Field Versus Lab: Situational Influences on Vehicle Sound Assessment,
David Maiberger, Ewald Strasser, Uwe Letens and Steven Van De Par,
Acta Acustica **vol. 105** (Number 2), 2019, pp. 401-411

DOI
https://doi.org/10.3813/AAA.919323

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 105 (2019) 401 – 411

DOI 10.3813/AAA.919323

Field Versus Lab: Situational Influences on Vehicle Sound Assessment

David Maiberger[1)], Ewald Strasser[2)], Uwe Letens[1)], Steven van de Par[2)]
[1)] Daimler AG, 71059 Sindelfingen, Germany. david.maiberger@daimler.com
[2)] Carl von Ossietzky University, Department of Medical Physics and Acoustics,
Acoustics Group, 26111 Oldenburg, Germany

Summary
Although it is agreed in literature that contextual aspects are effective moderators of sound quality evaluation, it is common practice to use results gained in laboratory studies to draw conclusions about customer experience in the real-life context. However, as the contexts differ considerably, such deductions are of potentially questionable value. For an effective improvement of experienced sound quality, understanding its evaluation process is crucial. For vehicle sound evaluation, a comprehensive model was presented in [1] that was obtained from data of a field study. However, it is unclear whether, and to what extent, the same model also can be transferred to listening experiments performed in a lab. In this work, a comparison between field and lab-test results of vehicle sound evaluation is carried out. Ratings of sound quality, comfort, sportiness, type conformity and make conformity were collected in a listening experiment under controlled conditions. The ratings were fed into the sound quality model of [1]. It was found that the sound quality model can also successfully describe lab-test results. However, an increased impact of comfort on sound quality was found, which is attributed to a greater focus of the participants on the mere acoustical aspects and an increased sensitivity to loudness in the lab. Certain specific driving conditions gain higher importance for comfort, such as standstill conditions or dynamic driving. As this can substantially influence the sound quality, the unreflected transfer of lab-test results to real-life contexts is not advised. Finally, evidence was found that the division of subjects into three driver types, as proposed in [2], makes meaningful distinctions with respect to the evaluation of vehicle sounds.

PACS no. 43.66.Ba, 43.66.Lj

1. Introduction

In research and development related to vehicle noise, vibration and harshness (NVH), the aim is to understand the relation between the emitted sound of a car and the subjective response to this sound [3, 4, 5, 6]. A thorough understanding of the evaluation process of vehicle sounds will assist NVH engineers in improving the sound quality experienced by the users in their everyday use.

However, sound quality as defined in [7] is not solely determined by properties of the sound. Along with auditory perception, also the *response-moderating factors* “cognition, action and emotion”, which are subject to “input from non-auditory modalities”, influence the judged product sound quality [7]. The “cognitive” situation encompasses aspects such as prior knowledge, expectation, experience, taste or social and cultural influences [7, 8]. Several authors point out the role of the spatial, time, semantic and response context of an experiment on the rated sound quality [9, 10, 11]. The semantic context, for example, is related to the meaning of the sound [12]. Additional visual information about the stimuli given in the instruction [13] or during the experiment [11] can have an impact on sound perception and judgement. The response context, as another example, involves the type of answering scale provided for the subjects, the selected set of stimuli presented in the experiment or the sequence of the sounds [9, 10]. The “actional” situation is characterized by the action a listener is engaged in when judging the product sound. For example, the user of a product is less annoyed by the product sound than a passive listener, as the latter has no control over the sound [8]. Furthermore, sound ratings can differ between evaluation in a listening lab as compared to a more realistic usage situation, where subjects are engaged in typical actions [14, 11]. Also the “emotional” situation of the subject influences sound quality evaluation. For example, a (positive or negative) frame of mind or attitude towards the product, established via priming tasks, mood or reading about others’ opinions, can significantly alter the sound quality judgements [15]. Sound quality evaluation can be influenced by “input from non-auditory modalities” [7]. For example, the simultaneous presentation of auditory and visual [11] or vibrotactile [16] stimuli can influence subjective reactions as com-

Received 1 March 2018,
accepted 18 December 2018.

pared to the case of unimodal auditory stimuli. Hence, it is crucial to understand how these non-auditory *context variables* affect sound quality evaluation of vehicle sounds. A valid model of the underlying subjective evaluation process is necessary to effectively improve sound quality.

A method frequently used to measure the sound quality of a car is by conducting listening experiments in a lab. This has several benefits, such as providing controlled conditions, requiring less time than a field test and thus enabling the coverage of a much broader spectrum of conditions. The disadvantage of lab-tests is the absence of the real-life evaluation context: neither do the listeners drive the car they listen to (so there is no feedback), nor do they really sit inside a car cabin. There are methods where subjects are placed in a car cabin or a virtual driving environment [17]. However, a great amount of studies are performed in ordinary listening labs. Lab experiments typically lack ecological validity, and thus, considering the context-dependency of sound quality, the transfer between lab-test results and real-life assessment as experienced by the customer may be problematic.

In a previous study, Maiberger *et al.* [1] developed a comprehensive model for vehicle sound quality evaluation from data of an exploratory, contextualized field study[1]. The model takes the acoustic factors *comfort*, *sportiness*, *type conformity* and *make conformity*, which are commonly treated in literature [3, 6, 19, 20], as input variables. These input variables are then related in a linear manner to *sound quality* as the output variable of the model. The non-acoustic contextual aspects *car* and *driver type* [2] moderate the strength of the relationship between input and output variables. According to [2], the driver type of a person is either the *sound-oriented fun-driver*, the *noise-sensitive aesthete* or the *sound-uninterested pragmatist*. Furthermore, it was found that comfort consists of two components that each relate to a specific set of driving conditions. These two components were found to be *dynamic comfort*, determined during overtaking or accelerating, and *static comfort*, determined when driving with constant speed. Sportiness, in contrast, could not be further resolved into separate components, but evidence was found that there is an emphasis on dynamic driving conditions [1]. As this model is based on data from a field study, it is assumed to be ecologically valid.

The aim of this paper is to answer the following two questions:

1. Can the model of [1], which was developed in the field, be transferred to the lab context?
2. Which differences are there between field and lab evaluation?

In order to answer these questions, a listening experiment was set up where cars were evaluated as a whole, comparable to the field situation. To be specific, each rated car was represented by a set of sound recordings in various driving conditions, and the car was identified to the evaluating subjects by name and picture. With the results, the model from [1] will be applied to the data in order to check its validity in the lab context. Afterwards, the regression coefficients will be compared, as the “strengths” of the individual dependencies.

Table I. Driving conditions used in the listening experiment.

