Predicting the perceived Quality of impulsive Vehicle sounds

Marius Hoechstetter, Jan-Michael Sautter
BMW AG, D-80788 Muenchen, Germany.

Ulrich Gabbert
Otto von Guericke University Magdeburg, Universitaetsplatz 2, D-39106 Magdeburg, Germany.

Jesko L. Verhey
Otto von Guericke University Magdeburg, Leipziger Str. 44, D-39120 Magdeburg, Germany.

Summary
A key element in the designing process of vehicle sounds is the evaluation of the customers' perception and cognitive expectation on the acoustic quality. As the human auditory perception comprises multiple dimensions, the degree of fulfilling the customers' requirements on acoustic quality is a complex task. In this respect, the perception of impulsive sounds is particular important since those are the sounds the customers listen to at their first contact when visiting a dealership. This study presents psychoacoustic metrics that allow predicting their perceived sound quality. In a first step, significant psychoacoustic parameters were derived from verbal descriptors assessing the sound quality of door handle snapback. Then, paired comparison tests and categorical scale judgments were carried out for door closing and indicator snapback sounds. Based on these results, linear regression analysis revealed a significant predictive accuracy for two psychoacoustic parameters. The first is the classical percentile loudness N_5 (see DIN 45 631) and the second is the newly considered measure duration of sharpness. Their relative contribution depends on the considered sound type. Whereas loudness appears to be the major predictor for the quality of indicator snapback sounds, the duration of sharpness has a large contribution when predicting the acoustic quality of door closing sounds. The influence of these parameters was confirmed in additional experiments measuring the quality of sounds where these parameters were explicitly varied. In summary, the data show that a linear combination of these parameters with signal type specific adjustments on their significance can serve as a good basis for the qualitative assessment of vehicle sound quality.

PACS no. 43.66.Lj, 43.66.Ba, 43.66.Cb, 43.50.Pn

1. Introduction

Since the technical progress in automotive developement reached a high availability of complex technical functions, the process of paradigm shift from rational to emotional aspects has been accelerated. In order to meet the customer's expectations, the vehicle's acoustic performance and appeal became decisive and important for the overall design process [1], [2]. In consequence, all perceivable quality criteria need to match with the general product design [3], [4]. As sound quality is determined on human perception, the car manufacturers take huge efforts to evaluate the perceptual characteristics of automotive sounds with respect to auditory sensations [5]. An highly emotional experience occurs the decisive moment of the customer's first contact when visiting a dealership. In this silent surrounding, no distracting aspects whilst driving are present and no engine is running. The relevant, mainly impulsive sounds result from an explicit action of the customer and induce sensitive action of the human auditory system. As they include both functional and qualitative information [6], [7], the inherent quality of this type of sounds appears highly critical to the subject's judgment of quality.
The aim of the present study was to describe and predict the perceived quality of impulsive vehicle sounds using psychoacoustic parameters. The parameters for the quality prediction were preselected purposefully to correlate with perceived sound quality according to factor loadings of verbal descriptors [8]. As the impartial prediction required an interval scaled classification, hearing tests were carried out by the method of relative comparison. These data were further vali-

