Classification Of Geometric Room Shapes Through Room Acoustic Parameters By Using Machine Learning Algorithms

Ina Medebach, Tobias Gutenkunst, Ajan Hannemann,
Athanasios Lykartsis, Stefan Weinzierl
Fachgebiet Audiokommunikation, TU Berlin, medebach@campus.tu-berlin.de

Introduction

Architects as well as acoustical consultants often operate with room typologies in order to distinguish different categories such as 'shoe box', 'vineyard' or 'fan shape' designs. These types can be described quite precisely as architectural designs, and it is often assumed that they can also be distinguished acoustically, i.e. that a 'shoe box' sounds different than a fan-shaped auditorium. To what extent these categories can be identified based on room acoustical parameters according to ISO 3382 [1] [2] has been investigated in the current study. Two supervised machine learning approaches, the k-Nearest Neighbor (KNN) and the Support Vector Machine (SVM) classifiers, were used to classify a set of synthetically generated rooms into six architectural design categories. The results bring new insights regarding the question of how meaningful these categories are from an acoustical point of view, which of these categories are most easy to identify, which combinations of room acoustical parameters are most suitable for this identification, and which of the two classification approaches is better suited to solve the task.

Methods

To find the proper parameters for predicting the room shape, it is necessary to have a dataset which includes state-of-the-art room acoustic parameters and all the different room shapes which are primarily in use. The dataset we used was created in connection with a master's thesis at the Audio Communication department of the TU Berlin in 2015 (Ackermann and Ilse, 2015 [3]), called Ground Truth for Room Acoustical Analysis and Perception (GRAP). The goal of this dataset is to cover a wide range of acoustic environments. It contains 49 rooms which are categorized in six different shapes and provides 12 different room-acoustical parameters. The categories (and therefore classes for the machine learning problem) of the six room-shapes can be found in Table 1. The rooms which are labeled as *complex*, are rooms with particular geometry which do not fit that of any other shape category.

The impulse responses on the basis of which all the parameters were calculated were acquired through digital models of the rooms. Each room was closely modeled to an existing original. There were one source (S) and two receiver points (R1 and R2). The positions were chosen based on the DIN EN ISO3382-1 (2009) and the DIN EN ISO 3382-2 (2008). As shown in Figure 1. This results in two datasets: The first one, for which the calculations are based on receiver position 1 (R1) lying on the same axis (x) as the source; and the second one, where he calculations are based on the receiver position 2 (R2) on the right side of the x axis to the source and further away from the source.

Table 1: Classes of the six room-shape categories and the numbers of rooms belonging to the class.

Shape	Count
Shoe Box	25
Church	6
Vineyard	3
Opera/ Horse Shoe	7
fan-shaped Auditorium	5
Complex	3

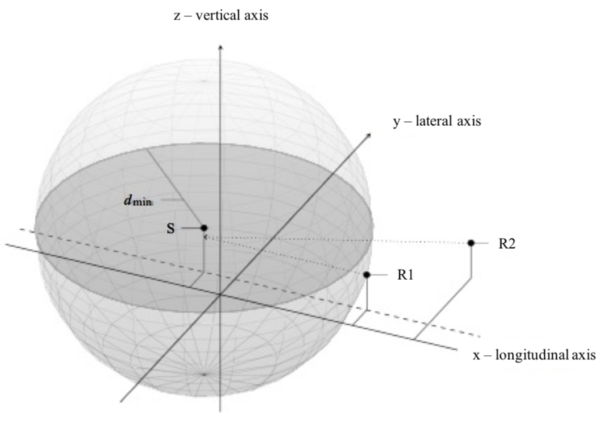

Figure 1: Model of the microphone positions used in GRAP.

The provided parameters are based on the ones recommended in the DIN EN ISO 3382-1 (2009) for describing the room acoustical impression and quality of performance rooms. Those are shown in Table 2: The parameters are available with different temporal integration bounds, resulting in a set of 53 parameters in total. For most of the parameters there is just one impulse response required, recorded with an omnidirectional microphone. For those who require separate lateral-sound information (J_{LFC}, J_{LF}, L_J), two impulse responses are necessary, one from an omnidirectional and one from a bidirectional microphone. For calculating the binaural IACC, impulse responses from an artificial head are used. Because of the room acoustic simulation there are no confounding factors in acquiring the impulse response, therefore this dataset leads to a verifiable testing environment and to consistent results.