No.	Driving condition	Location
1	Engine start	Exterior
2	Engine idle	Exterior
3	Engine idle	Interior
4	Partial load acceleration	Interior
5	Overtaking (partial-to-full load transition)	Interior
6	Accelerated pass-by	Exterior
7	Full load acceleration	Interior
8	Constant speed of 80 km/h, smooth road surface	Interior
9	Constant speed of 80 km/h, rough road surface	Interior
10	Constant speed of 140 km/h	Interior

2. Methods

2.1. Experimental setup

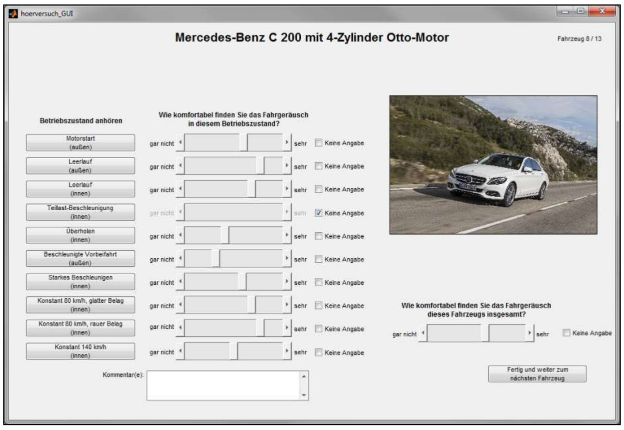
A listening experiment was conducted under laboratory conditions where subjects had to rate sound quality, comfort, sportiness, type conformity as well as make conformity of different cars. Each car was presented “as a whole”, i.e., as a set of sound recordings in 10 different driving conditions, such as engine start, idling, driving with constant speed or accelerating with various load levels. The selection of driving condition was made to resemble the selection used in [1]. Table I shows the driving conditions used in the experiment, for which a large sound database of different cars was available.

All sounds were recorded with an artificial head in the interior or the exterior of the vehicle. They were presented to the subjects with original sound pressure level over a calibrated setup consisting of an RME Fireface UC and Sennheiser HD 650 headphones. The graphical user interface (GUI) used to control the sound playback and to acquire the ratings was implemented in MATLAB [21]. It allowed to identify the vehicles to the subjects by make, name, motorization and a picture, in order to have the realistic frame of reference for the assessment (see Figure 1). Horizontal sliders ranging from “not at all” to “very much” were provided for each rating, as well as checkboxes for leaving particular ratings undecided (and greying out the corresponding slider). Subjects were provided with an additional comment field to optionally leave a comment for each presented car, for example in order to express peculiarities or thoughts. However, as these comments were very few, they did not help in the results analysis and will not be further detailed.

Thirteen *original cars* were included in the experiment, deliberately chosen to represent a broad range of vehicles of different German premium makes (i.e., Audi, BMW and

---

[1] As by the time of this submission reference [1] is still under revision, it should be noted that a preprint version of the manuscript can be found in [18].



hoerversuch_GUI

Mercedes-Benz C 200 mit 4-Zylinder Otto-Motor

Fahrzeug 8 / 13

Wie komfortabel finden Sie das Fahrgeräusch in diesem Betriebszustand?

Teillast-Beschleunigung (innen): ☑ Keine Angabe

Wie komfortabel finden Sie das Fahrgeräusch dieses Fahrzeugs insgesamt?

Figure 1. Layout of the graphical user interface used for the listening experiment. The interface for Pass 2, rating of comfort, is shown (cf. Section 2.2).

Mercedes) and car segments (i.e., compact car, mid-size (luxury) car, full-size luxury car, roadster, coupé, convertible, SUV), see Table II.

Since for nine cars the recordings of the constant driving conditions (8–10 in Table I) were not available, the corresponding recordings were taken from other cars. Depending on availability in the database, those replacement cars were chosen to be the same (or a follow-up) model (cars 1, 4, 5 and 9), the same model with a different engine (cars 3 and 7: Diesel instead of Otto), a comparable segment (car 10: SUV instead of sedan, car 13: mid-size luxury sedan instead of full-size luxury sedan, different engine), or the same segment of a different make (car 6: make C instead of B). Different engines were not regarded to be a problem, as the engine is not the dominant sound source in the considered constant driving conditions. No identical sounds were used for different cars, i.e., no replacements were taken from other cars used in this study. All replaced sounds were aurally checked by the authors for plausibility and similarity with the actual cars.

In addition, 12 *mismatched cars* were created by identifying the same sounds as a different car segment or car make (see Table III); in this context, “identifying” refers to the name and the picture shown to the participants in the GUI. To this end, the cars were divided into three major groups depending on their segment: *sports cars* (cars 5, 6, 7, 8, 9), *full-size luxury cars* (cars 10, 11, 12, 13) and all other cars that fall in neither of these categories (cars 1, 2, 3, 4); the latter group was termed *ordinary cars*. One car of make C was chosen from each group as the basis, i.e., cars 3, 5 and 12. These three cars were presented as each other (for the alteration of the car segment), making 6 mismatched cars. Furthermore, car 3 was presented as cars 1 and 2, car 5 was presented as cars 6 and 7, and car 12 was presented as cars 11 and 13 (for the alteration of the make), making another 6 mismatched cars. Just segment *or* make were altered, losing information about possible interaction between both categorical quantities, but drastically reducing the duration of the test. The mismatched car combinations are listed in Table III.

Table II. Original cars used in the listening experiment.

No.	Car segment	Cyl.	Engine type	Car make
1	Compact	4	Otto	A
2	Compact	4	Diesel	B
3	Mid-size	4	Otto	C
4	Mid-size luxury	6	Diesel	A
5	Roadster	4	Diesel	C
6	Roadster	4	Otto	B
7	Roadster	6	Otto	A
8	Roadster	6	Otto	C
9	SUV (coupé)	8	Otto	C
10	Full-size luxury (convertible)	8	Otto	B
11	Full-size luxury	8	Otto	B
12	Full-size luxury	8	Otto	C
13	Full-size luxury	6	Otto	A

Table III. Combinations forming the mismatched cars in the listening experiment; the numbers in the second and third column refer to the original cars from Table II.

No.	Sound of car no.	Identified as car no.
14	3	5
15	3	12
16	3	1
17	3	2
18	5	3
19	5	12
20	5	6
21	5	7
22	12	3
23	12	5
24	12	11
25	12	13

2.2. Procedure

The listening experiment consisted of five *passes* in total. Each pass consisted of a successive presentation (and rating) of all cars. When a participant had finished all ratings for a single car, he or she could proceed with the next car by pressing the designated button on the GUI. In order to prevent participants from forgetting ratings and to avoid logical inconsistencies (e.g., rating without prior listening), it was checked upon button press whether any of the sliders remained untouched or any sound was not played. If so, a dialog box appeared that asked the participant if he/she really wants to finish the rating of that car and proceed with the next car. This warning was integrated in order to hinder participants from skipping ratings, as it was not considered appropriate to force them to make a decision.

