25·29 August 2024
inter.noise
NANTES FRANCE

Predicting real indoor soundscapes based on auditory and non-auditory factors across different loudness ranges with linear and nonlinear models

Siegbert Versümer[1]
Patrick Blättermann[2]

University of Applied Sciences Düsseldorf, Institute of Sound and Vibration Engineering
Münsterstr. 156, 40476 Düsseldorf, Germany

ABSTRACT

Studies on the indoor sound environment have mainly focused on the impact of traffic, neighbor, or ventilation noise on health, annoyance, or sound quality. When conducting such studies in the laboratory, acoustic stimuli are often tailored to the research question, which presumably do not reflect the variation in the acoustic environment at home. Thus, we re-analyzed soundscape data gathered by participants at their homes several times a day. Since the recorded soundscapes differed notably between participants (and their dwellings) in terms of loudness, we investigated if cross-validated linear regression and Random Forest models predicting the soundscape dimensions based on four loudness quartile subsets achieve higher prediction performance compared to models using the full volume range. Our results, which are based on performance metrics of the statistical models, the correlation between the targets and acoustic and perceptual predictors, and the predictor importance, suggest the use of Random Forest models based on the entire loudness range. In our analysis of predictor importance, Pleasantness was understandably dominated by perceptual measures, whereas Eventfulness was driven through acoustic and psychoacoustic predictors. Interestingly, all models based on quiet soundscapes behaved differently, which provides scope for investigating the underlying causes in future.

1. INTRODUCTION

Soundscape research, following or adopting the soundscape standard [1–3], investigates the individual perception of the acoustic environment in context. While most soundscape studies focus on urban public spaces assessed in situ [4–6] or in the lab [6,7], recent studies also center on indoor soundscapes. Besides the investigation of workplaces and shopping malls, the investigation of the acoustic environment in people's dwellings is particularly important, as these should provide a restful space. While Torresin et al. [8] used a lab-based setting, Versümer et al. [9] aimed at high ecological validity and conducted a field study in which participants had to record and evaluate the acoustic environment in their home every hour during their 10-day involvement (without the presence of an experimenter). Whereas a steady city hum in urban public spaces leads to a constant noise level, the

[1] siegbert.versuemer@hs-duesseldorf.de

[2] patrick.blaettermann@hs-duesseldorf.de

Permission is granted for the reproduction of a fractional part of this paper published in the Proceedings of INTER-NOISE 2024 provided permission is obtained from the author(s) and credit is given to the author(s) and these proceedings.

sound that people experience in their homes can range from silence to very loud sounds (vacuum cleaner, music, playing children). In addition, it can be assumed that the sound environments differ substantially between dwellings, so that some people (predominantly) do not experience very loud environments at home, while others do not experience silence. It therefore seems questionable whether a generalized model is equally suitable for all people (and their homes), or whether models for partial sound pressure level ranges are better suited. Furthermore, it is of interest which factors influence the perception of the acoustic environment in terms of *Pleasantness* and *Eventfulness* of the soundscape, and whether they differ between sound pressure level subsets.

To address these questions, we reanalyzed the data from [9] and performed cross-validated (CV) linear regressions (LR) to model the *Pleasantness* and *Eventfulness* of indoor soundscapes based both on the whole dataset and four sound pressure level subsets. In addition, machine learning (ML) models based on random forests (RF) were also conducted and tuned inside a nested CV, since some nonlinearities have to be expected [10]. For the prediction of both target variables, we used acoustic and psychoacoustic variables as predictors for all models, as well as person-related variables and perceptual measures of the soundscape components and the situation. Finally, by identifying individual differences in sound perception, this study contributes to a deeper understanding of real indoor soundscapes.

2. METHOD

2.1. Field study description

In the original study from [9], participants answered the person questionnaire at the beginning of their participation. Four trained experimenters provided profound task descriptions together with explanations of the items and scales of the questionnaires used for each subsequent assessment at people's homes. There, the participants were notified on an hourly basis by a separate smartphone that also provided the questionnaires for the evaluation of the soundscape and the situation [11]. Participants rated their state of affect (valence) and arousal at the beginning of each assessment, immediately before making binaural recordings lasting 15 seconds. These recordings were made using two microphones positioned in the participants' ears and a specially designed low-noise device. Then, they evaluated (besides others) the soundscape they experienced, the soundscape composition, and further aspects of the situation, like the activity they performed directly before the assessment, or the perceived control over the sound environment. In addition to the hourly trigger, the participants could also start the evaluation manually, which many of the participants did.