Table 2: Provided acoustic parameters and their abbreviations.

Parameter	Abbreviation
Early Decay Time	EDT
Reverberation Time	T
Clarity	C
Deutlichkeit	D
Strength	G
Centre Time	T_S
Lateral Energy Fraction	J_{LFC}, J_{LF}
Late Lateral Sound Level	L_J
Interaural Cross-Correlation Coefficient	IACC, $IACC_A$, $IACC_E$, $IACC_L$
Bass-ratio (additionally)	BR

Firstly, the dataset has to be preprocessed prior to classification. As can be seen in Figure 1, the different classes had a very varying room count (ranging from 3 rooms of the class Vineyard to 25 rooms of the class Shoe Box), resulting in a highly unbalanced dataset which can lead to skewed classification results, due to the classifiers identifying every room as belonging to the greater class, since that would provide an overall higher accuracy. Therefore, we had to decrease the maximum number of rooms per class to 12, which left 36 usable rooms. As can be seen in Figure 2, the data of position 1 and 2 of the GRAP dataset are not identical. For that reason, we were able to use the second position as an independent room observation, and increase the number of rooms to a total of 58.

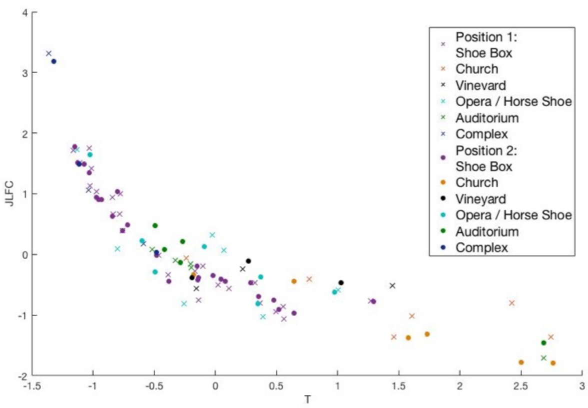

Figure 2: Acoustic parameters J_{LFC} and T of GRAP dataset, position 1 (×) and 2 (·). The distribution of the z-normalized data shows no overlapping, which indicates independent room observations of both receiver positions.

For the first part of our supervised machine learning task (classifying the room shape based on acoustic parameters), we tested all possible combinations of selected parameters (EDT, T, C, D, G, T_S, BR, $IACC_A$, $IACC_E$, $IACC_L$, J_{LFC}, J_{LF}, L_J). We allowed all parameter subsets, from a single parameter to all parameters combined. This led to a total of 8191 combinations. All data were normalized per z-Score. Figure 3 and 4 show the distribution of the dataset for two parameters. As shown in Figure 3, we reasoned that a non-linear Support Vector Machine would be a suitable classifier, based on some of the data being distributed in a way that a SVM with a polynomial kernel could provide good class separation while assuring a low complexity (e.g., lower than an SVM with a Radial Basis Function (RBF) Kernel).

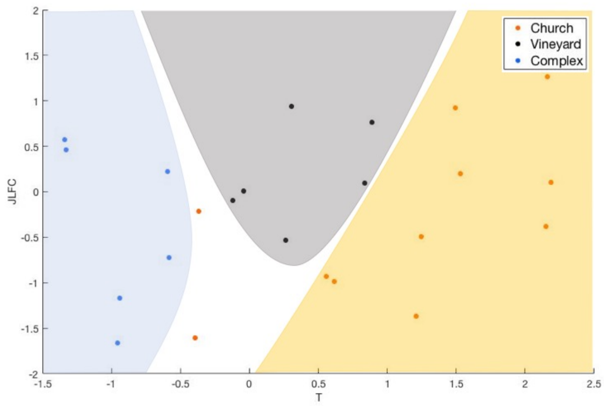

Figure 3: Scatter plot of the acoustic parameters J_{LFC} and T of GRAP (the data was z-normalized and reduced to the classes: Church, Vineyard and Complex). Distribution indicates a SVM classifier, the room data points are distributed in areas.

As shown in Figure 4, some data is spread around several centers of clusters, which suggests using a nearest neighbor approach. It must be mentioned that these are just examples for combinations for two parameters, whereas the distribution of the data in higher dimensions might be completely different. Therefore, to find the highest accuracy of all parameters and combinations, we choose to evaluate an SVM as well as an KNN approach. A SVM finds a hyperplane which separates the classes through a kernel transformation, while maximizing the distance between them. Different settings (such as the cost parameter C and the kernel type) define the form of the separation surface [4]. The KNN, a non-parametric method, examines the classes of surrounding samples and selects a class for a sample depending one the majority of its neighbors. The number of neighbors, the distance measure and a weighting of distance can be chosen [4].