In Pass 1, the participants rated the overall sound quality for each of the 13 original cars, making 13 ratings (“What is your overall impression of this car’s driving sound?”, for once with the slider ranging from “very bad” to “very good”). In Pass 2, they rated the comfort in each of the ten driving conditions (“How comfortable is the driving sound to you in this driving condition?"), as well as the overall comfort ("How comfortable is the overall driving sound of this car to you?") for each of the 13 original cars, making $(10+1)\cdot 13 = 143$ ratings. In Pass 3, the sportiness in every driving condition ("How sporty is the driving sound to you in this driving condition?") as well as the overall sportiness ("How sporty is the overall driving sound of this car to you?") were rated by the participants for each of the 13 original cars, making $(10+1)\cdot 13 = 143$ ratings. The order of Passes 2 and 3 was swapped for half of the participants. In Passes 4 and 5, the participants rated the overall sound quality, overall comfort, overall sportiness, overall type conformity ("To what extent does the driving sound match the car's type?") and overall make conformity ("To what extent does the driving sound match the car's make?") for each of the 25 cars, making $5 \cdot 25 = 125$ ratings. The 25 cars were spread pseudo-randomly between Passes 4 and 5 for every participant, resulting in 12 and 13 cars per pass, respectively. As these passes contained both original and mismatched cars, the randomization was constrained so that a successive presentation of the same sound or image was avoided; in addition, Passes 4 and 5 both were split in two halves between which the participants fulfilled a short diversionary task. In total, each participant was required to give 424 ratings. The five passes were spread over four dates. The first date consisted of the first two passes, the remaining dates of one pass each.

2.3. Sample

The participants were recruited via notice boards at the University of Oldenburg. Since it was found in a previous study [1] that *sound-oriented fun-drivers* are rare amongst the university members, additional participants were recruited from a vintage car club, where a larger fraction of *sound-oriented fun-drivers* was expected. This was done in order to have enough participants representing each of the three driver types. The participants completed a screening questionnaire that contained certain statements for determining their driver type (cf. [2]). As can be seen from Table IV, the vintage car club members that took part split equally into *sound-oriented fun-drivers* and *noise-sensitive aesthetes*, who are the two driver types that care the most about vehicle sounds [2].

Table IV. Distribution of participants, acquired via university and a vintage car club, over the three driver types *sound-oriented fun-driver* (SF), *noise-sensitive aesthete* (NA) and *sound-uninterested pragmatist* (SP).

	SF	NA	SP	Total
University	4	8	9	21
Vintage car club	3	3	0	6
Total	7	11	9	27

The age of the 21 participants acquired via the university ranged from 21 to 64 years (Ø = 30). The six participants stemming from the vintage car club were between 49 and 69 years old (Ø = 59). In total, 13 females and 14 males with an average age of Ø = 37 years took part.

2.4. Statistical methods

2.4.1. General linear models

For the statistical analyses in this paper, several of the family of *general linear models* (GLM) are used. These models are generalizations of *(multiple) linear regression*, which takes one (or more) independent variables to predict one (or more) outcome variables with the *least squared error*. The branch of *analysis of variance* (ANOVA) allows for consideration of the influence of categorical predictor variables on an outcome variable. *Analysis of covariance* (ANCOVA) combines an ANOVA with linear regression, which allows for a multiple linear regression with *regression coefficients* b_i that vary depending on one (or more) grouping variables. In this paper, an ANCOVA will be used to formulate the model for sound quality, which depends on the (continuous) predictor variables comfort, sportiness and conformity, and weights that vary depending on the categorical variables *car* and *driver type*. The regression coefficients b_i estimated in ANCOVA reflect the amount of change in the model outcome per unit change of the i-th predictor variable. Similarly, the *standardized coefficients* β_i reflect "the number of standard deviations that the outcome will change as a result of one standard deviation change in the predictor" [22]. The *effect size* of each predictor will be represented by *partial* η^2, denoted as η^2_p.

Multivariate analysis of variance (MANOVA) is an enhancement of ANOVA that simultaneously tests for mean differences in more than one dependent variable (see [22] for details). The closely related *repeated measure MANOVA* (rmMANOVA) will be used to test for mean differences when the dependent variables are measured on the same subjects more than once. WILKS' *lambda* is used as the test statistic for deciding on the significance of rmMANOVA.

A significance level of $\alpha = 5\,\%$ is used in this study, against which the *probability value* p of each analysis will be compared. The p-value represents the probability of observing the given result just by chance, and hence p-values below 5 % suggest an actual relationship. In order to avoid familywise error inflation in post hoc analyses, the BONFERRONI *correction* is applied.

2.4.2. Pearson's correlation coefficient

Pearson's correlation coefficient R measures the amount of linear association between two variables (see [22]). For significance testing in this paper, the one-tailed probability (i.e., testing whether there is a significant *positive* relationship) rather than the two-tailed probability (i.e., testing if there is a significant positive *or* negative relationship) is evaluated (if not stated otherwise). Furthermore, the predictive power of a regression model is typically assessed by the portion of *explained variance* R^2.

2.4.3. Log-likelihood statistic

As the predictive power of a regression model increases with its *degrees of freedom* (df), i.e., the number of predictors, the -2 *log-likelihood* (-2 LL) was used to compare competing regression models in terms of their goodness of fit (see [22]). The difference Δ-2 LL between competing models can be tested for statistical significance, using a *chi-square distribution*; the df of the chi-square distribution are equal to the difference Δdf between the models.

2.4.4. Exploratory factor analysis

Exploratory factor analysis (EFA) is a method for grouping a set of variables into common factors, based on the correlation among the variables (for an overview over EFA, see [22]). EFA will be used in Section 3.3 to determine which driving conditions are being rated similarly with regard to comfort and sportiness, respectively. The adequacy of the data for being analyzed with EFA was determined with the KAISER-MEYER-OLKIN *coefficient* (KMO) and the *measure of sampling adequacy* (MSA). The number of factors was determined with the *scree test* and the KAISER-GUTTMAN *criterion*. The initial factors were extracted with *principal component analysis*, and the rotation was done according to the *Varimax*-criterion. For reliability considerations, CRONBACH's α was used.

3. Results

3.1. Reliability of the ratings

In this section, the question is tackled whether the collected data can be considered reliable across participants as well as across passes.

3.1.1. Between-subject agreement

In order to determine to what extent the participants agreed in their judgements, the ratings from all five passes were compared between subjects. To be specific, the ratings were examined for outliers in the sense of a *boxplot*, with the whiskers extending to a maximum of 1.5 times the *interquartile range* from the box's edges (see [23]). This was done for each given rating separately, resulting in 424 boxplots. It is observed that one participant accounts for almost 29 % of all of such outliers, which equals 12 % of this participant's ratings. The participant following with 14 % of all outliers (6 % of his ratings) is the one who also shows peculiarities in the analysis in the following section. The other participants each account for less than 8 % of all outliers. At this point, the first-mentioned participant with the highest percentage of outlying ratings was excluded from the analyses starting in Section 3.3. This participant is a *sound-uninterested pragmatist.*

3.1.2. Test-retest reliability

With the two successive ratings of overall sound quality (Passes 1 and 4/5), overall comfort (Passes 2 and 4/5) and overall sportiness (Passes 3 and 4/5) given by each participant, the test-retest reliability was evaluated. When correlating the earlier with the later ratings for every participant, the correlation coefficients of all except for five participants are significantly positive. For one out of these five participants the correlation coefficient is negative ($R = -.458$), while it is positive for all other participants ($R = .488$ on average); as this participant was the one with the second most outlying ratings in the previous section, this participant was also excluded from further analyses. Also this participant is a *sound-uninterested pragmatist.*