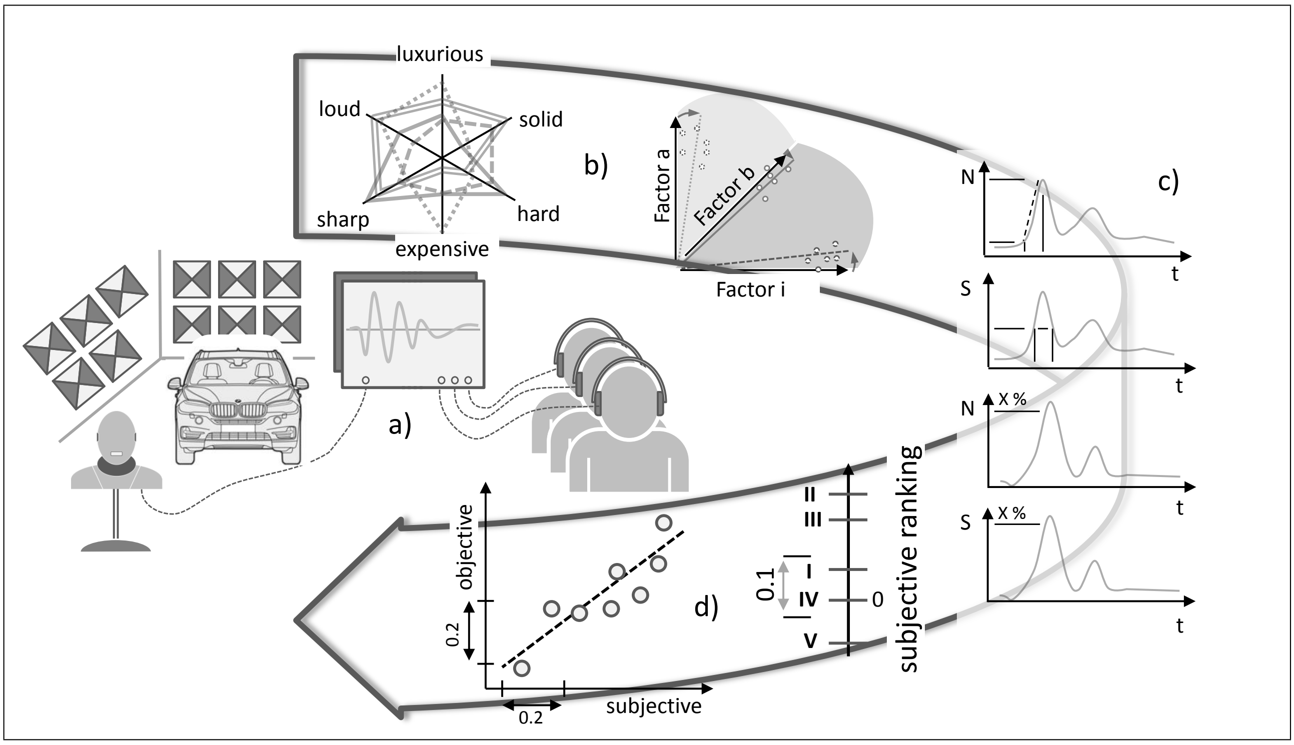

Figure 1. Schematic sketch of the course of the experiment. Recording of the stimuli in a semi anechoic chamber and play back by open headphones a), verbal classification of quality b), deriving psychoacoustical measures c), correlation analysis of subjective and impartial sound quality d).

dated by an absolute ranking using a categorical scale. Applying linear regression analysis, a significant predictive accuracy of two psychoacoustic parameters was revealed. The first is the percentile loudness N_5 and the second is the newly considered measure duration of sharpness, refered to T_S in the following [9], [10]. The relative importance of the parameters depend on the signal type: loudness appears to be the major predictor for the indicator snapback sound whereas the duration of sharpness seems to have an important role in the prediction of the sound quality of door closing sounds. As this approach included the semantic differential to assess the subjects' demands on a high quality sound, psychoacoustic parameters could be derived that describe the auditory sensation of the sound quality. Furthermore, a standard operating procedure to validate and impartially control the development process according to the customers' interests can be integrated.

2. General Methods

2.1. Recordings

Door handle snapback and door closing sounds have been recorded in a semi- anechoic chamber with the help of an artificial head (HEAD acoustics GmbH, type HMS IV) outside the vehicle, as schematically shown in Fig. 1a). The artificial head was placed with the microphones 165 cm over ground and 45 cm aside and back to the car's b-pillar with the face looking at the b-pillar. Regarding the door closing sounds, a closing velocity of about 1.2 m/s (tolerance: ±0.02 m/s) was chosen for the experiments. A light barrier was installed in close proximity to the vehicle to measure the closing velocity. For the recording of the indicator snapback sound, a second articificial head was placed inside the vehicle on the driver's seat. Considering the dimensions of each vehicle type, an identical position of the artificial head was assured with the microphones beeing placed in front of the car's b-pillar and above the steering wheel's upper edge.

