Acoustical and non-acoustical influences on vehicle sound evaluation

Von der Fakultät für Medizin und Gesundheitswissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von Herrn **David Maiberger**

geboren am 29.05.1989 in Gelnhausen

Gutachter:	Prof. Dr. Steven van de Par
Zweitgutachter:	Prof. Dr. Ercan Altinsoy
Tag der Disputation:	4. Dezember 2018

Contents

Summary

Vehicle sound evaluation, as any kind of product sound evaluation, is a complex process. Along with purely acoustical aspects, such as comfort, sportiness, conformity with vehicle type and with vehicle make, also non-acoustical aspects like expectation, personality and situation influence the perceived sound quality of a car. In this work, a model for the process of vehicle sound evaluation is developed that incorporates acoustical as well as non-acoustical aspects.

In the first part of this work, a typology of car drivers is developed that represents the typical real-world configurations of personality traits related to vehicles, sound perception as well as vehicle sound preferences. From results of an online survey among drivers of premium vehicles three clusters of participants are identified: the *sound-oriented fun-driver* with a dynamic driving style and high sound awareness, the *noise-sensitive aesthete* with an equally high sound awareness but also high noise sensitivity, and the *sound-uninterested pragmatist* with a low sound awareness, seeing a car as a tool rather than a status symbol.

In the second part, the relationship between the acoustical and non-acoustical aspects is analyzed with data gained in an extensive field study. From a linear modeling approach a model is developed that best describes the data. In this model, acoustic comfort and acoustic sportiness directly determine vehicle sound quality, but with weights depending on the combination of driver type and vehicle type. Furthermore, high type conformity is important for a high sound quality, but type conformity itself is determined by the particular combination of comfort and sportiness, depending on the vehicle type. In addition, it is found that acoustic comfort is determined mainly in static driving conditions such as driving with constant speed in city traffic or on rural roads. It is indicated that, conversely, acoustic sportiness is mainly determined in dynamic driving conditions such as accelerating and overtaking.

In the third part, it is investigated to what extent there is congruence between the model as developed in the field and results gained in a listening experiment performed in the lab. For this purpose, a lab experiment is conducted where subjects are provided with non-auditory information about the rated cars, namely a picture and the name. It shows that in the lab comfort has much greater influence on the rating of vehicle sound quality, and in the field vehicle sound quality is generally rated higher. In the lab dynamic driving conditions as well as standstill driving conditions, such as engine start and engine idle, exert an influence on the comfort rating, which was not observed in the field. Furthermore, the dependence of the rating on the concrete task given to the subjects becomes apparent by a low test-retest reliability.

It is therefore concluded that future sound design processes should incorporate the model for vehicle sound assessment. Especially the connection to sound metrics and thus the relation to the actual sound should be established, in order to allow for a meaningful subject-centered sound engineering. The driver typology has proven to be a successful tool for describing individual preferences and differences in vehicle sound evaluation.

Zusammenfassung

Wie jede Art von Produktgeräusch-Bewertung ist auch die Fahrgeräusch-Bewertung ein komplexer Prozess. Neben rein akustischen Aspekten wie Komfort, Sportlichkeit und Konformität mit dem Fahrzeugtyp sowie der Fahrzeugmarke beeinflussen auch nicht-akustische Aspekte wie Erwartungshaltung, Persönlichkeit und Situation die wahrgenommene Fahrgeräusch-Qualität. In dieser Arbeit wird ein Modell für den Fahrgeräusch-Bewertungsprozess entwickelt, das akustische sowie nicht-akustische Aspekte berücksichtigt.

Im ersten Teil der Arbeit wird eine Typologie von Autofahrern entwickelt. Diese repräsentiert typische Kombinationen derjenigen Persönlichkeits-Eigenschaften, die Fahrzeuge, Geräusch-Wahrnehmung sowie Fahrgeräusch-Vorlieben betreffen. Anhand der Ergebnisse einer Online-Befragung von Fahrern von Premium-Fahrzeugen werden drei Personengruppen identifiziert: *soundorientierte Spaßfahrer* mit einem dynamischen Fahrstil und hohem Fahrgeräusch-Bewusstsein, *lärmempfindliche Ästheten* mit ebenso hohem Fahrgeräusch-Bewusstsein bei gleichzeitig hoher Lärmempfindlichkeit, und *Sound-uninteressierte Pragmatiker* mit geringem Fahrgeräusch-Bewusstsein, für die ein Auto eher Gebrauchsgegenstand als Statussymbol ist.

Im zweiten Teil wird der Zusammenhang zwischen akustischen und nicht-akustischen Aspekten anhand von Daten aus einer umfangreichen Feldstudie analysiert. Mit einem linearen Modellierungs-Ansatz wird das Modell entwickelt, das die Daten am besten beschreibt. In diesem Modell hängt die Geräusch-Qualität direkt vom akustischen Komfort und der akustischen Sportlichkeit ab, mit Gewichten die von der Kombination aus Fahrertyp und Fahrzeugtyp abhängen. Weiterhin ist Typ-Konformität wichtig für die Geräusch-Qualität, wobei Typ-Konformität selbst von der fahrzeugtyp-spezifischen Kombination von Komfort und Sportlichkeit abhängt. Zusätzlich wird festgestellt, dass der akustische Komfort im Wesentlichen von statischen Betriebszuständen wie Konstantfahrten innerorts oder auf der Landstraße abhängt. Es deutet sich an, dass akustische Sportlichkeit hingegen im Wesentlichen von dynamischen Betriebszuständen wie Beschleunigen und Überholen abhängt.

Im dritten Teil wird untersucht inwieweit das im Feld entwickelte Modell mit Hörversuchs-Ergebnissen kompatibel ist. Dazu wird ein Labor-Experiment durchgeführt, in dem die Probanden nicht-akustische Informationen zu den bewerteten Fahrzeuge erhalten, nämlich deren Bild sowie Bezeichnung. Komfort hat im Labor einen größeren Einfluss auf die Geräusch-Qualität, und im Feld wird die Geräusch-Qualität generell höher bewertet. Im Labor zeigen dynamische sowie Stillstands-Betriebszustände, wie Motorstart und Leerlauf, einen Einfluss auf die Komfort-Bewertung, der im Feld nicht beobachtet wurde. Weiterhin verdeutlicht die geringe Test-Retest-Reliabilität die Abhängigkeit der Bewertung von der konkreten Aufgabenstellung.

Es wird gefolgert, dass das Fahrgeräusch-Bewertungsmodell künftig in den Sound Design-Prozess integriert werden sollte. Speziell die Verbindung mit Geräusch-Metriken und damit zum eigentlichen Fahrgeräusch sollte hergestellt werden, um sinnhaftes, fahrerbezogenes Sound-

Engineering zu ermöglichen. Die Fahrer-Typologie hat sich als erfolgreiches Werkzeug zur Beschreibung individueller Vorlieben und Unterschiede bei der Fahrgeräusch-Bewertung erwiesen.

1 Introduction

1.1 General background

The sound produced by a vehicle is an important quality factor. For example, a car with a loud sound will likely be perceived as less pleasant than a comparable car with a quiet sound. However, the mere distinction between loud and quiet cars is not sufficient for describing the holistic perception of vehicle sounds. This insufficiency has become more apparent as a result of decreasing sound levels that occurred in almost all types of passenger cars over the last decades. The reduction of emitted sound levels gave rise to an increased audibility of sounds from other sources than the engine. This includes ancillary components such as the electrical steering aid, valves, pumps, fans, electric seat adjustment, electric window lift, electric sunroof or heating, ventilation and air conditioning (HVAC) (see Engel et al., 2010, p. 306), which had formerly been masked by wind, rolling and engine noise. The process of "sound cleaning", i.e., the removal of such disturbing sounds, is just the predecessor to "sound engineering", which describes the tuning of relevant sound components in order to achieve a desired sound as defined by the "sound design" (Pfäfflin et al., 2010, p. 430). The sound design usually aims at achieving the "sound of quality", i.e., the sound that conveys an impression of quality. It is of particular relevance because many customers relate the experienced sound quality to the actual quality of the car (Miśkiewicz and Letowski, 1999), which is a distinguishing feature. Thus, the sound of the car influences "the customer's satisfaction with the vehicle" and "is an important criterion in choosing which product to purchase" (Miśkiewicz and Letowski, 1999, p. 646). Hence, it is important for the sound engineer to know what constitutes a high *sound quality*.