2.2. Variables

Both target variables, the soundscape dimensions *Pleasantness* and *Eventfulness*, which range between [-1; 1], were based on the participants' ratings of these eight soundscape items according to ISO 12913-3 [2]: *pleasant*, *annoying*, *calm*, *chaotic*, *vibrant*, *monotonous*, *eventful*, *uneventful*. The independent variables used in this reevaluation as predictors are shown in Table 1.

The ArtemiS SUITE from HEAD acoustics [12] was used to calculate the acoustic and psychoacoustic parameters for both recorded channels, whereby we selected the maximum value of both channels to calculate the models.

Table 1: Predictors

Variable	Description		
LAeq	A-weighted sound pressure level. Only used for building the *LAeq* quartiles. Not used for modeling.	Volume	Acoustic and psychoacoustic measures
Loudness	DIN ISO 532-1 [13]		
LAF10-LAF90	$L_{AF10} - L_{AF90}$, difference of percentage exceedance levels	Time variability	
Rel.Approach	Relative approach [14]		
LCeq-LAeq	$L_{Ceq} - L_{Aeq}$, low frequency components	Spectral content	
Sharpness	DIN 45962 [15] high frequency components		
Roughness	ECMA-418-2 [16]		
Tonality	ECMA-418-2; ECMA-74 [16,17]		
SC_....	The perceived intensity of sounds from …*nature*, *human*, *traffic*, or *other* sounds	Soundscape composition	Perceptual measures
Valence	Core affect [18]	Situation	
Arousal			
Control	Perceived control over the environment		
Activity_active	[0 = other; 1 = active]. Performing an active activity (before the assessment) like housework, workout, socializing, instead of a relaxing, sleeping, or concentrating		
Mental demand	Mental and physical demand experienced during the activity. NASA TLX [19]		
Physical demand			
Age		Person	Person-related measures
Gender_male	[0 = female or non-binary; 1 = male]		
Noise sensitivity	[0; 3] NoiSeQ-R [20,21]		
Well-being	[0; 100] WHO5 [22,23]		
nPeople	Number of people living together	Household	

2.3. Sample

From the original study, 96 adult participants (*Age*: $M = 35$, $SD = 14$) were selected who made 40 or more observations. We found an even gender distribution (56% male; 44% female or non-binary), low *Noise sensitivity* ($M = 1.7$, $SD = 0.6$), and medium *Well-being* ($M = 59$, $SD = 17$).

2.4. Analysis

The center of the analyses is a 5-fold CV, programmed in Python [v3.10; 24], that facilitates an unbiased evaluation of all models by reducing the selection bias. Due to the hierarchical data structure (repeated measures within participants), a customized group-k-fold was used to assign the data to the folds. In doing so, we kept all observations of a participant within one fold to avoid group leakage,

while seeking to preserve the distribution of the target variables within each fold. In addition to LR, we also computed RF models [v2.1.1; 25] whose hyperparameters were optimized in the inner loop of a nested CV, for which the fold assignment followed a normal group-k-fold approach. For the RF models, we chose the hyperparameters shown in Table 4 (in the Appendix) and optimized them using the Optuna search engine [26], that follows a Bayesian approach and thus finds optimal values faster than grid-search. To answer the research questions, we created models based on the full data set and four subsets (Q1 to Q4) based on the four quartiles (with equal number of observations) of the standardized *LAeq* values.

We standardized all predictors independently both in the outer loop and inner loop to avoid data leakage. To prevent overfitting and underfitting, both of which are frequently observed when hyperparameters are tuned only by minimizing the prediction error (e.g., in terms of MSE), we applied this objective function:

$$\mathrm{OF} = MSE + 0.5 * \mathrm{abs}(MSE_\Delta) + 2 * \max(0, (MSE_\Delta - Error_\Delta)) + \max(0, -MSE_\Delta) \quad , \tag{1}$$

where $MSE_\Delta = MSE - MSE_{\mathrm{Training}}$, and $Error_\Delta = 0.05$ serve as a threshold for the maximum allowable MSE_Δ value. Thereby, we minimized the prediction error, but also achieved what we consider to be an acceptable MSE_Δ in the range of [0; 0.05]. In addition to the *MSE, w*e calculated the (in-sample) coefficient of determination, the

$$R^2 = 1 - MSE / VAR(Y) \quad , \tag{2}$$

which can become negative in case of prediction errors (*MSE*) larger than the variance of the target variable ($VAR(Y)$). With linear models, the coefficient of determination can be interpreted as the proportion of the explained variance of the target variable, which is not the case with nonlinear models. In general, the coefficient describes the fraction by which the variance of the errors is less than the variance of the dependent variable. Thus, higher values indicate a higher model performance.