2.2. Setup and procedure

The aim of this study was to impartially predict the perceived sound quality of impulsive vehicle sounds. To reach this target, the methods of the semantic differential, the paired comparison and the categorical judgment were used. Starting with a semantic differential test on the door handle snapback sounds, the semantic space was reduced to significant dimensions of perception. As a criterion to choose, these factors were used to preselect several psychoacoustic parameters to predict the perceived sound quality of the door closing and the indicator snapback sounds.
Semantic differential tests allow to assess the subject's demands on a sound of high quality when measuring their auditory perception [8], [11]. This is schematically described in Fig. 1b). A principal component analysis was used to unite identically associated ad-

Table I. Pairs of adjectives, choosen context specifically for door handle snap back sounds according to twelve experts of vehicle acoustics.

loud	quiet
high frequency	low frequency
good	bad
sharp	soft
instable	stable
disturbing	desirable
low value	high value
valuable	worthless
tinny	not tinny
uncomfortable	comfortable
luxurious	simple
hard	soft
irritating	nice
popping	not popping
dull	metallic
hollow	solid
cheap	expensive

jectives to overall dimensions of perception. To derive significant factors, the scree plot's graphical turning point combined with eigenvalues of $\lambda > 1$ were applied as criteria as their own variance exceeds the one of the containing variable. On a statistical basis, Barlett's test of sphericity and the Kaiser-Meyer-Olkin criterion were applied. A factor rotation (varimax) completed the analysis to raise the loadings.

In advance to rank the perceived sound quality on an interval scale (see Fig. 1d)), reliable and ordinal scaled data result applying the paradigm of paired comparison [12]. To ensure high concentration during the evaluation, the sounds were combined to each other by the algorithm of Ross [13]. This setup included all stimuli to appear equally often on the first and the second position and to maximize the number of pairs between the first and the following playback of a sound. Furthermore, the individual preference in this form does not provide the actual difference in quality as the decision was made on the psychological continuum. This has to be derived from the frequency, which listeners prefer one stimulus to another [14]. Thus, the cumulative results were transformed to an interval scale via dominance matrix [14], [15].

Regarding the method of categorical judgment, a seven-point Likert scale was used to assess the sound quality of door closing. Considering the listeners' demands for the indicator snap back sounds, the scaling was enlarged to a nine-point Likert scale, also specified in equidistant steps. Based on the opposing adjectives low quality and high quality, the centre of both bipolar scales was defined as neutral. As this scaling was printed on a sheet of paper, the participants had to indicate the perceived sound quality. Each stimulus had to be randomly evaluated two times (test and retest) in order to check consistency of the answers.

The sounds were presented to the participants over open head phones, type STAX SR-202, as schematically shown in Fig. 1a). Prior to an experiment, a written, standardized text was handed out explaining the corresponding method and the test procedure. At the initial phase of the experiment, the test manager presented the sounds to the participants. Afterwards, each question of the participants was answered by the test manager before the main experiment started.

2.3. Participants

In total, sixteen women aged between 23 and 40 years and 26 men from 19 to 52 years participated in the experiments. The average age of these listeners was 30.2 years with all of them reporting a normal hearing ability. Listeners had different backgrounds: 28 were laymen and 14 experts working in the field of acoustic development. Additionally, the test design of relative comparison considered the restrictive criterion of context independence whilst making a choice [16]. It is valid as context independent, if the entire scale values of the stimuli are taken into consideration when making a judgment. Based on the χ^2-Test, participants with inconsistent responsiveness were excluded from further analysis by a confidence level of α=0.05.

2.4. Psychoacoustical parameters

To predict the perceived quality of singular impulsive vehicle sounds, the present study is focused on psychoacoustical parameters. Considering the factor loadings of the sematic differential test, the focus was on loudness (ANSI S3.4 [17], DIN 45631/A1 [18]) and sharpness (DIN 45692 [19], [20]). In particular, the maximum gradient N'_{max} and average gradient N'_{av} of loudness N, percentile loudness N_5 (see Fig. 3) and the maximum sharpness S_{max} were analyzed. An important derived measure is the duration of sharpness T_S. The duration is calculated as long as the signal exceeds a constant threshold of sharpness level specified to $1 acum$ [8], [9] (see Fig. 5). Due to the short duration and impulsiveness of the sounds, tonality and roughness were not considered in the present study according to the signals' characteristics.