In general, quality evaluation of product sounds is a complex process. The complexity arises from the interplay between physical properties of the sound waves, sensory properties of the human auditory system, as well as psychological processes involved in the judgement. The path from the sound-emitting product to product sound quality, according to Blauert and Jekosch (1997), is depicted in Figure 1.1. In the course of the depicted processing the total amount of available information is reduced to finally arrive at a single sound quality judgement (Blauert and Jekosch, 1997). At the beginning, in the leftmost part of Figure 1.1, the complete (acoustic) information is carried by the emitted "acoustic waves". These waves could be physically measured with a microphone and analyzed with appropriate hardware or software tools, which

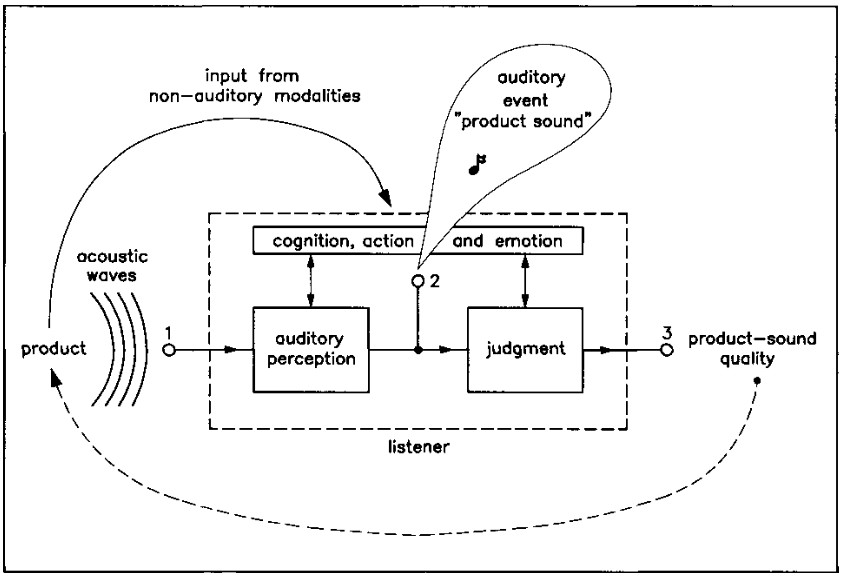


Figure 1.1: Schematic of a listener in the process of product-sound-quality assessment [taken from Blauert and Jekosch (1997)].

is why this stage can be called the domain of *physical acoustics*. In the next stage the sound is being perceived by the listener, as indicated by the block labeled "auditory perception", causing an auditory event. The unbiased description of auditory events is the scope of *psychoacoustics*. At this stage the first information reduction has already taken place: the sound waves have propagated through the outer ear, have been transformed by the eardrum, transmitted over the middle ear into the inner ear, where they have evoked nerve impulses. These nerve impulses have been transmitted to the brain, where they are being interpreted, eventually causing the auditory event. The final block in the horizontal chain is the process of "judgement", the output of which finally is product sound quality. This stage of conscious judgement can be called the domain of *psychological acoustics*. In total, the described path characterizes the *acoustical* contribution to sound quality, originating from the sound emitted by the product.

In Figure 1.1, however, a further block called "cognition, action and emotion" interacts with both auditory perception and judgement. These response-moderating factors are *non-acoustical* in nature as they relate to personal attributes of the evaluating subject, to the situational context, to the subject's current mood, and they incorporate "input from non-auditory modalities" (such as the visual, tactile or olfactory senses). The definition of product sound quality given by Blauert and Jekosch (1997, p. 748) is the following:

> "*Product-sound quality is a descriptor of the adequacy of the sound attached to a product. It results from judgements upon the totality of auditory characteristics of the said sound — the judgements being performed with reference to the set of those desired features of the product which are apparent to the users in their actual cognitive, actional and emotional situation.*"

According to this definition, the evaluation of product sound quality must be viewed with regard to the product itself, as also indicated by the dashed arrow in Figure 1.1 pointing back at the product.

The "totality of auditory characteristics" is typically referred to as the *sound character*. In contrast to sound quality, "sound character judgment does not imply anything about the listener's preferences or satisfaction" (Miśkiewicz and Letowski, 1999, p. 647). Sound character rather "refers to the unbiased sensory properties of the sound implying that these are neutral descriptions to the largest extent" (Zeitler et al., 2006, p. 5624). One can think of the sound character as a "profile" of a sound with respect to psychoacoustic properties without regard to its source, i.e., the product.

Following the above definition, in order to determine a product's sound quality its sound character is compared against a reference set of desired features of the product. Thus, the product or the product class defines a certain reference, which could be specified with various degrees of abstraction, e.g., as a concrete target sound, as a particular sound character or as a certain "meaning" of the sound (see Jekosch and Blauert, 2005, Jekosch, 1999, Zeitler et al., 2006). In any case, the closer the experienced product sound matches that reference, the higher is its sound quality (Jekosch and Blauert, 2005). Since this reference in sound quality evaluation is product-specific, the judgement of product sound quality is made, as stated earlier, with regard to the product, not just to its sound.

However, the reference further depends on the users' "cognitive, actional and emotional situation". The "cognitive situation" encompasses prior knowledge (Blauert and Jekosch, 1997) as well as "individual expectation, habit, experience, personal taste, collective taste (zeitgeist), attitude and motivation, social and cultural influences", as pointed out by Bednarzyk (1999, p. 38). These factors vary from listener to listener and can therefore be characterized as being *personal influences*. For example, the sound quality evaluation of electric screwdrivers differs between professional and free-time users due to "different expectations of the groups" (Fiebig and Kamp, 2015, p. 125). Also the sounds from every-day situations and electrical devices can be judged to be more annoying when they are not identifiable by the listener (Zeitler et al., 2006). This emphasizes the role of sound as a *sign*, the "meaning" of which is associated by the listener factoring in his/her knowledge and experience (Jekosch, 1999). The action a listener is engaged in when using the product defines the "actional situation" (Blauert and Jekosch, 1997). Product sound evaluation can differ between situations in which the subject is just listening to the sounds as compared to situations where the subject is engaged in another activity. Bednarzyk (1999) asserts that a product user is less annoyed by the product sound than a passive listener, because the user has control over the sound while the listener is helplessly exposed to it. To the user the sound has a positive meaning as it provides feedback about the correct functioning of the product. For example, Steffens (2013) showed that the sounds of household appliances (laundry dryers, washing machines, refrigerators) are evaluated less “pleasant” in a listening experiment than in a launderette, where the subjects washed and dried their clothes. Similarly, Fiebig (2015) found differences in the evaluation of vacuum cleaner sounds that were due to the test environment. To be precise, in a laboratory context the assessments with regard to the attributes “powerful”, “dull”, “sharp” and “acceptable” were (almost) significantly different from assessments made in a living room environment, where subjects watched TV while evaluating the sounds[1]. Finally, also the “emotional situation”, i.e., the emotional state of the listener, causes variation in sound quality judgements. In a series of experiments Västfjäll (2004) showed that a positive or negative frame of mind or attitude towards the product influences the rating of sound quality in a valence-congruent manner, meaning that a positive attitude leads to more positive evaluations and vice versa. In addition, this short-term frame of mind interacted with the more stable personality trait of *noise sensitivity* (see Section 1.2.3.1). The influence of “input from non-auditory modalities” has been observed, for example, in the form of audiovisual interactions affecting sound quality judgements of kettle sounds when pictures of the rated kettles were presented to the participants (Fiebig, 2015). Similarly, an accelerating sports car can be judged louder if a picture shows it in red rather than blue or light-green (Menzel et al., 2008). But also other modalities can influence subjective reactions to sounds. For example, Ho et al. (2007) found that audiotactile warning signals are more effective in capturing driver’s attention than purely auditory or vibrotactile signals, as seen from significantly shorter reaction times.

In conclusion, Blauert and Jekosch (1997, p. 748) summarized the following:

> “*Sound-quality, hence, is not an inherent property of the product, but rather something which develops when listeners are auditorily exposed to the product and judge it with respect to their desires and/or expectations in a given situational context.*”

Consequently, for improving sound quality, it is necessary to understand the influence of these non-acoustical aspects on sound quality. This includes having background knowledge about the typical product users, about their expectations with regard to the product sound, about the typical usage situations and about the influence of emotional aspects connected to the product. In theory, if all relevant contributions were understood, the process of product sound quality evaluation could be modeled according to the depiction in Figure 1.1, and the judgement of product sound quality could eventually be predicted from a set of acoustical and non-acoustical inputs. To develop such a framework for the realm of vehicle sound is the primary goal of this work.