When just considering the remaining 25 participants, the average correlation coefficient $R = .500$ is still quite low and indicates a general difficulty of the participants to give reliable results. Therefore, a rmMANOVA was performed with their ratings of sound quality, overall comfort and overall sportiness as the dependent variables, indeed revealing a significant overall effect of the measurement ($\Lambda = .82$, $F(3, 312) = 23.16$, $p < .05$, $\eta_p^2 = .18$). This effect affects sportiness ($F(1, 314) = 69.56$, $p < .02$, $\eta_p^2 = .18$) but not sound quality ($F(1, 314) = 1.05$, $p > .02$) or comfort ($F(1, 314) = 0.19$, $p > .02$). It is found that the later ratings (Pass 4/5) of sportiness are higher. It must be pointed out that the tasks were different in the respective passes. The earlier ratings (Passes 1, 2 and 3) were made individually for each quantity, i.e., sound quality, comfort or sportiness. For comfort and sportiness, this overall rating was requested together with ratings in the driving conditions. That way the participants were forced to deal with these quantities on a more detailed level. The later ratings (Pass 4/5), however, were requested all together (along with type and make conformities), making it possible to just roughly tune into the sounds in the different driving conditions, and to distribute a general impression of a car over the ratings of all five quantities. Therefore, a low reliability measure does not necessarily mean a low reliability, but should be interpreted in the light of the experimental setup.

3.2. Influence of car identification

As described in Section 2.1, half of the cars presented in Passes 4/5 were mismatched cars in the sense that the same sounds were identified as different cars. The influence of this car identification, i.e., the presented name and picture, on the overall ratings of the cars was assessed by a MANOVA with the dependent variables sound quality, comfort, sportiness, type conformity and make conformity (all from Passes 4/5), and with two binary factors capturing whether the identification was type conform and whether the identification was make conform, respectively. For example, although the three cars 5, 18 and 20 consisted of the same sounds (namely that of car 5), the identification of car 5 was both type and make conform, the identification of car 18 was not type conform but make conform and the identification of car 20 was type conform but not make conform (see Tables II and III).

The overall effects of both type conform identification ($\Lambda = .94$, $F(5, 600) = 8.12$, $p < .05$, $\eta_p^2 = .06$) and make conform identification ($\Lambda = .94$, $F(5, 600) = 7.07$, $p < .05$, $\eta_p^2 = .06$) were found to be significant. Follow-up analyses showed that the type conform identification affects sportiness ($F(1, 604) = 7.10$, $p < .05$, $\eta_\mathrm{p}^2 = .01$) and type conformity ($F(1, 604) = 18.11$, $p < .05$, $\eta_\mathrm{p}^2 = .03$), but neither sound quality ($F(1, 604) = 0.02$, $p = .89$) nor comfort ($F(1, 604) = 2.94$, $p = .09$) nor make conformity ($F(1, 604) = 0.99$, $p = .32$). The make conform identification has a significant influence on sportiness ($F(1, 604) = 14.64$, $p < .05$, $\eta_\mathrm{p}^2 = .02$), on type conformity ($F(1, 604) = 10.20$, $p < .05$, $\eta_\mathrm{p}^2 = .02$) and on make conformity ($F(1, 604) = 5.74$, $p < .05$, $\eta_\mathrm{p}^2 = .01$). However, no effect was found on sound quality ($F(1, 604) = 0.04$, $p = .84$) or comfort ($F(1, 604) = 2.85$, $p = .09$).

It is observed that the non-conform identification of cars lowers the affected ratings. That means that an identification that is not type conform lowers type conformity, while an identification that is not make conform lowers make conformity, which could consistently be expected. However, an identification that is not make conform also lowers type conformity. It might be easier for the participants to rate type conformity than make conformity, and that therefore the non-conform identification also manifests in type conformity. Finally, both kinds of non-conform identification lower the sportiness rating. It is possible that sportiness is one of the major acoustical distinguishing features of car types and car makes. All of the mismatched cars' sounds stemmed from make C, which is generally considered rather comfortable than sporty. Therefore, the identification as another make might result in a general lack of sportiness with regard to participants' expectation, and therefore a lower sportiness rating.

3.3. Factor analysis of driving conditions

In [1], a distinction between *dynamic comfort* and *static comfort* was found, which differ in the driving conditions they relate to. In this section, an EFA is performed in a similar manner to group the driving conditions with regard to correlations between the ratings of comfort and sportiness, respectively.

3.3.1. Comfort

An EFA was carried out in order to find groups of driving conditions that form common factors of rated comfort. With a KMO of .826 and MSA between .752 and .889, the data were considered adequate for the analysis. Scree test and Kaiser-Guttman criterion pointed at three factors, which together explain 79.9% of the observed variance. The rotated factor loadings for the comfort ratings are shown in Table V. The factors can be interpreted as 1) *standstill comfort*, 2) *static comfort* and 3) *dynamic comfort*.

A reliability analysis showed that "engine start", "engine idle (e)", "engine idle (i)" and "partial load acceleration" are the most reliable items for *standstill comfort*. For *static comfort*, "constant speed 80 km/h" on both smooth and rough road surface are the most reliable items, and exclusion of "constant speed 140 km/h" led to a better reliability. For *dynamic comfort*, "overtaking", "accelerated pass-by" and "full load acceleration" are all reliable items.

Table V. Rotated factor loadings of the comfort ratings, together with communalities (Com.) and explained variances. (e) and (i) denote exterior and interior sound recordings, respectively. Loadings < .4 are printed in grey; bold values indicate the reliable items forming each factor.

	1	2	3	Com.
Engine start (e)	**.833**	-.023	.291	.779
Engine idle (e)	**.859**	.001	.264	.808
Engine idle (i)	**.842**	.256	.034	.776
Part. load acc. (i)	**.694**	.394	.168	.666
Overtaking (i)	.296	.360	**.757**	.791
Acc. pass-by (e)	.179	.070	**.887**	.824
Full load acc. (i)	.203	.350	**.815**	.828
Const. 80 smooth (i)	.249	**.875**	.133	.846
Const. 80 rough (i)	.162	**.915**	.189	.899
Const.140 (i)	-.019	.827	.303	.776
% variance explained	28.7	27.7	23.5	
Cronbach's α	.867	.922	.874	

To examine how the three factors relate to rated overall comfort, a regression analysis was performed with the three factors as predictors to estimate overall comfort. For this purpose, each factor is represented by the average rating of its reliable items. For example, *static comfort* is the average of the comfort ratings of "constant speed 80 km/h, smooth road surface" and "constant speed 80 km/h, rough road surface". In the regression analysis, all three factors *standstill comfort* ($b = 0.275$, $\beta = .236$, $p < .05$, $\eta_\mathrm{p}^2 = .120$), *static comfort* ($b = 0.340$, $\beta = .288$, $p < .05$, $\eta_\mathrm{p}^2 = .172$) and *dynamic comfort* ($b = 0.495$, $\beta = .506$, $p < .05$, $\eta_\mathrm{p}^2 = .366$) were found to be statistically significant, with $R^2 = .704$. From the standardized regression coefficients β_i it can be concluded that overall comfort is best predicted by *dynamic comfort*, followed by *static comfort* and *standstill comfort*.