3. Experiments

3.1. Experiment 1: Semantic space of door handle snapback sounds

Stimuli and Methods

For the sematic differerntial test of door handle snapback, eight sounds were recorded from different vehicle types. With the help of twelve acoustic specialists, context specific adjectives were selected including documented terms [21], [22]. Further pretests with laymen resulted in a questionnaire consisting of the 17 pairs as seen in Tab. I. In order to collect valid results and to minimize habits, positive and negative associated adjectives were irregularly distributed on both sides of a bipolar,

Table II. Factor loading for Semantic Differential test of door handle snap back. Factors: (1) Quality, (2) Intensity, (3) Dynamic, (4) Timbre

	Factor			
Pairs	*1*	*2*	*3*	*4*
irritating - nice	-0,90			
disturbung - desirable	-0,84			
cheap - expensive	-0,81			
low value - high value	-0,81			
valuable - worthless	0,77			
uncomf. - comfortable	-0,77			
luxurious - simple	0,73			
loud - quiet		0,88		
hollow - solid		-0,84		
hard - soft		0,79		
high freq. - low freq.			0,87	
sharp - soft			0,82	
instable - stable			0,67	
good - bad				0,83
dull - metallic				0,80
tinny - not tinny				-0,68
popping - not popping				

seven-point scale. In addition to that, the sequence of the adjective pairs was varied and the questionnaires were changed depending on the stimulus.

Results
In total, the auditory sense included the four dimensions of perception shown in Tab. II. The eight sounds invetigated covered a cumulated variance of $\sigma^2 = 79.07\%$. Based on the Kaiser-Meyer-Olkin criterion ($0.77 < KMO < 0.84$) and the Barlett test of sphericity ($sig. = 0.00$), the data could be deemed as suitable and originate from a population only consisting of correlations to zero.
The factor *Quality* shows the highest contribution ($\sigma^2 = 45.40\%$) to explain the perceptual space. Based on a variance of $\sigma^2 = 14.55\%$, the adjectives *loud - quiet*, *hollow - solid* and *hard - soft* describe the second factor named *Intensity.* The thrid one is called *Dynamic* as it includes adjectives related to frequency aspects. The fourth dimension *Timbre* statistically offers an eigenvalue of $\lambda = 1.39$ and a variance of $\sigma^2 = 8.19\%$. As the combination *popping - not popping* did not meet the criterion of normal distribution, it could not be assigned to any factor.

3.2. Experiment 2: Sound quality of door closing sounds

Stimuli and Methods
The perceived sound quality of door closing was subjectively classified using the method of paired comparison. Therefore, fourteen sounds were recorded and devided into two groups of eight stimuli. Two stimuli were equal in each group. With the help of linear regression analysis, an overall psychoacoustical prediction model based on loudness and sharpness was derived to correlate with perceived sound quality.

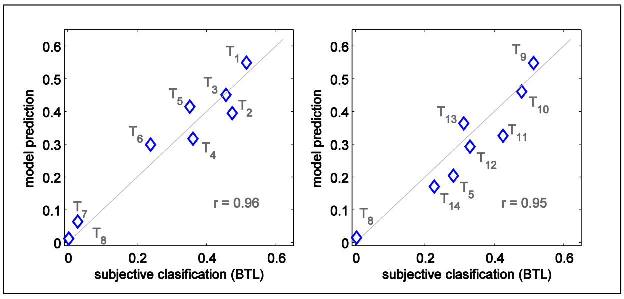

Figure 2. Correlation analysis of subjective and impartial sound quality of the door closing sound based on Eq. 1. Group 1 (left), group 2 (right).

For further validation, additional fourteen sounds of door closing were evaluated in an absolute manner on a seven-point likert scale.