In the specific context of vehicle sound, there are three acoustical aspects that are commonly treated in literature as constituting vehicle sound quality. These aspects can be termed as

[1] These two latter studies just serve as examples since the specific evaluation situation differed between the experiments, relating to the concept of *ecological validity* of such studies (see Section 1.2.2).

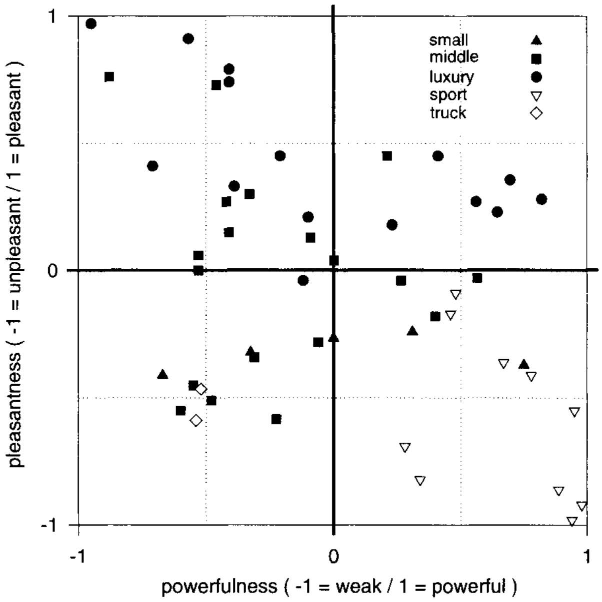


Figure 1.2: Scattering of 51 different car interior sounds within the four quadrants of pleasantness and powerfulness (from Bisping, 1995) [taken from Bisping (1997)].

comfort, *sportiness* and *conformity* (Bisping, 1997, Zeitler and Zeller, 2006, Letens, 2010, Köber et al., 2016). Although the terms differ between authors, it can be assumed that these three constructs are identical across the cited literature. To be precise, comfort in this context is described as "freedom from disturbing noise", "perceived value[2]", "pleasantness", and as being inversely related to loudness. Sportiness is related to "powerfulness" and to "emotional" as well as "aesthetic" aspects of vehicle sound. Conformity corresponds to the adequacy of the sound with respect to the car's market positioning, i.e., the extent to which the sound matches the vehicle character and brand values ("corporate sound"). Therefore, conformity can be considered to be twofold, divided up into *type conformity* and *make conformity*, referring to the vehicle type and the vehicle make, respectively. This definition of conformity encompasses the aspects adequacy and expectation, both brought up by Blauert and Jekosch (1997) in the above definition of product sound quality. Also Zeitler and Zeller (2006) state that "sound quality can be defined as the extent to which the sound character matches the overall vehicle character and supports brand values". However, in this work the construct of conformity is distinguished from sound quality: a car whose sound perfectly matches the sound character required for its type and make will not necessarily be evaluated as the best sounding car when that particular sound character does not match the evaluating subject's personal preference. Instead, vehicle sound quality is assumed to be dependent on the acoustical aspects comfort, sportiness and (type and make) conformity, in accordance with Letens (2010).

[2]German "Wertanmutung" in the original article (Letens, 2010).

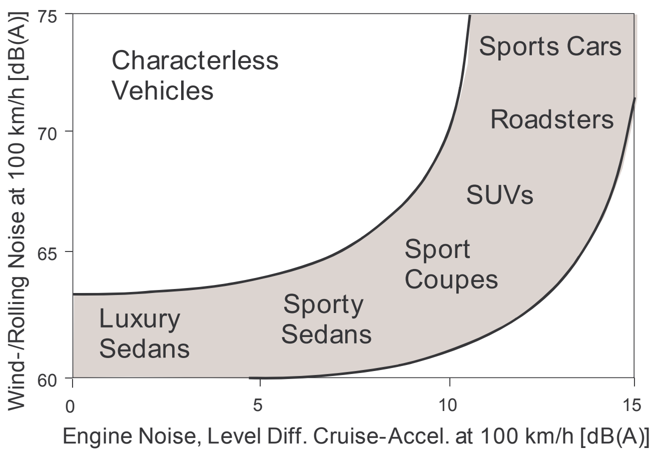


Figure 1.3: Acoustic vehicle positioning. Different sound characters as defined by the noise levels during cruise (y-axis) and acceleration (x-axis) are required depending on the vehicle type [taken from Zeitler and Zeller (2006)].

These four principal aspects of vehicle sound, however, are not completely independent from one another. For example, comfort and sportiness are linked by the “trade-off hypothesis of pleasantness and powerfulness” proposed by Bisping (1997, p. 815). This trade-off means that as sportiness increases comfort will eventually be lowered and vice versa, which Bisping could confirm in listening experiments with sounds that had been manipulated accordingly. Consequently, in Bisping’s depictions of ratings of different car sounds hardly any car combines a high degree of comfort and sportiness at the same time (see Figure 1.2). Zeitler and Zeller (2006) relate sportiness to the level difference of engine noise between cruise and acceleration at 100 km/h, and comfort (inversely) to the wind and rolling noise at 100 km/h (cf. Figure 1.3). Also in their corresponding schematic depicted in Figure 1.3, no vehicle type typically combines a high degree of comfort *and* sportiness. Another connection between the four acoustic aspects is the relationship between type conformity on the one hand and comfort and sportiness on the other hand. Depending on the vehicle type, cars have to exhibit specific sound characteristics in terms of the combination of both comfort and sportiness, as can be seen from the differentiation between vehicle types in Figure 1.3. For example, a sports car could be expected to sound rather sporty and less comfortable, while for a luxury sedan it is the other way round. Similarly, from Bisping’s (1997) depictions of ratings of car sounds in Figure 1.2 it can be concluded that different sound profiles in terms of pleasantness and powerfulness are required for different car categories, as the distinguished categories “small”, “middle”, “luxury”, “sport” and “truck” occupy different quadrants. Along with the connection between comfort, sportiness and type conformity, this indicates an influence of the non-acoustical factor *vehicle type* on type conformity and thus on sound quality. Analogous to this, an influence of *vehicle make* (which determines together with vehicle type the market positioning) on make conformity can be assumed. Hence, in order to develop a model of vehicle sound evaluation, the connection

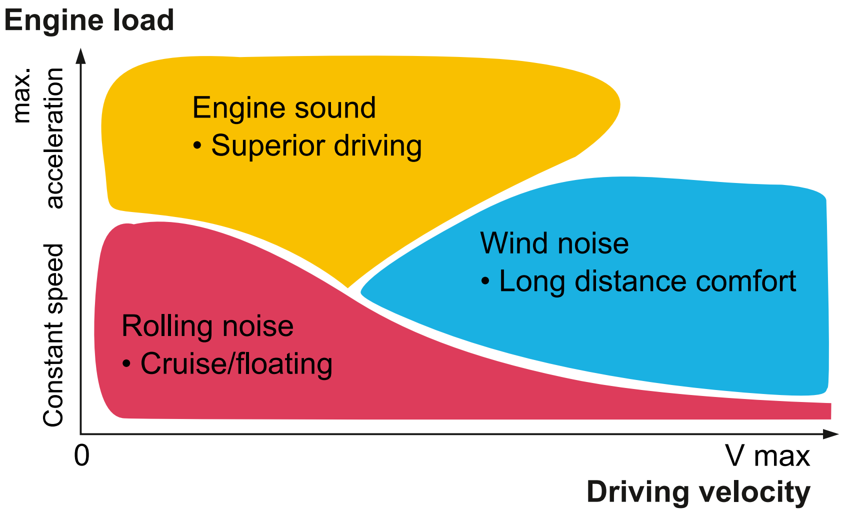


Figure 1.4: Sound spectrum — differing dominance depending on the driving condition (source: Daimler AG) [taken from Pletschen (2010), translated from German by the author].

between the four aspects comfort, sportiness, type conformity and make conformity and sound quality will be investigated in this work by means of statistical modeling, while taking care of the influence of different vehicle types and vehicle makes.