To discuss the research questions, we report the mean model fit value, i.e., the average of the 5 model instances of the outer CV, as well as the corresponding standard deviation range. To assess the predictor importances, the (absolute values of the) standardized coefficients for the LR models and the Shapley values [27] for the RF models were used by summing up their rank over the 5 instances of the CV for each predictor, each target, and each method (LR and RF).

3. RESULTS

3.1. *LAeq* subset model performance measures

To answer the first research question, whether a generalized model is equally suitable for all people (and their homes), or whether models for partial sound pressure level ranges are better suited, Table 2 shows the mean in-sample R^2 values (of the R^2 values of all 5 CV instances of each model). We observed that about 20% of the variance in the *Pleasantness* ratings can be explained or reduced for both methods (LR and RF) for the full set and the louder quartile subsets. The decrease of variance was lower for quieter quartiles, especially for RF and for the quietest quartile Q1. Thus, no general advantage of the quartile subset models could be observed while model performances tend to be comparable to the full set model performance except for Q1.

Looking at the *Eventfulness* ratings, the LR full set model achieved also about 20% variance explanation, while the RF model based on the full set stands out with the highest predictive power, reducing 27% of the *Eventfulness* ratings variance. In the quartile subsets, we generally observed smaller amounts of explained or decreased variance compared to the full set models and a decreasing

performance trend from Q1 to Q4, while Q1 again stands out as significant exception. Q2 to Q4 fall especially behind, with the LR model of the loudest Q4 showing the lowest predictive power. Thus, the quartile subsets were again no improvement over the models based on the entire dataset, of which the RF version should be preferred due to its superior predictive power.

Table 2: Mean in-sample R^2 values (of the R^2 values of all 5 CV instances of each model). LR: Linear regression. RF: Random Forest. Qn: Model based on the nth *LAeq* quartile subset. SD($R^2_{LR_i}$): [0.016; 0.122] and SD($R^2_{RF_i}$): [0.024; 0.074] was observed for all 5 instances of the CV of all data (sub/full) sets.

Target	Method	Q1	Q2	Q3	Q4	Full set
Pleasantness	**LR**	0.15	0.21	0.20	0.20	0.21
	RF	0.05	0.16	0.19	0.18	0.19
Eventfulness	**LR**	0.17	0.12	0.10	0.04	0.21
	RF	0.18	0.10	0.08	0.09	0.27

