGA),* pp. 1405-1408, March 2022.

[13] G. Stolz, S. Werner, F. Klein, L. Treybig, A. Bley and C. Martin, "Autonomous Robotic Platform to Measure Spatial Room Impulse Responses," *Fortschritte der Akustik - DAGA 2023; Deutsche Gesellschaft für Akustik e.V. (DEGA),* pp. 59-61, ISBN: 978-3-939296-21-8, URL: https://pub.dega-akustik.de/DAGA_2023, March Berlin, 2023.

[14] 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.

[15] L. Treybig, S. Werner, F. Klein and A. G. S. V. Gari, "Robust reverberation time estimation for audio augmented reality applications," in *154 AES Convention*, Aalto, Finland, May 2023.

[16] N. Meyer-Kahlen, S. Schlecht and T. Lokki, "Clearly audible room acoustical differences may not reveal where you are in a room," *The Journal of the Acoustical Society of America 152(2),* pp. 877-887, August 2022, DOI:10.1121/10.0013364.

[17] H. Kuttruff, Room Acoustics, 5 ed., Spoon Press, 2009.

[18] I. T. Jolliffe, Principal Component Analysis, New York: Springer, 2. edition, 2002, ISBN 978-, 2002.

[19] S. P. Driscoll, Y. S. MacMillan and P. D. Wentzell, "Sparse Projection Pursuit Analysis: An Alternative for Exploring," *Analytical Chemistry, 92 (2),* pp. 1755-1762, DOI: 10.1021/acs.analchem.9b03166, 2020.

[20] S. P. Driscoll, *Sparse Projection Pursuit Analysis (SPPA),* URL: https://github.com/S-Driscoll/SparseProjectionPursuit.

[21] K. Backhaus, B. Erichsons, W. Plinke and R. Weiober, Multivariate Analysemethoden, Berlin Heidelberg: Springer-Verlag, 2018, ISBN 978-3-662-56654-1.

[22] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence,* vol. 97, p. 273–324, 1997.

[23] D. C. Blair, "Information Retrieval, 2nd ed. C.J. Van Rijsbergen. London: Butterworths;," *Journal of the American Society for Information Science,* vol. 30, p. 374–375, 1979.

[24] T. K. Ho, "Proceedings of 3rd International Conference on Document Analysis and Recognition," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995.

[25] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, p. 5–32, 2001.

[26] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, New, York: Springer New York, 2009.

[27] L. Breiman, Classification And Regression Trees, New, York: Routledge, 1984.

[28] T. G. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization," *Machine Learning,* vol. 40, p. 139–157, 2000.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, p. 2825–2830, 2011.

[30] P. Geurts, D. Ernst and L. Wehenkel, "Extremely randomized trees," *Machine Learning,* vol. 63, p. 3–42, 2006.

[31] S. Janitza, E. Celik and A.-L. Boulesteix, "A computationally fast variable importance test for random forests for high-dimensional data," *Advances in Data Analysis and Classification,* vol. 12, p. 885–915, 2018.

CONTACTS

M.Sc. Lukas Treybig email: lukas.treybig@tu-ilmenau.de
M.Sc. Juliane Höbel-Müller email: juliane.hoebel@ovgu.de