To evaluate the performance of the classifier, we calculated the accuracy (ACC), the ratio between correctly classified rooms to the total amount of rooms. To reduce the risk of overfitting, we wanted to utilize classifiers with less specific criteria (higher amount of neighbors and lower degrees of functions). After testing different kinds of these two classifiers, we chose a SVM with a cubic kernel function, and a weighted 10-NN with Euclidean metric and a weighting squared inversely with respect to distance.

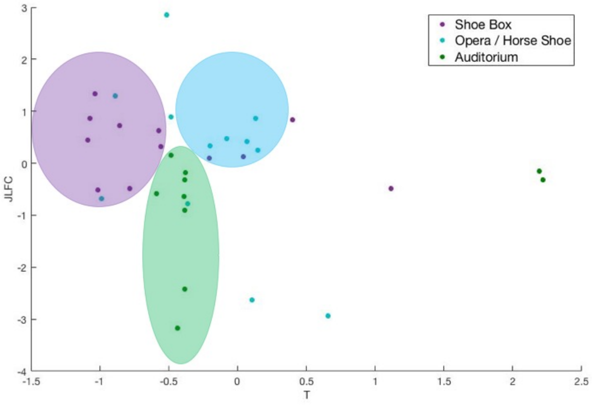

Figure 4: Scatter plot of the acoustic parameters J_{LFC} and T of GRAP (the data was z-normalized and reduced to the classes: Shoe Box, Opera / Horse Shoe and fan-shaped Auditorium). Distribution indicates a KNN classifier, the room data points are distributed in around centers.

In order to further reduce overfitting, we conducted a 5-fold cross-validation. This validation divides the data into 5 random groups, in which every group is once the test set compared to a training set out of the remaining 4 groups. The accuracy is thereby the average accuracy of those 5 iterations. To even out accuracy variations, we repeated this test 20 times and generated an average accuracy out of the 100 trials in total.

Results and Discussion

After the cross-validation, the accuracies varied by up to 29 percentage points for KNN and 62 percentage points for the SVM per combination. For this reason we compared not only the maximum, but also the average value for each set. In the following tables (Table 3 and Table 4), the best 5 combinations to classify the room shapes, resulting from a direct comparison, are shown. As previously mentioned, the rooms were distributed to 6 classes with a total of 58 rooms. Shoe Box, Church and Opera comprised 12 rooms each, Auditorium had 10, Vineyard and Complex had 6 each.

Table 3: Accuracies for the 5 best parameter combinations for predicting the shape based on KNN, showing minimum to maximum (min - max) and average accuracy (avg ACC).

Parameters	min - max	avg ACC
EDT,G, BR	0,64 - 0,85	0,74
T,G, BR	0,64 - 0,79	0,73
EDT,T, G,BR	0,60 - 0,79	0,71
T,C,G, $IACC_E$,BR	0,62 - 0,76	0,69
EDT,T,C, G,BR	0,62 - 0,76	0,68

Table 4: Accuracies for the 5 best parameter combinations for predicting the shape based on SVM, showing minimum to maximum (min - max) and average accuracy (avg ACC).

Parameters	min - max	avg ACC
EDT, T, C, G, T_S, $IACC_E$, $IACC_L$, BR	0,21 - 0,72	0,65
EDT, C, G, $IACC_E$, $IACC_L$, BR	0,17 - 0,72	0,65
EDT,C,G, T_S, $IACC_E$, $IACC_L$, BR	0,17 - 0,71	0,65
T,C,G, T_S, $IACC_E$, $IACC_L$, BR	0,17 - 0,72	0,65
EDT, T, C, G, $IACC_E$, BR	0,17 - 0,72	0,64

As can be seen in Table 3, KNN-weighted creates the best accuracy with an average of 74 % when using EDT, G, BR followed by T, G, BR with 73 %. As can be seen in Table 3, the parameters EDT, T, G and BR are able to classify the room shape with an average of 74 % accuracy. The first and third best combination, as shown in Table 3, only differentiate in 3 percentage points and one additional parameter: T. As a result, we are able to conclude that T does not significantly increase or decrease the accuracy of a combination which already included the parameter EDT, therefore one of those parameters need not be used. A similar result can be observed for the second combination, were EDT is exchanged with T, leading to a 1 percentage point lower accuracy. With an range of 21 percent points, these differences are negligible.