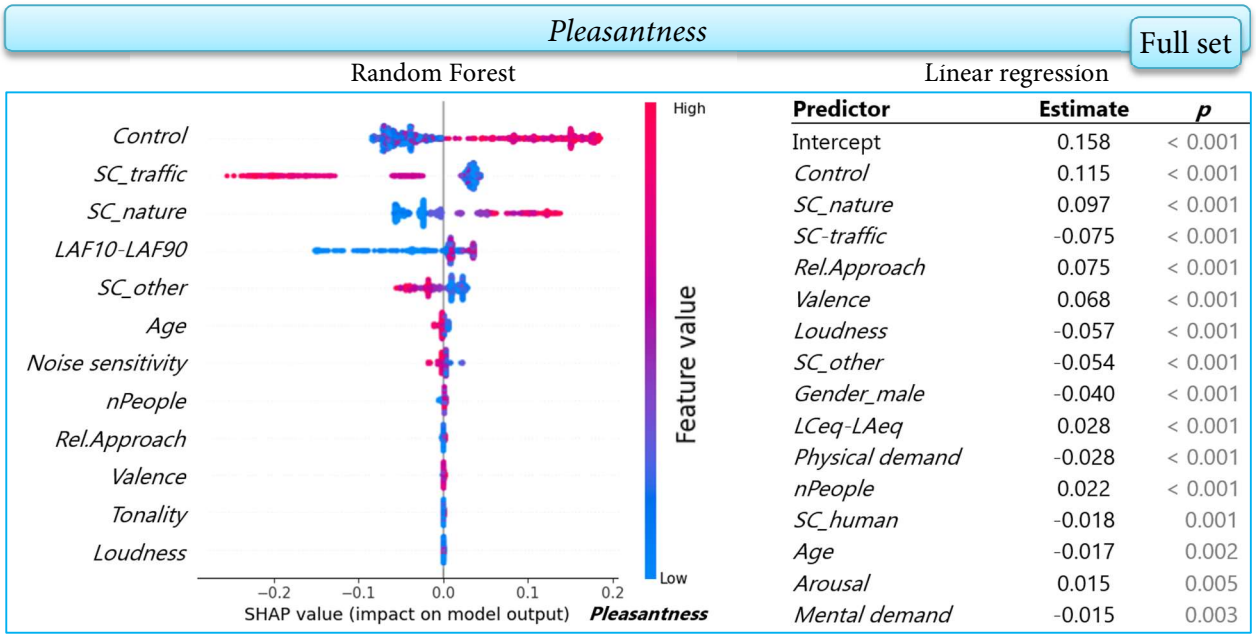
3.2. Predictor importances

Looking at Table 3, which shows the predictor importances, we observed for both *Pleasantness* models based on the entire dataset that perceived *Control* was the most important predictor, followed by the perceived intensity of natural sounds and traffic noise. Therefore, results suggest that the *Pleasantness* ratings are dominated by perceptual measures, assessed by the participants. Similar results can be observed for the *Pleasantness* quartile models. However, the order of the important predictors there was slightly different, and other perceptual (*Valence* in both Q1 models and the RF Q4 version) and acoustical measures (*Rel.Approach* as the most important predictor of both Q4 models) showed to be effective.

Table 3: Predictor importance, showing the three most important predictors in descending order for the models for both targets, both methods, and for the full set and all subsets. (Intercepts not shown.)

Target	Method	Q1	Q2	Q3	Q4	Full set
Pleasantness	**LR**	*SC_nature*	*SC_traffic*	*SC_nature*	*Rel.Approach*	*Control*
		Control	*SC_nature*	*Control*	*Control*	*SC_nature*
		Valence	*Control*	*SC_traffic*	*SC_traffic*	*SC_traffic*
	RF	*SC_nature*	*SC_nature*	*Control*	*Rel.Approach*	*Control*
		SC_other	*SC_traffic*	*Rel.Approach*	*Control*	*SC_traffic*
		Valence	*Control*	*SC_traffic*	*Valence*	*SC_nature*
Eventfulness	**LR**	*SC_human*	*SC_human*	*Gender_male*	*Rel.Approach*	*Rel.Approach*
		SC_traffic	*SC_traffic*	*SC_traffic*	*Noise sens.*	*SC_traffic*
		LAF10-LAF90	*Rel.Approach*	*Rel.Approach*	*SC_human*	*SC_human*
	RF	*LAF10-LAF90*	*LAF10-LAF90*	*LAF10-LAF90*	*Rel.Approach*	*Rel.Approach*
		SC_human	*Tonality*	*Tonality*	*Tonality*	*LAF10-LAF90*
		Rel.Approach	*SC_human*	*Rel.Approach*	*SC_human*	*Loudness*

Contrary, the *Eventfulness* model with the highest predictive power, based on RF and the full set, showed only acoustic measures as the three most important predictors. The *Rel.Approach* was first, followed by the second measure for time variability (*LAF10-LAF90*), and *Loudness*. Thus, results suggest that the *Eventfulness* ratings in the observed data are dominated by acoustic measures. Since the predictive power of most quartile models was low, we do not describe their dominant predictors.