In addition to this integrated view on vehicle sound quality, also more differentiated considerations are necessary as car drivers typically experience their cars in a variety of operating conditions, or *driving conditions*. For example, the sound emitted by a vehicle will be different depending on whether it is performing an engine start, engine idle, acceleration, coast down or driving with constant speed. Accordingly, different driving conditions are typically distinguished in studies concerning vehicle sounds (Bisping, 1997, Letens, 2000, Ishii et al., 2001, Fuhrmann and Garcia, 2001, Schulte-Fortkamp and Genuit, 2005, Zeitler and Zeller, 2006, Altinsoy et al., 2010, Jung and Grützmacher, 2011, Sukowski et al., 2013, Busse, 2014, Fiebig and Kamp, 2015, Alenius et al., 2016). In Figure 1.4, it can be seen that depending on driving velocity and engine load either engine sound, rolling noise from the tires or wind noise are the dominant sound source (taken from Pletschen, 2010). As it can be suspected that overall sound evaluation incorporates assessments in a variety of driving conditions, those driving conditions relevant for sound evaluation will be identified in this work in order to fully understand vehicle sound evaluation.

Besides the two aforementioned aspects market positioning (i.e., vehicle type and make) and driving condition, there are further non-acoustical factors influencing vehicle sound evaluation, such as *situational* and *personal* aspects. For example, from exterior vehicle pass-by sounds an uninvolved listener can experience annoyance as well as physiological reactions, which might affect health conditions (Notbohm et al., 2013). In contrast, a more involved listener such as the driver might enjoy the vehicle sound as "in a sports car, an adequate engine sound can increase driving pleasure and the emotional state of the driver" (Jung and Grützmacher, 2011, p. 2). This shows that the relationship of the listener to the car plays a role: depending on the situation in which the sound is perceived (in this example as an external, uninvolved listener or as the driver in the car interior) it can evoke vastly different affective reactions. Miśkiewicz and Letowski (1999, p. 647) assert that "an engine sound that is perfect for a young driver of a sports car will, most likely, be unacceptable for an older person driving in a luxury sedan", as sound quality judgements do not just depend on the character of the vehicle, but also on user expectations. Thus in order to create meaningful target sounds that achieve acceptance, subject-centered methodology with regard to the target customer has to be applied (Schulte-Fortkamp et al., 2006). Also physiological and affective reactions to vehicle pass-by sounds are moderated by the personal aspects gender, age, noise sensitivity and sensation seeking (Notbohm et al., 2013). This underlines that factors lying within the individual, such as gender, age or personality traits, influence the response to vehicle sounds. Hence, in a model for vehicle sound evaluation as the one aimed at in this work, the driver's personality needs to be handled in an adequate way, which will be achieved with a driver typology. Furthermore, it has to be determined whether such a model is compatible with different rating situations.

Although the non-acoustical factors market positioning, driving condition, personality and rating situation exert an influence on the sound evaluation, the acoustical signal itself is typically considered as the prime basis for sound quality. Even though disregarding non-acoustical factors limits the ecological validity of the results, various psychoacoustic *metrics* have been developed as objective sound descriptors. With such metrics, a scalar value reflecting the average subjective rating of a particular sound attribute can be calculated directly from sound recordings, which is then used to objectively compare the sounds originating from different sources. The best-known psychoacoustic metrics probably are the calculation methods for *loudness* (DIN 45631/A1, 2010), *tonality* (DIN 45681 Berichtigung 2, 2006), *sharpness* (DIN 45692, 2009) and the *speech intelligibility index* (SII) (ANSI S3.5, 1997), all of which have been standardized. Loudness describes the sensation of intensity of a sound, tonality describes the presence of audible tonal components in a sound, sharpness relates to the amount of high frequency content of a sound, and the SII indicates to what extent speech can be masked by a sound. Software packages for performing psychoacoustic analyses typically offer further metrics, including those for temporal characteristics such as *fluctuation strength*, *roughness* or *impulsiveness* (for more information on these psychoacoustical attributes, see Fastl and Zwicker, 2007, Oetjen et al., 2016). In recent years much effort has been made in order to develop specific metrics for perceptual attributes related to vehicle sounds, such as comfort (Noumura and Yoshida, 2003, Zeitler and Zeller, 2006, Köber et al., 2016), sportiness (Ishii et al., 2001, Noumura and Yoshida, 2003, Zeitler and Zeller, 2006, Altinsoy, 2015, Köber et al., 2016, Maiberger et al., 2017), engine presence (Alenius et al., 2016), diesel knocking (Keppens and van de Rostyne, 2015), impulsive as well as tonal engine sound phenomena (Pischinger et al., 2006), annoyance (Beidl and Stücklschwaiger, 1997, Mutalik et al., 2017) or sound quality (Ingham et al., 1999). These approaches mostly depend on linear regression models, in which typically standardized psychoacoustic estimates as well as more specific calculations are being correlated to subjective ratings. The metric is then deployed as the linear model that predicts the perceptual attribute as a weighted sum of the inputs such that the quadratic error between the model output and the subjective ratings will be minimal (for an overview regarding metric development and common pitfalls, see Fiebig and Kamp, 2015). However, as the subjective ratings used for this kind of regression are typically collected in the listening lab, the stimuli are taken out of their typical context, leading to "biased results, which often cannot be generalized and confirmed in reality" (Schulte-Fortkamp et al., 2006, p. 1303). Therefore, it must be questioned whether the results can be considered to be ecologically valid (cf. Section 1.2.2). In addition, such metrics cannot account for the influence of non-acoustical aspects as discussed above, including the typical usage situations, prior knowledge about the car and individual expectations and preferences.

In order to provide access to more purposeful "sound engineering", a working understanding of the subjective evaluation process of sound quality is necessary. Sound quality evaluation involves both acoustical and non-acoustical factors, and hence subjective assessments of vehicle sounds can vary substantially due to factors that do not lie within the sound. This variation considerably reduces the predictive power of current sound metrics, and it makes a "good sound" seem a quite fuzzy design target. It is therefore desirable to have a model that describes the subjective evaluation process of vehicle sound quality as experienced by the driver during everyday use, also considering non-acoustical aspects. The aim of this work is to establish such a model that indicates at which stage and in which manner the non-acoustical aspects driver type, vehicle type, vehicle make, driving condition and rating situation influence the assessment of the acoustical aspects comfort, sportiness, type conformity and make conformity, and especially the sound quality judgement.

In the following sections of the introductory chapter, the most important aspects of the methodology used in this PhD thesis will be reviewed. As there is no prevailing classification of car drivers with regard to sound preferences, a typology will be developed in Chapter 2, in order to get a handle on subject-related influences on vehicle sound evaluation. This typology will be constructed from data of a survey among owners of premium vehicles, which deals with the attitude of the participants with regard to vehicles and sounds (including noise sensitivity, see Section 1.2.3.1), as well as with the assessment of, and wishes for, the sound of vehicles. The three resulting driver types will be further characterized with independent items of the same study, providing a first inspection of their plausibility.

In Chapter 3, the subjective evaluation process of vehicle sounds will be investigated in the close-to-reality context of a field study. The focus lies on how the abovementioned acoustical aspects comfort, sportiness, type conformity and make conformity relate to the overall sound quality of a car. To this end, the data of a field study, providing a high degree of ecological validity, will be used to determine a model for the evaluation process by means of statistical methods. The driver typology will be included in the modeling procedure in order to validate its significance in vehicle sound evaluation.

As stated above, results of lab tests cannot be easily transferred to the field situation. Therefore, in Chapter 4 the vehicle sound evaluation process as investigated in Chapter 3 will be compared to an experiment in a lab test setting as a non-realistic evaluation context that is commonly established in research. In order to provide the subjects with the same contextual information as in the field study, however, the cars will be identified to them by name and a picture. The evaluation model presented in Chapter 3 will be tested on these data in order to assess its applicability as well as differences that are due to the situational context.

Finally, in Chapter 5 the results will be summarized and implications on future research will be discussed.