The factor structure is similar to the factors found in the previous field study [1]. However, as the first factor, *standstill comfort*, was not observed there it seems that this factor is less relevant in the real driving context. The strong influence of *dynamic comfort* on overall comfort is contrary to the significance of the predictor *static comfort* that was found in [1]. However, the N_5 diffuse field loudness[2] [24] of the driving conditions comprising dynamic comfort is higher than for the driving conditions comprising static comfort, as can be seen from Table VI. Since higher loudness probably leads to lower comfort, the dynamic factor might develop a greater impact on comfort, as participants in a purely acoustical test tend to overestimate loudness as compared to a close-to-reality context [14].

3.3.2. Sportiness

As in the previous section, also for the sportiness ratings an EFA was carried out. The KMO was .829 and MSA

---

[2] The N_5 value is the loudness value exceeded only 5 % of the analysis duration, i.e., the 95th percentile.

Table VI. N_5 diffuse field loudness ranges and means for the driving conditions comprising static and dynamic comfort.

Factor	Driving condition	N_5 in sone (GD) Min	Max	Mean
Static comfort	Const. 80 smooth (i)	12.6	22.3	17.4
	Const. 80 rough (i)	17.0	29.3	23.0
Dynamic comfort	Overtaking (i)	18.6	46.1	31.1
	Acc. pass-by (e)	29.3	122.9	49.6
	Full load acc. (i)	24.4	56.9	36.7

Table VII. Rotated factor loadings of the sportiness ratings, together with communalities and explained variances. (e) and (i) denote exterior and interior sound recordings, respectively. Loadings < .4 are printed in grey; bold values indicate the reliable items forming each factor.

	1	2	Com.
Engine start (e)	.507	.531	.539
Engine idle (e)	**.778**	.325	.711
Engine idle (i)	**.719**	.022	.518
Part. load acc. (i)	**.683**	.341	.583
Overtaking (i)	.225	**.888**	.839
Acc. pass-by (e)	.162	.839	.730
Full load acc. (i)	.070	**.912**	.837
Const. 80 smooth (i)	**.868**	.085	.760
Const. 80 rough (i)	**.894**	.134	.818
Const. 140 (i)	**.674**	.215	.500
% variance explained	39.3	29.0	
Cronbach's α	.881	.916	

ranged from .733 to .958, which is why the dataset was considered adequate for the analysis. The scree test hinted at two to three factors, while the Kaiser-Guttman criterion suggested two factors; as the latter criterion is usually considered an upper bound [22, p. 641], just two factors were extracted explaining 68.3% of the observed variance. The rotated factor loadings for the sportiness ratings are shown in Table VII. Similar to the second and third comfort factor, the two factors can be interpreted as 1) *static sportiness* and 2) *dynamic sportiness*.

For *static sportiness*, all items "engine idle (e)", "engine idle (i)", "partial load acceleration", "constant speed 80 km/h" on both smooth and rough road surface and "constant speed 140 km/h" are reliable. For *dynamic sportiness*, exclusion of "accelerated pass-by" leads to a higher Cronbach's α for the remaining items "overtaking" and "full load acceleration". As the driving condition "engine start" did not show clear-cut loadings, it was not further analyzed.

In a regression analysis to predict overall sportiness, both factors *static sportiness* (b = 0.299, β = .211, p < .05, η_p^2 = .088) and *dynamic sportiness* (b = 0.794, β = .675, p < .05, η_p^2 = .497) were found to be statistically significant, with R^2 = .617. It is concluded that overall sportiness is mainly determined by *dynamic sportiness*.

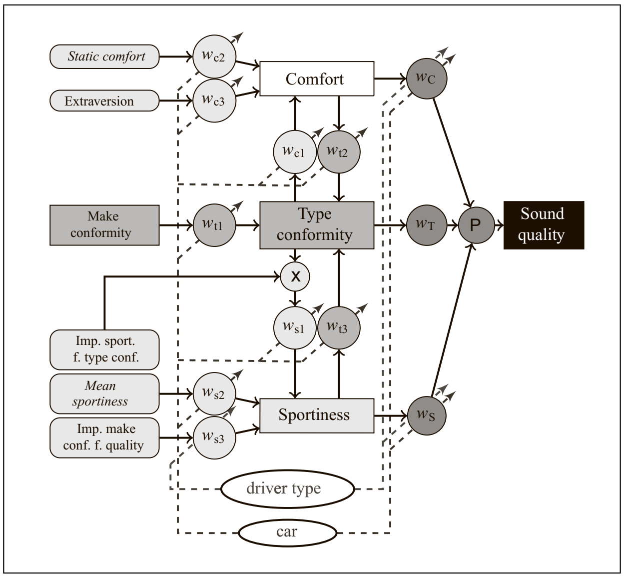


Figure 2. Model for sound quality assessment, taken from [1].

In the previous field study, just a single factor of sportiness was found, but it was hypothesized that this was just due to the segment of sports cars being underrepresented [1]. As the significance of the dynamic driving conditions for sportiness was concluded from the participants' answers in [1], the differentiation of dynamic against static aspects of sportiness seems reasonable. This is further underlined by the weight put on this factor in the regression analysis.

3.4. Validation of sound quality model

In [1] a model for vehicle sound assessment was proposed. The model was developed to predict the overall sound quality of a car based on both ratings of individual sound attributes as well as inclusion of context variables, such as the driven car and the driver type. The model was developed based on the data of an extensive field study where subjects drove six different cars and rated various perceptual attributes, such as comfort, sportiness, type conformity, make conformity and sound quality. The model is visualized in Figure 2.

The model is essentially a linear model that applies weights w_i to the input variables to predict the three constituent components of sound quality, i.e., comfort, sportiness and type conformity. These three components themselves are then weighted linearly to predict the overall sound quality. To obtain an accurate modeling it was found that some weights needed to be adapted such that they depend on the driver type and the car, which was realized with ANCOVA models.

In the current paper only the rightmost part of Figure 2 will be considered. This part describes the formation of sound quality from comfort, sportiness and type conformity depending on the car and the driver type, and will be denoted as the *sound quality model*. In order to assess the predictive power of the sound quality model in the context of a laboratory study, it is applied to the data from the listening experiment. Just like the research questions posed in the introduction, also the validation will be twofold: in the first step, the overall prediction accuracy will be examined in terms of explained variance R^2 and log-likelihood statistic -2 LL. In the second step, the resulting models will be compared in terms of their parameters, which describe the dependencies within the model in both contexts. For these analyses, both ratings given by the subjects with respect to sound quality, overall comfort and overall sportiness, respectively, were averaged in order to increase the robustness of the yielded scales.