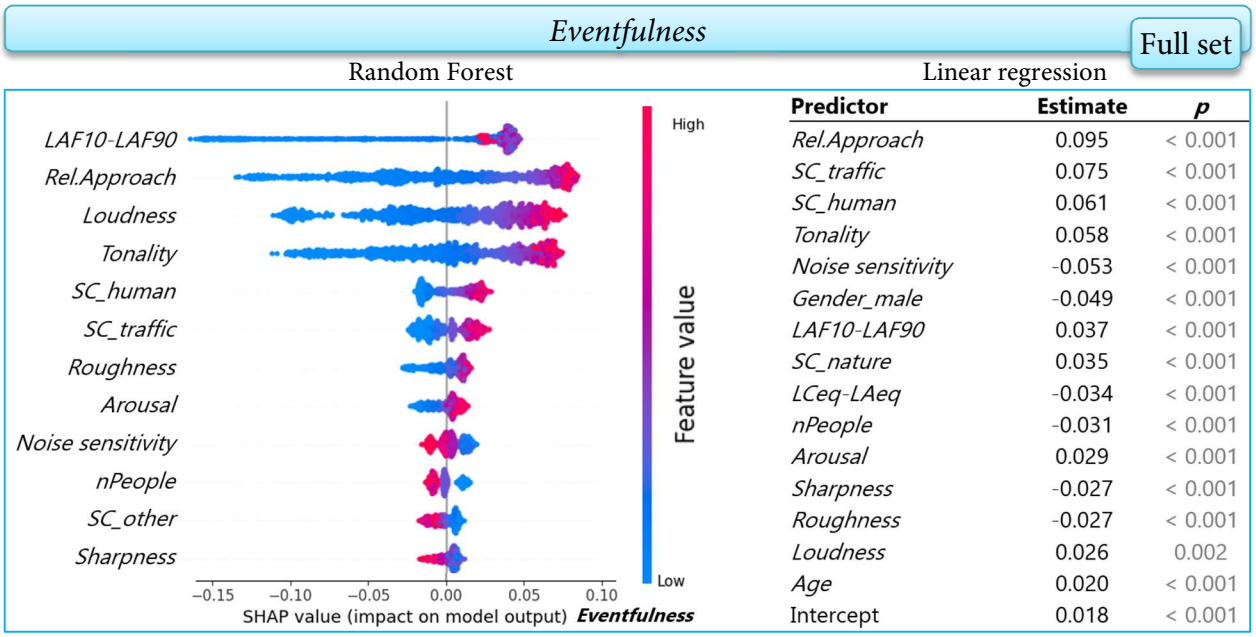


Figure 1: Importance and direction of effects for (median) full set models. RF: SHAP values [27] show the impact of the variable ("feature") values on the target. LR: Significant ($\alpha = 0.05$) standardized estimates are given in descending order (of their absolute values) together with their *p*-values.

To present the direction and the intensity of the effects, Figure 1 shows the significant estimates of LR and the SHAP values for the RF models for both targets (of the model instances with the median prediction error, which is why the order of the predictors can vary compared to these of Table 3 whose ranks were summed across all 5 CV instances). Expectedly, higher levels of perceived *Control* and

higher perceived intensity of natural sounds were associated with increasing *Pleasantness* values, while this was reversed for traffic noise. Looking now only at the RF model for *Eventfulness* (which offered the highest prediction performance), higher levels were, expectably, related to increasing levels of all acoustic and psychoacoustic measures.

3.3. Distribution of the targets, and cross-correlation coefficients of selected predictors

For the interpretation of the results above, some additional analysis of the target variables and the similarity of the acoustic predictors are useful and are therefore presented below.

Both standardized targets appeared nearly normally distributed (not shown here), both for the entire dataset and the four quartiles, suggesting that both soundscape dimensions were largely independent of the sound pressure level of the environment (exception: the Eventfulness model of the quiet subset Q1). This was corroborated by the uncorrelated variables *Pleasantness* and *LAeq* or *Loudness*, shown in Figure 2 based on the full set. Furthermore, *Pleasantness* ratings were not (or not noteworthy) correlated with any of the acoustical and psychoacoustical predictors used, which was observed for the full dataset and all quartile subsets within the correlation matrices. Contrary, some perceptual predictors (*Control, SC_nature*, and *SC_traffic*) were identified as weakly correlated with *Pleasantness,* again consistent over all five datasets. Additional weak correlations were found between *Eventfulness* and most of the acoustic and psychoacoustic predictors, indicating a certain effect of these predictors on the *Eventfulness* perception. Further results show that *Loudness* correlated expectably strong with the sound pressure level (*LAeq*), whereas moderate correlations were observed between *LAeq* and the *Rel.Approach* and most spectral content measures (*LCeq-LAeq*, *Roughness*, and *Tonality*). Interestingly, the aforementioned correlations were particularly evident for Q1 of the quieter soundscapes, as can be seen in Figure 3, while *Sharpness* showed weak negative correlations with *Roughness* and *LCeq-LAeq* in the upper three quartiles. In contrast to the target variable *Pleasantness*, correlation coefficients of acoustic and psychoacoustic measures were found to be not consistent over all five datasets for *Eventfulness* and tended to decrease for most quartile subsets, with the exception of Q1, compared to the full dataset. Finally, the correlation analyses also revealed that perceptual predictors showed no noteworthy correlation with *Eventfulness* across all 5 datasets.

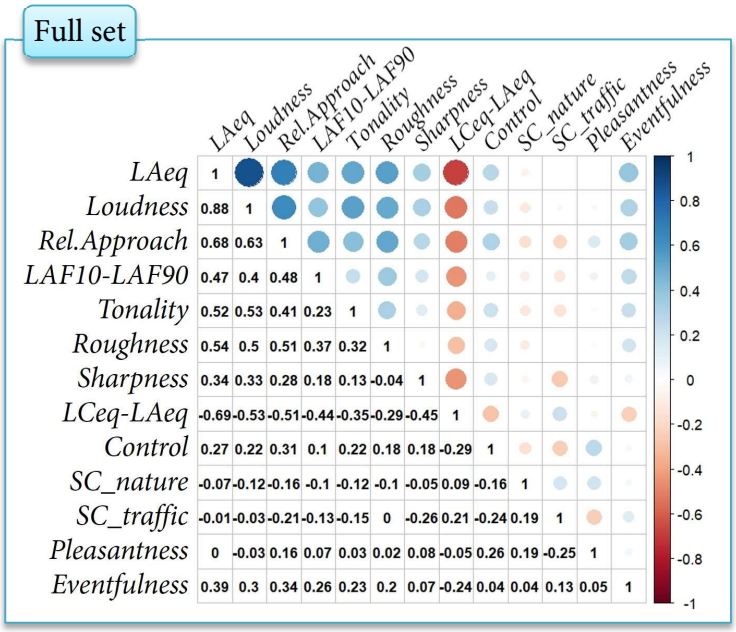


Figure 2: Pearson's cross-correlation coefficients for selected predictors, based on the entire dataset.