1.2 Methodology

1.2.1 Psychometry

The central question of this work is how humans perceive and evaluate sounds produced by vehicles. However, perception itself is a personal and subjective process and thus cannot be measured "directly", i.e., by using a physical instrument for measurement, as for example a voltage could be measured using a voltmeter. If perception is to be measured, subjects are typically asked to describe, quantify or rate their perception in a certain manner, which constitutes the process of *psychometric measurement*. As such, each subject is object of measurement and measuring instrument at the same time (Blauert and Jekosch, 1997), which is why techniques from psychology form the adequate tool set for the measurement of perception. Some of the most common subjective measurement techniques will be described in Section 1.2.1.2. Prior to that, the important concept of *scale levels* will be introduced in the following section.

1.2.1.1 Scale levels

Especially for subjective measurement techniques, there has been quite some debate about what actually constitutes a "measurement". The answer given by a committee of the British Association for the Advancement of Science in 1946 is relatively simple: "measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules"

(Stevens, 1946, p. 677). Depending on these rules, four different *scale levels* are distinguished, namely *nominal*, *ordinal*, *interval* and *ratio scale level.* which differ in the operations that are meaningful to the data.

A variable with nominal scale level has discrete outcome values that do not have a meaningful natural order. Typical examples are categorical variables, such as numerals for coding gender, nationality or names. The basic operation is determination of equality or inequality of outcomes. One can evaluate counts or frequencies in every category, and the mode, i.e., the category with the most counts, is the measure of central tendency.

A variable with ordinal scale level also has discrete outcome values, but in contrast to a variable with nominal scale level the values can be ordered in a meaningful way (e.g., rank ordered data, sorting words by initial letter). The basic operation is the determination of whether an outcome is greater or less than another outcome. However, no meaningful distance between successive categories can be determined. Percentiles can be evaluated, and the median (as the 50 %-percentile) indicates the central tendency.

For a variable with interval scale level, also the distance between successive categories is meaningful (e.g., temperature on the Celsius scale, the intelligence quotient (IQ) or the sound pressure level in decibels (dB)). The basic operation is the determination of equality of differences between outcomes. Mean and standard deviation can be calculated, but no absolute "zero" exists.

In addition to the properties of the interval scale level, a variable with ratio or metric scale level has a meaningful zero, which allows for division and multiplication of items described by such a scale (e.g., age, length, mass or loudness in sone). The basic operation is the determination of equality of ratios between outcomes, making it possible for example to denote a sound as being "twice as loud as" another sound.

Depending on the psychometric method used to collect subjective data the results will exhibit different scale levels. It is therefore important to be aware of the correct measuring technique and the allowed operations with which data may be processed.

1.2.1.2 Measurement of sensation

Psychometric methods are used to measure subjective sensations. Depending on the method, this can yield qualitative or quantitative results. In this section, some of the most commonly used ways to measure auditory perception will be reviewed.

Free verbalization Free verbalization is a technique that is least restrictive with regard to the responses subjects can give. The subjects are typically asked to verbalize their thoughts on a specific topic, or to answer an open question without predefined answering categories.

This can be realized, for example, in the form of a task given to the subjects, in a (guided) interview or a questionnaire. As the subjects' responses are not forced into a certain rating scale or scheme, the resulting answers have to be further processed in order to draw conclusions. After transcription, several analysis methodologies are available for such qualitative data, like *quantitative* or *qualitative content analysis* (Mayring, 2010) or further techniques as reviewed in Hossain (2011). If the researcher wishes to analyze relationships between subjects' responses using mathematical tools from statistics, the subjects' responses need to be captured in the form of quantitative values. To this end, categorization can be done, which requires the investigator (or any other rater) to develop an appropriate categorization scheme, ending up with data that have nominal scale level. This method, however, is also prone to subjective influences as it relies on the objectivity of the rater, which asks for the inter-rater reliability, i.e., whether or not the results depend on the investigator. However, other methods are available that leave the task of producing quantitative results to the subject, as will be presented in the following sections.

Paired comparison In paired comparison, two stimuli are presented to the subject who has to decide which of the two exhibits more of a given attribute (e.g., which one is louder, brighter, rougher). It reflects a "greater than" (>) or "less than" (<) decision. This is probably the easiest task for the subject, at the cost of a "reduced" measurement level, i.e., ordinal scale level. However, the probabilistic *Bradley-Terry-Luce* (BTL) *model* can be used to deploy a ratio scale, using a sufficiently large dataset of paired comparisons (see Bortz and Döring, 2006). It is based on the assumption that the judgements are intrinsically "noisy", and that the further apart the two stimuli A and B are on the attribute continuum, the greater will be the proportion of responses where stimulus A is rated over B rather than B over A (or vice versa). If, on the contrary, A and B are perceived (almost) equal with regard to the requested attribute, stimulus A will be rated over B in approximately 50 % of the cases. Paired comparison can lead to inconsistent results like the intransitive *cyclic triads*. When three stimuli A, B and C are compared pairwise, it is possible that a subject rates $A > B$, $B > C$ and $C > A$. However, such constellations might happen by chance due to the assumed noisiness in the judgement.

A limitation of paired comparison is that the total number of possible pairings $P = N(N-1)/2$ grows quadratically with the number N of stimuli to be compared, making it infeasible for a complete comparison if N is great. In this case, incomplete designs can be used (Bortz and Döring, 2006).

Magnitude estimation Magnitude estimation requires the subjects to rate each stimulus with regard to a given attribute, such as loudness, brightness or roughness. If no reference is given, the subjects have to make an absolute judgement on an arbitrary scale (*absolute*

(a) quiet

1	2	3	4	5	6	7	8	9	10

loud

(b) dark

−3	−2	−1	0	1	2	3

bright

(c)

not at all	slightly	moderately	quite	very much

Figure 1.5: Exemplary scales used for category scaling.

magnitude estimation). In *relative magnitude estimation*, an “anchor” stimulus is given that corresponds to, e.g., 100 %, and the subjects are asked to rate each stimulus relative to this anchor on the percentage scale defined by it. Hence, the results have ratio scale level.

Both variants of magnitude estimation require the subjects to make a judgement on a continuous scale. This, however, implies an accuracy that is unlikely to reflect the subjects’ ability to differentiate between stimuli. A method to overcome this limitation is *categorical scaling*.

Categorical scaling In categorical scaling, as compared to magnitude estimation, the space of possible answers is discretized to a finite number of (sorted) categories. Typically, the endpoints of the scale are labeled. For example, in the case of loudness these labels could be “quiet” and “loud”, as illustrated in Figure 1.5(a). Such scales are often assumed to be perceived as equidistant, thus exhibiting interval scale level. Typical measures to strengthen the implication of equal distances include the equidistant visual arrangement of the levels and the assignment of consecutive integers to the categories (see Figure 1.5, (a) and (b)). However, interval scale level cannot be guaranteed, especially when each category itself has a verbal label rather than just the endpoints. For example, the four-point loudness scale with labels “(1) not at all, (2) a little, (3) rather, and (4) unbearably loud” has unknown metric properties, as Guski (1997, p. 769) points out. Thus, at least ordinal scale level can be achieved in categorical scaling by using adequate labels (or omitting them). For verbal labels as in Figure 1.5(c), there is literature suggesting that certain sets of verbal scale point labels (VSPLs) are perceived as being equidistant (Rohrmann, 1978, 2007) and can thus be considered as having interval scale level. If there is any doubt on the measurement level of categorical scaling, the *method of successive intervals* (MSI) can be used to transform the data in order to ensure interval scale level (Diederich et al., 1957) and to test the assumption of interval scale level (as done in Schütte et al., 2009).

A disadvantage of categorical scaling as compared to magnitude estimation is the loss of “resolution” as the psychometrical continuum is dissected into discrete bins. A compromise is the *category partitioning scale*, in which each category is again subdivided into several categories (see Guski, 1997).

A measuring technique closely related to categorical scaling is the *semantic differential* (see Bortz and Döring, 2006). Initially developed to measure the connotative meaning or affective qualities of certain objects or words, it consists of a set of antonym pairs, i.e., adjective pairs with opposite meaning, as in Figure 1.5, (a) and (b). Subjects are then asked to rate the object of study, e.g., a certain acoustic stimulus, with respect to this set of adjective pairs, giving a rating on each of the scales. This is typically repeated over several stimuli. The results can be visualized in a *polarity profile*, or clustered by means of an *exploratory factor analysis* (EFA) (see Section 2.2.4.1, or Field (2009)).