3.4.1. Applicability of sound quality model

The sound quality model explains $R^2 = 80.6\,\%$ of the variance in the listening experiment, and all of its constituents (except for the intercept, i.e., the constant offset) are statistically significant (see Table VIII). Compared to the field data from [1] where $R^2 = 82.5\,\%$, this is considered a good result given that the participants, the cars as well as the contexts were different.

For the sake of generalizability, however, it seems adequate to categorize the cars rather than referring to the 25 individual (original and mismatched) cars. Therefore, the three car segments *sports cars*, *full-size luxury cars* and *ordinary cars* (as defined in Section 2.1) were distinguished. In order to be in the participants' reference frame, the fixed factor "car" in the original model was then replaced by the segment as identified to the participants rather than the segment of the played sound. The resulting model can account for $R^2 = 74.0\,\%$ of the variance; however, even considering the smaller number of degrees of freedom of this model, it is significantly worse than the original model (Δ -2 LL = 177.6, Δdf = 132, $p < .05$). In a similar manner, the categorization of cars into the three makes A, B and C (see Table II) was tested by replacing the factor "car" by the make (as identified to the participants). The resulting model ($R^2 = 72.6\,\%$) is also significantly worse than the model that considers each individual car (Δ -2 LL = 210.1, Δdf = 132, $p < .05$). The replacement of the car by the interaction between segment and make (both as defined before) also yields a model that is significantly worse than the initial model ($R^2 = 75.4\,\%$, Δ -2 LL = 143.0, Δdf = 96, $p < .05$).

By examining the data, however, a promising categorization was found that groups the cars into the following five segments: *compact cars* (cars 1, 2, 16, 17), *mid-size (luxury) cars* (cars 3, 4, 18, 22), *full-size luxury cars* (cars 11, 12, 13, 15, 19, 24, 25), *sports cars* (cars 5, 6, 7, 8, 10, 14, 20, 21, 23) and *SUV* (car 9). In comparison to the earlier categorization of the car type from Section 2.1, the ordinary cars have split in two groups. Car 10, which is a coupé, switched from a full-size luxury car to a sports car. Car 9 dropped out of the sports cars and forms its own group. These categories are well in line with the ones used in [2]. However, it should be noted that these five categories are unbalanced in the present study, e.g., there is only one SUV, but nine cars presented as sports cars. The model resulting from this categorization can explain $R^2 = 75.7\,\%$ of the variance, and is as good as the model that includes the individual cars (Δ -2 LL = 136.3, Δdf = 120, $p = .147$). The model terms are listed in Table IX.

Table VIII. Contributions to the sound quality model with the individual cars; $R^2 = .806$, $R^2_{\text{adj}} = .741$.

Source	p-value	η_p^2
Intercept	.103	.006
car * driver type * comfort	.000	.571
car * driver type * sportiness	.000	.260
type conformity	.000	.113

Table IX. Contributions to the sound quality model with five car segments; $R^2 = .757$, $R^2_{\text{adj}} = .744$.

Source	p-value	η_p^2
Intercept	.061	.006
segment * driver type * comfort	.000	.521
segment * driver type * sportiness	.000	.169
type conformity	.000	.149

Table X. Effect sizes η_p^2 of sound quality model applied to field test data [1] and lab-test data (this study), respectively.

	η_p^2	
Predictor	Lab	Field
Intercept	.006	.270
segment * driver type * comfort	.521	.249
segment * driver type * sportiness	.169	.242
type conformity	.149	.195

To sum up, the sound quality model that was developed in the field is also applicable in the lab context. This was concluded from a similarly high portion of explained variance R^2 in both studies, which means that the model describes the ratings reasonably well. Hence, participants do not seem to have substantially different evaluation schemes in both contexts.

3.4.2. Equality of sound quality models

Up to now the principal structure of the sound quality model (as found in [1]) has proven successful for describing the data of the laboratory experiment. However, if these models are to be considered equal across contexts, also the portion of explained variance η_p^2 of each predictor as well as the regression coefficients b_i should be of comparable magnitude. To compare the models, the data from [1] were reanalyzed with the five car segments established in the previous section; accordingly, one *compact car*, one *sports car* and four variants of a *mid-size (luxury) car* were considered for [1].

The effect sizes η_p^2 of both sound quality models are given in Table X. It is observed that η_p^2 of the sportiness term and type conformity are quite similar between the studies, but the comfort term has a much greater influence in the lab-test than in the field test.

The regression coefficients b_i, which give the amount of change in predicted sound quality per unit change of the predictors (i.e., comfort, sportiness or type conformity) for each combination of segment and driver type, are shown in Figure 3 together with their confidence intervals. The sign of the significant regression coefficients b_i can be used to determine the direction of the influence of the model predictors. For the lab experiment it is found that a higher type conformity generally leads to higher sound quality. Except for *sound-oriented fun-drivers* in the SUV, higher comfort always leads to higher sound quality. For *sound-oriented fun-drivers* in all but the *mid-size (luxury) cars*, higher sportiness leads to higher sound quality; for *noise-sensitive aesthetes* in *compact cars* and *sports cars*, higher sportiness also leads to higher sound quality; for *sound-uninterested pragmatists* in *sports cars*, however, higher sportiness leads to lower sound quality.

It is observed that for almost all coefficients the confidence bounds between studies overlap, implying that the corresponding values do not differ between lab and field, statistically speaking. The only exceptions are the intercept, the regression coefficients of comfort for *sound-oriented fun-drivers* and *noise-sensitive aesthetes* in *sports cars* and the regression coefficient of sportiness for *noise-sensitive aesthetes* in *compact cars*, which therefore are not considered equal between studies. The greater intercept in the field study means that the participants gave higher sound quality ratings on average. The two regression coefficients related to comfort that are significantly greater in the lab imply that comfort is more important there than in the field context, which has already been concluded earlier from the larger effect size η_p^2 of comfort. The sound quality of *compact cars* benefits from sportiness in the lab experiment but obviously suffers from it in the field context, as indicated by the negative sign of the corresponding coefficient for *noise-sensitive aesthetes*. It seems that the role of sportiness can change due to the evaluation situation as already indicated by the low test-retest reliability for the sportiness ratings in Section 3.1.2.

4. Discussion

This paper investigated to what extent a model developed to predict vehicle sound quality using field tests (cf. [1]) can be transferred to the listening lab. Listening tests coupled with an extensive analysis showed that this is indeed possible. The sound quality model was equally successful in the lab as compared to the field, being able to explain $R^2 = 75.7\,\%$ of the variance observed in the lab. However, the findings also demonstrate that there is a shift in the importance of certain aspects between lab and field experiments.

Furthermore, it was found that participants judged differently depending on the specific task. For example, overall sportiness was rated lower when requested together with the individual sportiness ratings for each driving condition (Pass 3), as compared to being rated together with sound quality, comfort, type and make conformity in Pass 4/5. The different response contexts in the passes, i.e., the specific sets of ratings requested together, may have led to this discrepancy. For example, in Passes 2 and 3 just a single acoustic attribute (comfort or sportiness, respectively) was requested, while in Passes 4 and 5 the participants were asked to rate their overall impression of all five acoustic attributes at once. In addition, the participants were obliged to listen to all provided driving conditions in Passes 2 and 3 as they had to give a rating for each of these, while in Passes 4 and 5 no ratings of the individual driving conditions were requested. This difference between “local perception” and “global perception” might be a form of a more general principle of perception. For example, it has been shown that attending to local or global properties of sounds can result in different auditory processing and responses [25]. Depending on the task, the participants’ focus can differ, resulting in different ratings of individual aspects.