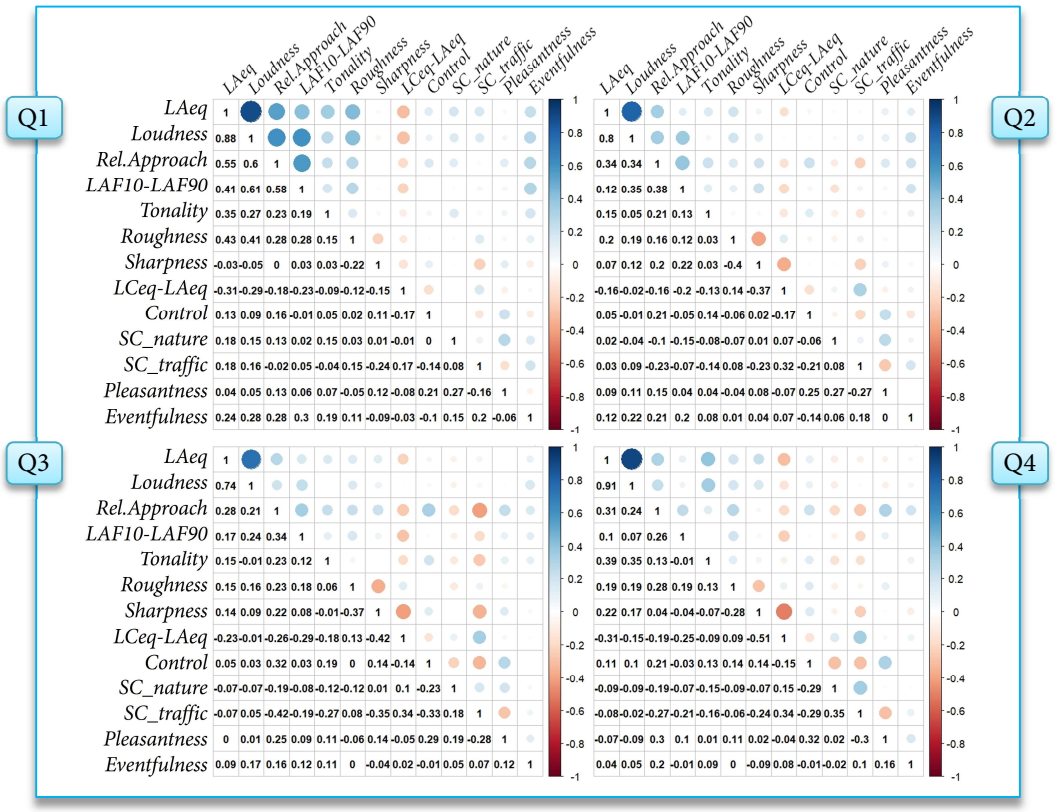


Figure 3: Pearson's cross-correlation coefficients for selected predictors, based on the *LAeq* quartile subsets Q1 to Q4.

Despite the moderate correlations, no multicollinearity issue was observed in all LR models, as indicated by small Variance Inflation Factor values (VIF < 3). The other (perceptual, situational, and person-related) variables not shown here had no or only weak correlations.

4. DISCUSSION

In this study on modeling the soundscape dimensions *Pleasantness* and *Eventfulness* for different quartiles of the sound pressure level distribution, we reanalyzed an existing indoor soundscape dataset for which participants recorded and rated the sonic environments they experienced in their homes. Acoustic, psychoacoustic, perceptual, situational, and person-related variables served as independent predictors for LR and RF models that were used to identify relevant predictors for both targets, the soundscape dimensions.

Despite the initial assumption that models based on *LAeq* subsets, the quartiles, can enable higher predictive performance in each quartile, they were found to offer no advantage over models based on the entire dataset as measured by their variance explanation or reduction ($R^2_{Pleasantness} \approx 20\%$ and $R^2_{Eventfulness} \leq 27\%$). Instead, most quartile models achieved even worse performance. No systematic reason could be found in the methods used (LR vs. RF), as there were cases where one or the other method was superior or where one or the other method performed notably worse. RF, which is known to be more sensitive to sample size compared to LR, performed similarly well and even better in case of *Eventfulness* if the entire dataset has been used. Finally, RF should preferably be used, as it is the method with the higher potential compared to LR and, therefore, also considers nonlinear relationships and complex interaction effects.