An important question in the context of categorical scaling is the number of categories and its parity. If too few categories are used, the results are too coarse and might not reflect subjects' perception. If too many categories are used a similar problem as in magnitude scaling arises, namely the subjects' ability to differentiate between stimuli is exceeded, resulting in the use of only a reduced set of categories (Bortz and Döring, 2006). Whether the researcher chooses an odd number of categories, offering a neutral category in the middle, or whether he/she forces the subjects to decide on the left or the right side by choosing an even number of categories depends on the specific research question. In some cases, one of these two possibilities could be preferred. A disadvantage of the forced choice resulting from an even number of categories is that it suppresses truly neutral responses, and it can "mask the information about the difficulty of an answer", as Guski (1997, p. 769) points out. However, offering a neutral category by choosing an odd number introduces a problem with the interpretation of the responses as the middle category could express indifference ("neither nor"), ambivalence ("it depends") or avoidance of a decision ("don't know" or "don't want to answer") of the subjects (see Bortz and Döring, 2006, Porst, 2014). Also the scale edges can cause problems, in the sense that *floor* or *ceiling effects* can occur, or conversely that subjects try to avoid using the extreme categories and tend towards the center of the scale (Bortz and Döring, 2006). A possible countermeasure is to present the whole set of stimuli to the subjects prior to the actual rating, so that they already know the occupied scale range of the attribute.

Categorical scaling and the semantic differential will be used for most of the ratings acquired in the course of this work, as they constitute a reasonable compromise between ease of use for the subjects, accuracy of the results and resulting scale level. Depending on the specific task, five-point, seven-point or ten-point scales will be used, as will be indicated in the corresponding chapters.

1.2.2 Ecological validity

When performing an experiment, subjects are usually placed in a specific test environment, as for example a listening lab, a room in a certain institution or a driving simulator. In general, this test environment will not be equal to the "natural" environment in which the subject would usually encounter the object of study, e.g., in public places, in their household

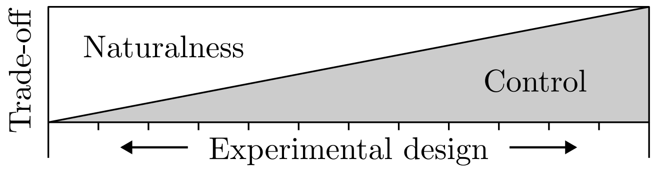


Figure 1.6: Illustration of the trade-off between naturalness of and control over an experiment along the horizontal grading of different experimental designs.

or in the road traffic. Subjective perception and evaluation, however, is moderated by aspects such as "context, ambiance, interactivity, occurrence of several sensory inputs at the same time (combined stimuli)" (Schulte-Fortkamp et al., 2006, p. 1303), which do differ between lab and field test situations. As Fiebig (2015, p. 2558) points out, "the congruence of experimental results obtained in different environments is limited". The property of experimental results to be valid also in real-life situations is known as *ecological validity*. With reference to Wandmacher (2002), Steffens (2013, p. 1473) states:

> "*The degree to which experimental conditions and factors match with factors and conditions which determine our psychological processes in everyday life is called ecological validity*".

When subjects are asked to rate the sound of household appliances in a lab as compared to a close-to-reality context, significant differences of the sound evaluation between both situations can be observed. In Steffens (2013), the sounds of laundry dryers, washing machines and refrigerators were judged more pleasant in the real-life setting of a launderette since there was more distraction by other modalities, whereas loudness seemed to be overestimated in the corresponding lab experiment. In Fiebig (2015), the sounds of different vacuum cleaners were rated differently with respect to the attributes "powerful", "dull", "sharp" and "acceptable" when comparing the data gained in a living room environment to those gained in a listening lab. Ecological validity of laboratory experiments is also important for research regarding soundscape. Guastavino et al. (2005) found that the background noise of a representative city noise environment was attributed to the reproduction system in the lab instead of belonging to the urban ambiance as in the corresponding field study. This result was accompanied with the participants reporting on a lack of envelopment. Even if an experiment is performed in-situ, i.e., with the subjects being tested as part of the real environment, already the mere knowledge of taking part in an experiment can alter the subjects' behavior and responses, which is known as the *observer effect* (see Monahan and Fisher, 2010).

Hence, there seems to be a trade-off between control over experimental conditions on the one hand and providing a realistic context for subjects on the other hand. When setting up an experiment, the investigator has to decide on the balance between the two aspects, as illustrated by the horizontal grading in Figure 1.6. With increasing control over experimental conditions, naturalness typically decreases. Listening experiments in a lab offer a high degree of control over the experimental conditions, and are thus located towards the right side of Figure 1.6. In a listening experiment the investigator has the choice over the exact stimuli, the arrangement of the experimental setting and a reproducible experimental procedure. In addition, lab experiments typically are less time-consuming than investigations in the real-life context, which is why they are often preferred to extensive field studies. Field experiments, on the other hand, allow for observing participants in their usual environment, being subject to the combination of external factors as they also occur in reality, and judging the stimuli as they "happen" in that very moment. Hence, field experiments represent the left part of Figure 1.6. For example, participants might well be in the same mood and mental situation as when they are usually exposed to the stimulus, engage in the same activities and experience the corresponding input from other modalities and integrate it into their judgement. In addition, also interactivity with the stimulus (e.g., as feedback) is typically given in field experiments.

There surely will be situations in which a reasonable compromise between control and naturalness can be achieved. The researcher has to make a conscious decision and deliberately choose an experimental design that fixes some (but probably not all) of the conditions in order to be able to draw ecologically valid conclusions about particular aspects.

The difference between the rating of vehicle sound quality in a field experiment as compared to an experiment performed in a listening laboratory will be the subject matter of Chapter 4.

1.2.3 Personality traits

As pointed out earlier, it is supposed that certain personality traits are able to affect the rating of vehicle sound quality. Along with gender and age, two important concepts regarding personality will be considered in this work, namely *noise sensitivity* and the *Big Five dimensions of personality*, which will be introduced in this section.

1.2.3.1 Noise sensitivity

The annoyance a person experiences from being exposed to noise does not only depend on the noise exposure itself, but also on the person's susceptibility to noise, the so-called *noise sensitivity*. Job (1999) gave the following definition:

> "*Noise sensitivity refers to the internal states (be they physiological, psychological [including attitudinal], or related to life style or activities conducted) of any individual which increase their degree of reactivity to noise in general.*"

Miedema and Vos (2003) found that the effect of noise exposure on noise annoyance is moderated by noise sensitivity. Ellermeier et al. (2001) showed that noise sensitivity is largely unrelated to psychoacoustic indices, meaning that there are no perceptual differences between noise sensitive and insensitive persons. However, small effects of noise sensitivity were found on tasks more closely related to annoyance ratings, such as loudness scaling and unpleasantness ratings, i.e., differences concerning the judgement of sounds. Ellermeier et al. (2001, p. 1471) thus concluded that noise sensitivity reflects “a judgemental, evaluative predisposition towards the perception of sounds” rather than “sensory components” (p. 1472).

In order to assess the extent to which a person is noise sensitive, a questionnaire originally developed by Weinstein (1978) is commonly used. The questionnaire comprises 21 statements (or *items*) that “emphasize affective reactions and avoid general inquiries about noise as an environmental problem” (Weinstein, 1978, p.459). For each statement the person has to indicate his/her level of agreement with this statement on a six-point scale. Weinstein’s questionnaire has also been translated into other languages, such as Swedish (Ekehammar and Dornic, 1990), German (Zimmer and Ellermeier, 1997), Italian (Senese et al., 2012) or Chinese (Fong et al., 2017), with all versions exhibiting good psychometric properties.

Notbohm (2010) extended the one-dimensional construct to the whole “experience of the acoustical environment” in a broader sense than just the relation to noise annoyance. He developed a questionnaire with initially 40 items concerning aspects like “habits of listening to music, attitude towards calmness and silence, perception of sounds from nature, personal habits of causing noise, and, of course, sensitivity to traffic noise as well as to disturbing noises nearby” (Notbohm, 2010, p. 5341). The answers are collected on a five-point scale of agreement adapted from Rohrmann (1978), who found (the German version of) the scale to be perceived as equidistant. It ranges from “not at all”, “slightly”, “moderately”, “quite” to “very much”. After psychometric analyses of the answers of more than 1,100 subjects, there finally remained 28 items that group into four factors of noise sensitivity, which Notbohm calls (i) “critical attitude towards (environmental) noise”, (ii) “activation by music”, (iii) “noise sensitivity and disturbance by noise” and (iv) “appreciation of silence and natural sounds”. The factors (i) and (iii) show a differentiation of the traditional construct of noise sensitivity, while (ii) and (iv) characterize additional aspects of experiencing and responding to the acoustical environment.