It was found that mismatched cars, i.e., cars that have been identified to the participants as different cars, are rated lower with respect to type conformity and make conformity, but also with respect to sportiness. It is indicated that participants can give consistent ratings upon the conformity of the cars, which depends on their expectation. This expectation also seems to affect the given sportiness ratings. However, no direct influence on sound quality was found.

A factor analysis of the comfort ratings in different driving conditions revealed the factors *standstill comfort*, *static comfort* and *dynamic comfort*. As the factor *standstill comfort* was not observed in the field study [1], it might just form in the context of this laboratory experiment where the corresponding stimuli are presented in the same manner as all other driving conditions, lifting them on a higher level of notice. It is therefore questionable if this factor is of practical relevance, especially against the background of spreading automatic start-stop systems and electric starters in modern cars. *Dynamic comfort*, *static comfort* as well as *standstill comfort* contribute to overall comfort in the lab, in contrast to the field study where *static comfort* was finally found to be the best predictor. In the present study, the dynamic driving conditions were louder than the static driving conditions. As Steffens found in [14], participants in a purely acoustical test tend to overestimate loudness of household appliances as compared to a close-to-reality context, which results in lower rated pleasantness. Since higher loudness probably leads to lower comfort, an overestimation of loudness might result in a considerably higher weight of the dynamic driving conditions for comfort as compared to the field context. A factor analysis of the sportiness ratings yielded two factors entitled *static sportiness* and *dynamic sportiness*, which were already assumed in [1]. Accordingly, the rating of overall sportiness is mainly dependent on *dynamic comfort*. Hence, care must be taken when interpreting results from lab-tests that involve standstill conditions, such as engine

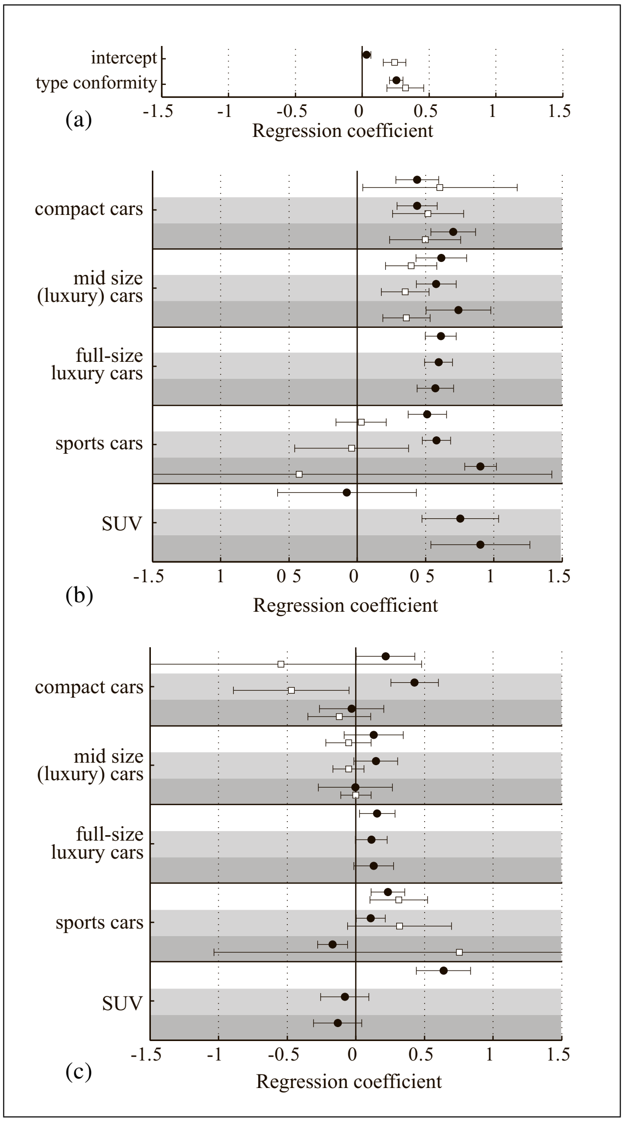


Figure 3. Regression coefficients b_i of the sound quality model, for the lab-test (●) and the field test data (□), respectively. Estimated means and 95 % confidence intervals are shown. In panels (b) and (c), white regions correspond to *sound-oriented fun-drivers*, light grey regions correspond to *noise-sensitive aesthetes* and dark grey regions correspond to *sound-uninterested pragmatists*. (a) Intercept and regression coefficient b_i for type conformity, (b) Regression coefficients b_i for comfort, (c) Regression coefficients b_i for sportiness.

start or idling. These driving conditions gain higher importance in a listening lab due to their mere presence, but seem negligible for overall assessment in the field. Also dynamic driving conditions gain greater importance for comfort than they do in the field because of increased sensitivity to loudness.

When applying the sound quality model from [1] to the data of this study, a higher importance of comfort was found. This was indicated by the considerably larger effect size of comfort than in the field experiment, and was further emphasized by the significantly positive regression coefficients in the lab for *sound-oriented fun-drivers* and *noise-sensitive aesthetes* in *sports cars*. As discussed before, the missing context of sitting in a car and the focus on the mere sound in the lab makes the participants more sensitive to vehicle noise, and thus more focused on comfort.

The negative contribution of sportiness to sound quality of *compact cars* observed in the field was opposed by the positive impact in the lab-test situation. As discussed above, also the specific task given to the subjects can alter their responses, as seen by the low test-retest reliability of the sportiness ratings. The rating of sportiness in the field might have turned out differently because it was requested retrospective, in contrast to the rating of sportiness in the lab where the driving conditions were available all at once during evaluation.

A generally higher rated sound quality in the field context was indicated by the greater intercept. Similarly, in [14] a generally higher pleasantness rating was found (for household appliances) in a close-to-reality context as compared to a lab study. Along with the overestimation of loudness in the lab, this was attributed to distraction in the real-life context, which can also be assumed for vehicle sound evaluation. Hence, care must be taken when extrapolating lab results to real-life assessment.