Results
The percentile loudness N_5 and the duration of sharpness T_S showed significant predictability of the sound quality derived form the paired comparison data (see Eq. 1). The euqation (1) indicates that T_S (-0.21) is more important to describe the perceived quality of door closing sounds than N_5 (-0.11). The correlations for both measures are negative, i.e., a soft and moderately sharp door closing sound is associated with high quality.

$$SQ_{dcs} = 0.28 - 0.11 \cdot N_5 - 0.21 \cdot T_S, \qquad (1)$$

Figure 2 shows the high correlation results of the subjective and the impartial sound quality $r = 0.96$ (group 1) and $r = 0.95$ (group 2). Rated by the methods of the $t-$ and the $F-$test on a confidence level $\alpha = 0.05$, the hypothetical assumption is considered to be significant. For the psychoacoustical measures loudness and duration of sharpness, a threshold criterion was used. Figure 3 shows how the overall correlation of sound quality change with this criterion, i.e., in this case with the loudness percentile. For the door closing sound, the percentile value of N_5 had the highest correlation to the perceived quality. Figure 5 shows the prediction accuracy for various sharpness levels which was the criterion for this measure. The threshold of $1acum$ offered the highest values to correlate with human perception. Moreover, the derived model was used to predict the perceived quality of additional fourteen door closing sounds, classified absolutely via the categorical judgment. As the relation shows a high coefficient of correlation $r = 0.94$, this further supports a reliable and representative predictability of this signal type.

3.3. Experiment 3: Sound quality of indicator snapback sounds

Stimuli and Methods
To classify the fourteen indicator snapback sounds,

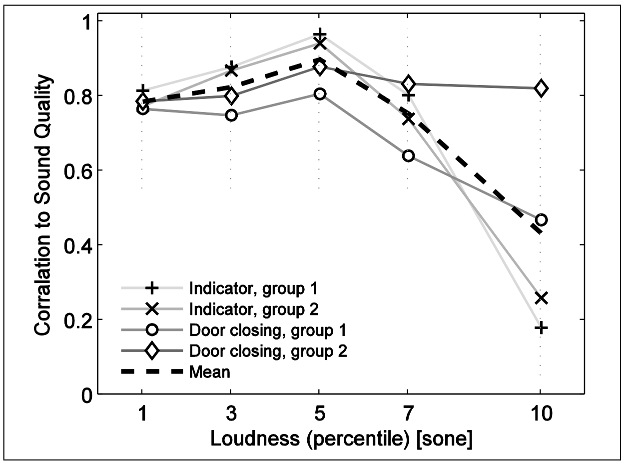

Figure 3. Correlation of loudness to sound quality over loudness percentile. Each symbol indicates the correlation of loudness (absolute value) to sound quality based on the corresponding percentile of loudness. The data of the door closing and indicator snapback sounds as well as their arithmetic mean are shown.

the same method of paired comparison was used for the door closing sounds. The data were further validated by additional experiments which questioned the sound quality on a nine-point Likert scale. Again, a correlation analysis was used to evaluate the predictive quality of the significant impartial measures considering loudness and sharpness.

Results
The correlation between the interval scaled results and the absolute ranking for the indicator snapback sound was high ($r = 0.94$). Thus, the data are regarded as valid. Based on the interval scaling, linear regression analysis led to a psychoacoustical prediction model with the same measures N_5 and T_S already used for the door closing sounds. For the present signal type, loudness is assigned with a weighting value of -0.21 while the duration of sharpness only had a weighting of 0.06 (see Eq. 2).

$$SQ_{iss} = 0.25 - 0.21 \cdot N_5 - 0.06 \cdot T_S, \tag{2}$$

Using this model to predict the sound quality of indicator snap back sounds (Fig. 4), the correlation values of $r = 0.98$ and $r = 0.97$ concerning group 1 and 2 were even more significant than for the door closing sounds. One major aspect is the very high coefficient of correlation for the percentile loudness N_5 as seen in Fig. 3. The duration of sharpness is less important to predict the perceived quality, since the calculated values of -0.65 and -0.66 in Fig. 4 correlate low with the sound quality. Equal to the door closing sound, both loudness and duration of sharpness reach the highest correlation at percentile N_5 and sharpness level of $1 acum$ to describe the perceived quality of indicator snapback sounds.