The general drop in performance of the *Eventfulness* models for the three louder quartiles could be attributed to the correlations between *Eventfulness* and the acoustic and psychoacoustic measures, which tended to decrease from moderate to loud sounds compared to the full dataset, as the formation of the quartiles splits, i.e., reduces the information available in the entire data set. For *Pleasantness* models, this reduction of information seems to be less relevant since model performances and correlations of perceptual features with the target were more stable across the quartile subsets compared to the full dataset with the exception of quiet sounds. Therefore, we conclude that building models based on quartile subsets might be more suitable for the prediction of *Pleasantness* and less so for *Eventfulness*, while for quiet sounds, however, the opposite seems to be case.

Since the four models based on the entire dataset yielded the highest performance for both soundscape dimensions, we now focus on these to discuss the importance of the predictors used. Results suggested that the ratings of *Pleasantness* were dominated by perceptual measures, whereas for (the superior RF model for) *Eventfulness* these were acoustic measures. Thus, compared to *Eventfulness*, *Pleasantness* appeared to be more susceptible to contextual influences and individual appraisal processes, which were not measured, and which are also difficult to measure altogether. However, the important predictors found in the full set models (*Pleasantness*: perceived *Control* and the intensity of natural sounds and traffic noise. RF *Eventfulness*: measures for time variability) were also found to be important in the quartile models and, thus, can be assumed to be robust important measures for the assessment of indoor soundscapes. Only for the quartile LR *Eventfulness* models, perceptual predictors were also found to be important in contrast to the full set model which could be caused by statistical suppression effects within the linear models since these results cannot be observed for the RF method.

In our study on the prediction of *Pleasantness* and *Eventfulness* for indoor soundscapes, we conclude, that here is no general advantage of models based on smaller volume ranges over models that use the full volume range. Consequently, models should be created using the entire available data. However, *Pleasantness* could also be investigated at smaller volume ranges, provided that perceptual measures serve as independent variables and that sounds are not quiet. Furthermore, our study points out that perceptual predictors seem to be crucial for modelling *Pleasantness* while a combination of acoustic and psychoacoustic measures are more important for predicting the soundscape dimension *Eventfulness*. Finally, our results were congruent across all measures used, namely in terms of the performance of the statistical models, the correlations between predictors and targets, and the importance of the predictors, which emphasizes the robustness of the study presented. Nevertheless, future research could further investigate underlying reasons for the different perception of quiet indoor sound environments in complex everyday context compared to louder soundscapes.

ACKNOWLEDGEMENTS

We would like to thank Fabian Rosenthal for programming the CV for the automated calculation of models of different statistical methods. A generalized version is available as the Python package *flexcy*: https://github.com/radlfabs/flexcv.

REFERENCES

[1] ISO. *Acoustics – Soundscape. Part 1: Definition and conceptual framework*. Standard DIN ISO 12913-1, International Organization for Standardization, Geneva, Switzerland, 2018.

[2] ISO. *Acoustics - Soundscape. Part 3: Data analysis*. Standard DIN ISO/TS 12913-3, International Organization for Standardization, Geneva, Switzerland, 2021.

[3] ISO. *Acoustics - Soundscape. Part 2: Data Collection and reporting*. Standard ISO/TS 12913-2, International Organization for Standardization, Geneva, Switzerland, 2018.

[4] A. Mitchell; T. Oberman; F. Aletta; M. Erfanian; M. Kachlicka; M. Lionello; J. Kang. *The International Soundscape Database: An integrated multimedia database of urban soundscape surveys—questionnaires with acoustical and contextual information.* https://zenodo.org/records/6331810, 2022. Last accessed 2024-04-12.

[5] F. Aletta, L. Molinero, A. Astolfi, S. Di Blasio, L. Shtrepi, T. Oberman and J. Kang. *Exploring associations between soundscape assessment, perceived safety and well-being: a pilot field study in Granary Square, London, Proceedings of 23th International Congress on Acoustics, ICA 2019,* 2019.

[6] C. Tarlao, D. Steele and C. Guastavino. *Assessing the ecological validity of soundscape reproduction in different laboratory settings, PloS one,* **17**(**6**), e0270401, 2022.

[7] K. Ooi, Z.-T. Ong, K. N. Watcharasupat, B. Lam, J. Y. Hong and W.-S. Gan. *ARAUS: A Large-Scale Dataset and Baseline Models of Affective Responses to Augmented Urban Soundscapes, IEEE Transactions on Affective Computing*, 1–17, 2023.

[8] S. Torresin, R. Albatici, F. Aletta, F. Babich, T. Oberman, S. Siboni and J. Kang. *Indoor soundscape assessment: A principal components model of acoustic perception in residential buildings, Building and Environment,* **182**, 107152, 2020.