As in [1], the driver typology from [2] has proven to make meaningful distinctions with regard to sound evaluation. Sportiness was rated very differently across segments and driver types. Most notably, *sound-oriented fun-drivers* appreciate sportiness in almost every car, while *noise-sensitive aesthetes* differentiate between segments. *Sound-uninterested pragmatists* have shown to be quite insensitive to sportiness, which is in accordance with the observations made in [1]; *sports cars* even get penalized for their sporty sound character in the lab, maybe due to being perceived as noisy and thus lowering their comfort. An interesting case is car 9, which in this study is the only car in its segment and thus has its “own” regression coefficients. This car, however, is a mixture of SUV and coupé, and it belongs to the sports car division of make C. The sound quality of this car is determined by comfort for *noise-sensitive aesthetes* and *sound-uninterested pragmatists*, but by sportiness for *sound-oriented fun-drivers*. It is possible that the *sound-oriented fun-driver*, who is assumed to be especially interested in cars, is the only driver type who recognizes this “crossover” car, and appreciates its sporty character; in contrast, the other two driver types both require it to be comfortable, most probably seeing it as a standard SUV. Three participants expressed after the experiment that there was a mismatch between some cars and the respective sounds; all of those have been *sound-oriented fun-drivers*. This was not a systematic measurement, but nevertheless gives a hint that the driver types differ in their ability to recognize mismatched cars. Furthermore, the two participants excluded due to inconsistent ratings have been *sound-uninterested pragmatists*, who are generally considered uninterested in vehicle sounds. The inconsistent ratings might originate from their insensitivity to certain aspects of vehicle sounds. Hence, it is emphasized that such a distinction of subjects into driver types should be considered or even extended in future research dealing with vehicle sound quality and metric development.

The results presented in this paper demonstrate that the transfer between lab and field studies is not straightforward. Sound quality metrics, which are usually based on lab experiments, can at best predict what they were trained with: lab-test evaluation. However, extrapolation of these results to field evaluation, or even “customer satisfaction”, is doubtful. It may well happen that the evaluation process itself is similar in both contexts, as demonstrated by the equally high predictive power of the same sound quality model in the lab and in the field. However, the specific dependencies inside the model (which can be viewed as the “paths” depicted in Figure 2) can have different weight, and their effect can even reverse under certain conditions. This was demonstrated by the different effect sizes, and the regression coefficient that switched its sign, which of course will have implications when optimizing the sound character of a car.

However, there was also much congruence between the models in terms of their general applicability and most of the regression weights. This goes along with the “common practice” of acoustic engineers to use metrics developed in the lab to objectify customers’ “real-world” evaluations. This approach, although not accurate, can in fact capture a decent amount of variance that engineers are actually interested in. Though, it is essential for every user of sound quality metrics to be aware of their restricted scope.

Since only a limited range of German premium makes was considered in this work, it will be interesting to verify the findings of this paper with other makes in future research.

References

[1] D. Maiberger, E. Strasser, U. Letens, S. van de Par: Contextual aspects in subjective vehicle sound assessment. Acta Acust united Ac **105** (2018) submitted.

[2] D. Maiberger, U. Letens, R. Weber, S. van de Par: Individual influences on the evaluation of vehicle sounds: a typology of premium car drivers with regard to attitude towards cars and sounds. Acta Acust united Ac **104** (2018) 509–520.

[3] R. Bisping: Car Interior Sound Quality: Experimental Analysis by Synthesis. Acta Acust united Ac **83** (1997) 813–818.

[4] D. Västfjäll, M.-A. Gulbol, M. Kleiner, T. Gärling: Affective evaluations of and reactions to exterior and interior vehicle auditory quality. Journal of Sound and Vibration **255** (2002) 501–518.

[5] A. Sköld, D. Västfjäll, M. Kleiner: Perceived Sound Character and Objective Properties of Powertrain Noise in Car Compartments. Acta Acust united Ac **91** (2005) 349–355.

[6] A. Zeitler, P. Zeller: Psychoacoustic Modelling of Sound Attributes. SAE Technical Paper 2006-01-0098, 2006, doi: 10.4271/2006-01-0098.

[7] J. Blauert, U. Jekosch: Sound-Quality Evaluation — A Multi-Layered Problem. Acta Acust united Ac **83** (1997) 747–753.

[8] M. Bednarzyk: Qualitätsbeurteilung der Geräusche industrieller Produkte — Der Stand der Forschung, abgehandelt am Beispiel der Kfz-Innenraumgeräusche. Doctoral thesis, VDI-Verlag, Düsseldorf, 1999.

[9] A. Zeitler, J. Hellbrück, W. Ellermeier, H. Fastl, G. Thoma, P. Zeller: Methodological approaches to investigate the effects of meaning, expectations and context in listening experiments. Proceedings of Internoise, Honolulu, Hawaii, USA, 2006.

[10] R. Guski, J. Blauert: Psychoacoustics Without Psychology? Fortschritte der Akustik — NAG/DAGA 2009, Rotterdam, 1550–1551.

[11] A. Fiebig: Influence of context effects on sound quality assessments. Proceedings of Euronoise 2015, Maastricht, 2015, 2555–2560.

[12] U. Jekosch: Meaning in the Context of Sound Quality Assessment. Acta Acust united Ac **85** (1999) 681–684.

[13] P. Susini, O. Houix, N. Misdariis, B. Smith, S. Langlois: Instruction’s effect on semantic scale ratings of interior car sounds. Applied Acoustics **70** (2009) 389–403. doi: 10.1016/j.apacoust.2008.06.008.

[14] J. Steffens: Realism and Ecological Validity of Sound Quality Experiments on Household Appliances. Fortschritte der Akustik — AIA-DAGA 2013, Merano, 1473–1476.

[15] D. Västfjäll: Contextual Influences on Sound Quality Evaluation. Acta Acust united Ac **90** (2004) 1029–1036.

[16] C. Ho, N. Reed, C. Spence: Multisensory In-Car Warning Signals for Collision Avoidance. Human Factors **49** (2007) 1107–1114. doi:10.1518/001872007X249965

[17] W. Krebber: Interactive Vehicle Sound Simulation. Proceedings of Internoise, 2000.

[18] D. Maiberger: Acoustical and non-acoustical influences on vehicle sound evaluation. Ph.D. thesis, Carl von Ossietzky University, Oldenburg, 2018.

[19] U. Letens: Von der Psychoakustik über die psychologische Akustik zur ganzheitlichen Geräuschbewertung in der Fahrzeugakustik. Fortschritte der Akustik — DAGA 2010, Berlin, 433–434.

[20] C. Köber, D. Bönnen, C. Huth, J. Bienert: Charakterisierung der Empfindungsgrößen Sportlichkeit und Komfort basierend auf der subjektiven Beurteilung von Fahrzeuggeräuschen. Fortschritte der Akustik — DAGA 2016, Aachen, 1202–1205.

[21] MathWorks: MATLAB R2014a.

[22] A. Field: Discovering Statistics using SPSS (and sex and drugs and rock ’n’ roll). Third edition. Sage Publications, 2009. ISBN 978-1-84787-907-3.

[23] M. Frigge, D. C. Hoaglin, B. Iglewicz: Some Implementations of the Boxplot. The American Statistician **43** (1989) 50–54.

[24] DIN 45631/A1: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum — Verfahren nach E. Zwicker — Änderung 1: Berechnung der Lautheit zeitvarianter Geräusche; mit CD-ROM. DIN standard, Deutsches Institut für Normung e.V., 2010.

[25] L. D. Sanders, D. Poeppel: Local and global auditory processing: Behavioral and ERP evidence. Neuropsychologia **45** (2007) 1172–1186. doi: 10.1016/j.neuropsychologia.2006.10.010.