With regard to the four factors Notbohm (2010) further identified nine different “types of listeners” that can be partitioned into three clusters. The first cluster combines three types that share a high “critical attitude towards (environmental) noise” together with a low “activation by music”, and are characterized as being *aware of noise*. Conversely, the second cluster combines three types that have high “activation by music” but both low “critical attitude towards (environmental) noise” and “noise sensitivity and disturbance by noise”, which is why they can be called *interested in music*. The third cluster contains types of listeners that do not have very accentuated profiles, and are therefore *without any emphasis*. The examination of further descriptive data about the subjects’ experience of their acoustical environment across the nine clusters revealed meaningful associations, thus confirming the plausibility of the typology (see Notbohm, 2007).

The fourfold concept developed by Notbohm will be used in this work to assess noise sensitivity. In addition, the methodology to develop the driver typology in Chapter 2 will be adopted from the clustering of types of listeners as done by Notbohm (2007).

1.2.3.2 Big Five dimensions of personality

The *Big Five dimensions of personality* are a model of personality widely used in psychology, which is based on the lexical description of interindividual differences (Rammstedt et al., 2013). The Big Five consist of the five scales (i) neuroticism, (ii) extraversion, (iii) openness to experience, (iv) agreeableness and (v) conscientiousness. McCrae et al. (2005) further describe each scale by several facets: neuroticism is comprised of the facets anxiety, angry hostility, depression, self-consciousness, impulsiveness and vulnerability. Extraversion consists of the facets warmth, gregariousness, assertiveness, activity, excitement seeking and positive emotions. Openness to experience incorporates the facets fantasy, aesthetics, feelings, actions, ideas and values. Agreeableness includes the facets trust, straightforwardness, altruism, compliance, modesty and tender-mindedness. Conscientiousness relates to the facets competence, order, dutifulness, achievement striving, self-discipline and deliberation. The Big Five model has meaningful associations with various individual and social processes and phenomena (see Rammstedt et al., 2013, for the corresponding overview in German).

A common questionnaire for assessing the Big Five, the NEO-Personality Inventory (NEO-PI-3), comprises 240 items (McCrae et al., 2005), and its short form, the NEO-Five Factor Inventory (NEO-FFI-3), still consists of 60 items (McCrae and Costa Jr., 2007). Especially for studies with limited time, Rammstedt et al. (2013) developed a short scale for assessing the Big Five in German, the 10 Item Big Five Inventory (BFI-10). In the BFI-10 each factor is covered by two items with opposite polarity. The answers are collected on a five-point scale with labels "does not apply at all", "does rather not apply", "neither nor", "rather applies" and "totally applies".

The Big Five dimensions of personality will be included in the questionnaire concerning personality used in Chapter 3.

2 Individual influences on the evaluation of vehicle sounds: a typology of premium car drivers with regard to their attitude towards cars and sounds[1]

It is widely agreed that apart from acoustical properties personal factors such as expectation, experience or attitude also influence the evaluation of product sounds. However, there is a lack of concrete knowledge about the nature of this influence on vehicle sound evaluation. Vehicle sound design or the development of sound metrics could benefit from a better understanding of the role of personality in sound evaluation. This article aims to develop a characterization of premium car drivers with regard to their personality. This characterization will be used for further analysis of personal influences on the perception and evaluation of vehicle sounds. For this purpose an analysis was made of data from a survey that dealt with the assessment of, and wishes for, the sound of vehicles, as well as the attitude of the participants with regard to vehicles and sounds. More than 1,800 participants who were familiar with vehicles of premium makes, i.e., makes with a high perceived quality and value, took part in the survey. The analysis of the results reveals three driver types that show significant differences in vehicle sound evaluation, namely the sound-oriented fun-driver, the noise-sensitive aesthete and the sound-uninterested pragmatist.

[1] This chapter is based on Maiberger et al. (2018a):
D. Maiberger, U. Letens, R. Weber, S. van de Par (2018): Individual influences on the evaluation of vehicle sounds: a typology of premium car drivers with regard to their attitude towards cars and sounds. Acta Acustica united with ACUSTICA 104(3), 2018, 509–520.

2.1 Introduction

Vehicle sounds belong to the class of product sounds. The evaluation of product sounds is in general considered to be a complex process that does not just involve acoustic factors, but is also subject to personal influences. Blauert and Jekosch (1997) introduced "cognition, action and emotion" as moderating factors in the quality assessment of product sound. According to these authors, a judgement also involves "information from non-auditory senses, prior knowledge and association to actions and emotions", and therefore "auditory perception is not at all predetermined solely by the acoustic input to the auditory system, but results from a complex interaction of auditory input, non-auditory input, expectation and mood". Similarly, Bednarzyk (1999) states that sound quality evaluation of product sounds is subject to cognitive influences, such as "individual expectation, habit, experience, personal taste, collective taste (zeitgeist[2]), attitude and motivation, social and cultural influences". But the relationship of the listener towards the sound source also plays a role. For example, the assessment of the sound of a vacuum cleaner through the (operating) user is different from that through a passive listener. Fiebig and Kamp (2015) emphasize the importance of the selection of participants in sound quality assessment studies and show an example of differing evaluation strategies for the sound quality of electric screwdrivers between professional and free-time users due to "different expectations of the groups".

Although it seems to be generally agreed that personal factors do also affect the assessment of vehicle sounds (see Bisping, 1994, Jung and Grützmacher, 2011, Letens, 2010), in most studies about vehicle sound evaluation the inter-individual variance is not further analyzed, but rather averaged out; at best, effects of age, gender or expertise (with regard to the specific task) are considered. This seems justified as long as the sole interest is to examine the variation introduced by properties of the sound or sound source, e.g., loudness, sharpness, roughness, or motorization, car segment (e.g., compact cars, full-size luxury cars or sports cars), car make or driving situation. When it comes to investigating the vehicle sound evaluation process as a whole, however, factors that lie *within* the individual should also be taken into account. For example, Strasser et al. (2016) point out that the personality traits "neuroticism" and "openness to experience" have a significant influence on the evaluation of vehicle sounds: the higher "neuroticism" or "openness to experience", the higher the rated vehicle sound quality[3].

When it comes to the practical application of such findings, e.g., in the process of vehicle sound design or the development of metrics for vehicle sound quality, there is another challenge: in addition to the general increase in complexity, it is usually not known how the actual drivers can be characterized in terms of their personality traits, or to be more specific: the sound

[2] Zeitgeist is "the general intellectual, moral, and cultural climate of an era" (Merriam-Webster).

[3] These two personality traits, along with "conscientiousness", "extraversion" and "agreeableness", form the *Big Five dimensions of personality* known from psychology (see Rammstedt et al., 2013).

designer does not know to what extent the drivers *really* are neurotic or open to experience — so for whom should the vehicle sound eventually be optimized?

It therefore seems useful to have a model for personality traits that can account for inter-individual variance in vehicle sound evaluation and which also offers an appropriate simplification in terms of complexity as well as real-world configuration of personality traits. In the automotive industry, a basic distinction is made between sports car fans and luxury car enthusiasts (as in Altinsoy, 2015). However, this distinction is based upon the possible car positionings as sporty or comfortable, and there are little substantiated findings about, e.g., the number of different driver types, how they relate to other personality traits, and how to determine the driver type of a particular person.

The aim of this study is to create a typology of car drivers with regard to personality as a basic concept for further analysis of personal influences on the evaluation of vehicle sounds. For this purpose, the results of a survey are examined as an exploratory approach. The associated questionnaire was designed to gather information about people's opinion and personality on the one hand, but on the other hand with respect to the perception, evaluation and wishes regarding the sound of their vehicle. The results of the first part of the questionnaire will be used to develop a typology of car drivers, and the second part will be used to test whether the developed typology provides a suitable partitioning of participants with regard to vehicle sound assessment.

2.2 Methods

The development of the driver typology was based on data from a survey among drivers of premium vehicles. This data was used in two ways: first, from certain personality traits a characterization (or typology) of car drivers was developed; second, with the participants' judgements on vehicle sounds, this typology was checked for plausibility. The methodology of the data acquisition is described in this section.