[9] S. Versümer; J. Steffens; F. Rosenthal. *Extensive crowdsourced dataset of in-situ evaluated binaural soundscapes of private dwellings containing subjective sound-related and situational ratings along with person factors to study time-varying influences on sound perception — research data.* https://zenodo.org/record/7858848, 2022. Last accessed 2024-04-12.

[10] M. Lionello, F. Aletta and J. Kang. *A systematic review of prediction models for the experience of urban soundscapes, Applied Acoustics,* **170**, 107479, 2020.

[11] movisens GmbH, Karlsruhe, Germany. *movisensXS.* https://www.movisens.com/en/products/movisensXS/. Last accessed 2022-10-06.

[12] HEAD acustics. *ArtemiS SUITE,* HEAD acoustics, 2022.

[13] ISO. *Acoustics - Methods for calculating loudness - Part 1: Zwicker method*. Standard DIN ISO 532-1, International Organization for Standardization, Geneva, Switzerland, 2017.

[14] W. R. Bray. *The "Relative Approach" for Direct Measurement of Noise Patterns, Sound and Vibration,* **2004**(**09**), 20–22, 2004.

[15] DIN. *Measurement technique for the simulation of the auditory sensation of sharpness.* Standard DIN 45692, DIN Deutsches Institut für Normung e. V., 2009.

[16] ECMA International. *Psychoacoustic metrics for ITT equipment — Part 2 (models based on human perception)*. Standard ECMA-418-2, ECMA International, 2020.

[17] ECMA International. *Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment*. Standard ECMA-74, ECMA International, 2019.

[18] J. A. Russell. *A circumplex model of affect, Journal of personality and social psychology,* **39**(**6**), 1161–1178, 1980.

[19] S. G. Hart and L. E. Staveland. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research.* In *Human Mental Workload,* Elsevier, 1988.

[20] M. Schütte, A. Marks, E. Wenning and B. Griefahn. *The development of the noise sensitivity questionnaire, Noise and Health,* **9**(**34**), 15–24, 2007.

[21] D. Schreckenberg, F. Faulbaum, R. Guski, L. Ninke, C. Peschel, J. Spilski and J. Wothge. *Verkehrslärmwirkungen im Flughafenumfeld Endbericht, Band 3: Wirkungen von Verkehrslärm auf die Belästigung und Lebensqualität,* 2015. http://www.laermstudie.de/fileadmin/files/Laermstudie/Lebensqualitaet_Wiss_Ergebnisbericht.pdf, 2015. Last accessed 2024-04-12.

[22] World Health Organization. Regional Office for Europe. *Wellbeing measures in primary health care/the DepCare Project: report on a WHO meeting: Stockholm, Sweden, 12–13 February 1998.* Technical documents, World Health Organization. Regional Office for Europe, Stockholm, Sweden, 1998.

[23] B. Kulzer, N. Hermanns, T. Kubiak, M. Krichbaum and T. Haak. *Der WH0 5: Ein geeignetes Instrument zur Messung des Wohlbefindens und zum Depressionsscreening bei Diabetikern, Diabetologie und Stoffwechsel,* **1**(**A97**), 2006.

[24] G. van Rossum and F. L. Drake. *Python 3 Reference Manual,* CreateSpace, Scotts Valley, CA, 2009.

[25] 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 E. Duchesnay. *Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research,* **12**, 2825–2830, 2011.

[26] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama. *Optuna.* In A. Teredesai (Ed.) *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* Association for Computing Machinery, New York, NY, United States, 2019.

[27] L. S. Shapley. *A value for n-person games.* In H. W. Kuhn und A. W. Tucker (Eds.) *Contributions to the Theory of Games, Volume 2.* (AM-28), Princeton University Press, Princeton, NJ, 1953.

APPENDIX

Table 4: RF hyperparameters and their ranges for the models based on all observations (full set) and the *LAeq* quartiles (Qn).

Hyperparameter	**Range (Full set)**	**Range (Qn)**	**Precision**
ccp_alpha	[1e-8; 0.005]	[1e-8; 0.005]	Float
max_depth	[5; 50]	None	Integer
max_features	[1; 22]	[1; 22]	Integer
max_leaf_nodes	[10; 40000]	[10; 40000]	Integer
max_samples	[0.0005; 0.9]	[0.002; 0.9]	Float
min_samples_leaf	[1; 120]	[1; 120]	Integer, log
min_samples_split	[2; 100]	[2; 100]	Integer, log
min_weight_fraction_leaf	[0; 0.2]	[0; 0.2]	Float
n_estimators	[2; 8000]	[2; 2000]	Integer