To evaluate the meaningfulness of the single categories, we analyzed the confusion matrix of the best combination: EDT, G, BR. A confusion matrix shows the predicted classes for all samples with respect to the actual class they belong to. At a single test, this combination created a accuracy of 70 % and a prediction as shown in Table 5. The model was able to classify all rooms of the category Auditorium (class 5). Followed by only one miss-prediction of a Church (class 2) and Vineyard (class 3). This indicates, that those geometrical shapes are different from the rest and easily to identify. On one side, rooms of the categories Complex (class 6) and Opera / Horse Shoe (class 4) are less likely to be classified correctly, which leads to a lower differentiation probability. On the other side, rooms with the shape of an Opera / Horse Shoe were likely to be miscategorized to Complex. This is also shown for the Complex rooms, which were misclassified to Operas and Horse Shoes. This questions the consistency of those two classes. Rooms of the shape Shoe Box (class 1) were the least accurate classified geometric room shape. This concludes a similar behavior to other rooms, which makes this category less reliable.

The results of the SVM approach (Table 4) are ranked

Table 5: Confusion Matrix for the best KNN combination EDT, G, BR. Classes 1-6: Shoe Box, Church, Vineyard, Opera / Horse Shoe, Auditorium, Complex

true	predicted class					
	1	*2*	*3*	*4*	*5*	*6*
1	**4**	1	2	1	2	2
2	0	**11**	0	0	1	0
3	1	0	**5**	0	0	0
4	0	0	1	**7**	1	3
5	0	0	0	0	**10**	0
6	0	0	0	3	0	**4**

by average accuracy. Results exhibit a variation up to 55 percentage points, which makes this approach less robust, although the average accuracy is still satisfactory (but lower than the weighted-kNN). As can be observed, the SVM works better with more parameters/ dimensions. The parameters EDT/T, G, BR are represented in all superior combinations in this case as well. In this ranking, the exchangeability of EDT and T can again be seen in the third and fourth position. The first and third position also show no advantage of including T to EDT. In contrast to the results of the KNN, the parameters $\mathrm{IACC_E}$, $\mathrm{IACC_L}$ and $\mathrm{T_S}$ are widely represented in the best combinations.

Conclusion

The results show that different architectural designs of musical venues can, to a certain degree, be identified by room acoustical properties, as represented by parameters according to ISO 3382–1. We tested two standard machine learning methods. With an achieved accuracy of 74 %, the KNN weighted approach was more successful than the cubic SVM. Auditorium, vineyard designs and churches produced the least misclassifications. The fact that a combination of EDT / T, G, BR provided the best results, however, suggests that the successful classification is, at least partly, due to the fact that certain designs are systematically correlated with higher or lower size and reverberation (and thus with T and G), while parameters assumed to be strongly linked to the geometry of the room such as $\mathrm{J_{LFC}}$, $\mathrm{J_{LF}}$ or $\mathrm{L_J}$, were not among the most successful combinations. Hence, the question whether different architectural designs can be identified by a different sound, can not be conclusively answered. Future work in this direction will not only require a considerable extension of the dataset by generating more rooms, but also more and / or better room acoustical parameters in order to fully represent the room acoustical impression. With respect to the methodology, future work will also include other classifiers, such as SVM with different kernels or neural networks.

References

[1] DIN EN ISO 3382–1: Measurement of Room Acoustic Parameters Part 1: Performance Rooms (2009).

[2] DIN EN ISO 3382–2: Measurement of Room Acoustic Parameters Part 2: Reverberation Time in Ordinary Rooms (2008).

[3] Ackermann, D. and Ilse, M.: The Simulation of Monaural and Binaural Transfer Functions for a Ground Truth for Room Acoustical Analysis and Perception (GRAP). Master thesis, TU Berlin (2015).

[4] Duda, R. O., Hart, P. E., and Stork, D. G.: Pattern classification. John Wiley & Sons, 2012.

Additional data

This QR-Code embeds all results of this Paper. Besides the results of all possible combinations for KNN and SVM (Tables 3 and 4), there is a digital copy of this paper. Alternative link: https://tubcloud.tu-berlin.de/index.php/s/5WSs6qbF4sHAH8E