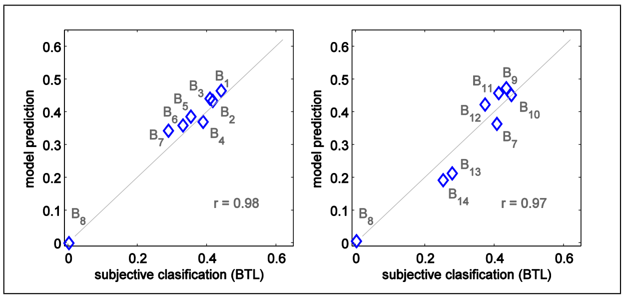

Figure 4. Correlation analysis of subjective and impartial sound quality of the indicator snapback sound based on Eq. 2. Group 1 (left), group 2 (right).

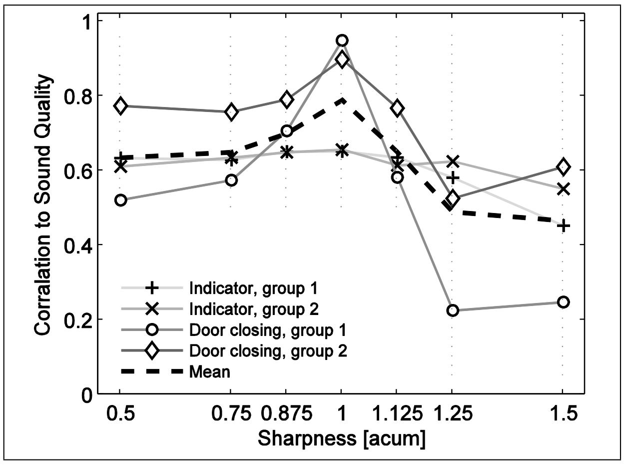

Figure 5. Correlation of the duration of sharpness T_S to sound quality over sharpness. Each symbol indicates the correlation of the duration of sharpness (absolute value) to sound quality based on the corresponding threshold of sharpness. The data of the door closing and indicator snapback sounds as well as their arithmetic mean are shown.

4. Discussion

As aesthetic and emotional sound aspects [2] contribute to the customer's whole picture of vehicle quality, the auditory perception of the sound quality is affected by multiple characteristics [1], [3], [5]. In order to impartially predict the related sound quality, relevant psychoacoustical parameters need to be selected purposefully. Considering the reduced dimensions of the perception of impulsive vehicle sounds, the customer's preferred sensation is reflected within four factors. Dependent on the share of variance, the factor *Quality* contributes to the subject's expectation on the quality of door handle snapback sounds, followed by *Intensity*, *Dynamic* and *Timbre*. Thus, the auditory quality represents a complex measure, which is not tangible by just individual factors [8], [11], [20]. As the dimensions *Intensity* and *Dynamic* include mainly adjectives related to loudness and frequency aspects of sounds, the focus was on the psychoacoustical measures of loudness and sharpness to predict the perceived quality of impulsive vehicle sounds.

The present study showed that for the impartial prediction of door closing and indicator snapback sounds, two major aspects are sufficient. In agreement with published results on a multidimensional semantic space [11], [23] and cognitive processing of impulsive vehicle sounds, a linear combination of two psychoacoustical parameters led to a higher predictability of the perceived quality compared to a single measure [9] (see Fig. 3, Fig. 5).The same parameters for the percentile loudness N_5 and the duration of sharpness T_S were derived reflecting the similarity of the two sound types. The relative impact of the two measures differed between the signal type. These results show that the derivation of a benchmark in sound desgin is a complex task. Since the approach is directly implementing the customers' expectations, multiple acousitc aspects have to be considered.

5. Summary and Conclusion

The empirical studies focused on the perceived quality of impulsive sounds when entering a vehicle. Based on these models, systematically testable targets can be derived and integrated in the development process of sound design. In future, specifically derived requirements can be determined for a certain vehicle category to precisely satisfy the customer's interests and to further increase the appeal of vehicle comfort.

References

[1] H. van der Auweraer, K. Wyckaert, W. Hendricx: From sound quality to the engineering of solutions for NVH problems. Acta Acust. Acust. 83, 796-804 (1997).

[2] U. Jekosch: Assigning Meaning to Sounds - Semiotics in the Context of Product Sound Design. in: J. Blauert (ed.): Communication Acoustics, 193âĂŞ219, (Springer, Berlin-Heidelberg-New York NY, 2005).

[3] M. Haverkamp: Synästhetische Wahrnehmung und Geräuschdesign. in: K. Becker (ed.): Subjektive Fahreindruecke sichtbar machen II, (Expert- Verlag, Renningen-Malmsheim, 2001).

[4] J. Blauert: Product sound design and assessment: An enigmatic issue from the point of view of engineering? Inter-Noise, Yokohama, Japan, 857-862 (1994).

[5] K. Genuit: Significance of psychoacoustic aspects for the evaluation of vehicle exterior noise. J. Acoust. Soc. Am. 123, 3134 (2003).

[6] G. Bobbert: Innengeräusche von Kraftfahrzeugen. in: Fortschritte der Akustik - DAGA 1988 (Dt. Ges. fuer Akustik, Braunschweig, 1988).

[7] K. Genuit, B. Schulte-Fortkamp, A. Fiebing: Neue Verfahren zum Benchmarking von Fahrzeuginnengeräuschen. in: K. Becker (ed.): Subjektive Fahreindruecke sichtbar machen III, chap. 8, 127-145, (Expert-Verlag Renningen-Malmsheim, 2006).

[8] M. Hoechstetter, M. Wackerbauer, J. L. Verhey, U. Gabbert: Psychoacoustic prediction of singular impulsive sounds. ATZ - Automobiltechnische Zeitschrift, (4/2015, Springer Verlag).

[9] M. Hoechstetter, M. Rolle, J. L. Verhey, U. Gabbert: Physikalische und psychoakustische Vorhersage von singular impulshaften Geäuschen. 8. Symposium Motor- und Aggregateakustik, Magdeburg, (2014).

[10] M. Hoechstetter, M. Rolle, J.-M. Sautter, B. A. Hemmrich, DE 102014212733.5. patent application (7/2014).

[11] M. E. Altinsoy, U. Jekosch: The semantic space of vehicle sounds: Developing a Semantic Differential with regard to customer perception. J. Audio Eng. Soc. 60, 13-20 (2012).

[12] N. Otto, S. Amman, C. Eaton, S. Lake: Guidelines for Jury Evaluations of Automotive sounds. Sound and Vib. 35, 24-47 (1999)

[13] R. T. Ross: Scaling: A sourcebook for behavioral statistics. Aldine Pub. Company (1974).

[14] F. Sixtl: Messmethoden der Psychologie. Verlag Julius Beltz, Weinheim (1967).

[15] M. G. Kendall: Rank Correlation Methods. Griffin, London (1970).

[16] R. D. Luce: The Choice Axiom after Twenty Years. in: J. Math. Psy., 15, 215-233 (1977).

[17] ANSI S3.4-2007 (R2012) American National Standard, Procedure for the Computation of Loudness of Steady Sounds (American National Standards Institute, New York, 2007).

[18] DIN 45631/A1:2010-03 Calculation of loudness level and loudness from the sound spectrum - Zwicker method- Amendment 1: Calculation of the loudness of time-variant sound; with CD-ROM (Deutsches Institut fuer Normung e.V., Berlin, 2010).

[19] DIN 45692:2009-08 Measurement technique for the simulation of the auditory sensation of sharpness (Deutsches Institut fuer Normung e.V., Berlin, 2009).

[20] E. Zwicker, H. Fastl. Psychoacoustics - Facts and Models. (Springer-Verlag Berlin Heidelberg, 3rd ed. (2006)).

[21] E. Altinsoy: Knocking Sound as Quality Sign for Household Appliances and the Evaluation of the Audio-Haptic Interaction. in: Haptic and Audio Interaction Design (Springer-Verlag Berlin Heidelberg, 121-130 (2012)).

[22] T. Hashimoto: Die japanische Forschung zur Bewertung von Innengeräuschen im PKW. in: Zeitschrift fuer Laermbekaempfung, 41, 69-71 (1994).

[23] S. Kuwano, H. Fastl, S. Namba, S. Nakamura, H. Uchida: Quality of door sounds of passenger cars. Acoust. Sci. & Tech. 27, 309-312 (2006).