2.2.1 Sample

The participants were taken from a panel that is frequently asked to take part in surveys concerning all sorts of automotive topics. More than 90 % of the participants own at least one premium vehicle, and are therefore presumably interested in cars. For this reason it cannot be expected that the sample is representative for the population as a whole. The panel members get a gift when taking part in a certain percentage of all surveys offered throughout the year. All participants reside either in Germany or the USA. In addition to the contents of the questionnaire described in Section 2.2.3, demographic data of all panel members were available.

2.2.2 Procedure

The questionnaire was implemented as an interface for web browsers. Prior to the strictly textual questionnaire which requested vehicle sound assessment of the participants, seven sound examples were offered in the introduction as preparation and focusing the participants on the topic matter of the survey. Those sound examples covered sports cars and sedans, which were recorded with an artificial head while driving off from standstill, accelerating, overtaking, idling or driving at a constant speed. Some sounds were recorded inside, others outside the vehicles. The actual reproduction quality of the sound examples strongly depended on the reproduction system the participants used, and thus was not controllable. Given that the sound examples only served as contextualization for the test, this was not considered to be a problem.

When answering the questions, the participants were asked to refer to the car they drive most often. Those participants owning a sports car besides their regular car were asked to refer to the sports car to ensure that there were enough assessments of sports cars in the survey. The car the participants referred to will be called their "own car" throughout this article.

2.2.3 Contents of the questionnaire

The questionnaire used in the survey contained open questions, multiple-choice questions, statements for which the participants could indicate to what degree they agree, as well as semantic differentials consisting of contrastive adjective pairs to describe vehicle sounds. The contents covered a broad range of aspects of vehicle sounds. The composition of questions within the survey (or *items*) was based on interviews with experts in vehicle acoustics, psychoacoustics as well as psychology, both from university and industry, as well as on an extensive literature research (e.g., Bednarzyk, 1999, Busse, 2014, Sukowski et al., 2013, Altinsoy et al., 2010, Hempel, 2001, Notbohm, 2010, Guski, 1999, Notbohm et al., 2013, Schulz et al., 2000, Schulz and Fröming, 2008, Taubman-Ben-Ari et al., 2004). The questionnaire was tested and refined in several pretests with naive participants, particularly with regard to comprehensibility and unambiguity. The final questionnaire consisted of the following parts:

1. Introduction to the topic "driving sounds", including sound examples;
2. Questions concerning the own car, e.g., make, model, motorization, assessment of the driving sounds of the participants' own car;
3. Questions concerning how a follow-up car should sound;
4. Questions concerning opinion and personality, e.g., opinion on noise in general, opinion on cars and driving, opinion on driving sounds in general;
5. Additional remarks, thanks and farewell.

Table 2.1: Adjective pairs of the semantic differentials used for the assessment of overall vehicle sound of the own vehicle and the ideal follow-up vehicle. The word `make` was replaced with the make of the participant's car.

No.	Adjective pair		
1	loud	–	quiet
2	calming	–	exciting
3	aggressive	–	defensive
4	rough	–	smooth
5	bright	–	dark
6	comfortable	–	sporty
7	restrained & elegant	–	striking & bold
8	beefy	–	mellow
9	howling	–	not howling
10	unpleasant	–	pleasant
11	high quality	–	low quality
12	untypical of `make`	–	typical of `make`
13	don't like it at all	–	like it a lot
14	familiar	–	innovative

In Part 1 the survey was introduced as dealing with the "future of driving sounds". The term "driving sounds" was defined to the participants as the "sounds generated by the engine, by the suspension, by the tires rolling on the road and by the airstream passing over and around the vehicle during travel", making the participants aware of the typical sound sources in a car. The participants were asked not to refer to "rattles and squeaks or other such noises that point to a defect in the vehicle", in order to avoid interference by these phenomena that were not of primary interest in this study.

In Parts 2 and 3 of the questionnaire the participants were asked to give information about their car by means of multiple-choice and open questions, and to rate its sounds by means of semantic differentials. For these semantic differentials, the adjective pairs listed in Table 2.1 were used with a seven-point bipolar category scale. For the composition of this list, some results of Busse (2014) were taken into account. In the cited work, several listening experiments were carried out to iteratively reduce an extensive list of contrastive adjective pairs applied to vehicle sounds. A factor analysis (cf. Section 2.2.4.1) finally yielded six factors: "evaluation", "perceived value", "driving dynamics", "impulsiveness", "frequency modulation" and "metallic character". Other studies dealing with semantic differentials for the evaluation of vehicle sounds are Bednarzyk (1999), Sukowski et al. (2013), Altinsoy et al. (2010), Hempel (2001). For the study presented here, the adjective pairs occurring most frequently in the aforementioned studies were chosen as a basis. As it was not the main goal to have a "complete" list of adjective pairs but rather to cover the main aspects of vehicle sounds and to avoid doublings while maintaining the participants' cooperativeness, the selection of adjective pairs was further reduced and adjusted. The final list covers all the aforementioned factors of Busse except "impulsiveness", which has shown to have a rather low consistency (Busse, 2014, p. 92). The items were used for determining whether the typology is capable of explaining inter-individual variance in the evaluation of vehicle sounds.

Table 2.2: Items used for the development of the driver typology, with their respective MSA (cf. Section 2.3.2).

No.	Item	MSA
1	I believe that noise is a major environmental problem.	.915
2	Music helps to put me in a good mood.	.824
3	If there's a lot of loud noise around me I can't concentrate on my work at all.	.802
4	I really love listening to the sounds of nature.	.832
5	I am generally very sensitive to noise.	.820
6	I very often get upset about traffic noise.	.862
7	I believe that cars are a major environmental problem.	.829
8	I think that road traffic has a lot of negative effects on people's health.	.818
9	I think cars should become quieter and quieter in the future.	.913
10	A car should emit sounds so that it gets noticed better by other road users.	.822
11	For myself, the driving sound is important in order to indicate how fast I am going and which gear I am in.	.894
12	I don't care how a car sounds as long as it runs well.	.858
13	I normally pay attention to how my vehicle sounds.	.882
14	To me it's important that car makers design a good driving sound for their cars.	.897
15	I really enjoy hearing a nice sounding car.	.921
16	I want to impress people with the sound my car makes.	.912
17	When driving I feel best when cruising down the road.	.809
18	I keep calm when other drivers speed or tailgate.	.697
19	When driving I am easily annoyed by other drivers that dawdle.	.742
20	I think it's fun to drive fast through curves and down serpentine roads.	.878
21	To me, a car is just a tool.	.869
22	I don't care if my friends and acquaintances don't like my car, as long as it's practical and gets the job done.	.855
23	To me, a car is also a status symbol.	.853
24	I want to drive cars that are unique.	.919
25	When driving alone it's important that I have a car with comfortable driving characteristics.	.745
26	When driving alone it's important that I have a car with sporty driving characteristics.	.899

In Part 4 of the questionnaire, a list of statements was presented to the participants. For each statement, the participants were asked to state how much they agreed on a five-point scale with labels "not at all", "slightly", "moderately", "quite" and "very much" (adopted from Notbohm, 2010). These scale labels were chosen to be perceived as equidistant (see Rohrmann, 1978, 2007), i.e., as having interval-scale level, which will be tested in Section 2.3.2 using the *method of successive intervals* (MSI) (Diederich et al., 1957). The statements are shown in Table 2.2, together with their respective *measure of sampling adequacy* (MSA, see Section 2.2.4.1). In studies concerning environmental noise and its effects on listeners (Notbohm, 2010, Guski, 1999, Notbohm et al., 2013), personality has shown to have a significant influence on the perception of noise. Statements 1 to 6, which deal with noise sensitivity in general, were adopted from Notbohm (2010). Statements 7 and 8, which deal with the general evaluation of the car, were inspired by Guski (1999). Regarding the perception of vehicle sounds, statements 9 to 16 were proposed in the expert interviews. They deal with aspects such as the wish for quieter cars, warning and feedback function of vehicle sounds, caring about and paying attention to vehicle sounds, importance of a good sound, enjoying of nice sounding cars and the wish of impressing people with vehicle sounds. Statements 17 to 20 were inspired by Schulz et al. (2000), Schulz and Fröming (2008), Taubman-Ben-Ari et al. (2004), and cover aspects of a relaxed, calm driving style as well as being easily annoyed by slow drivers and the tendency to drive fast. Also inspired by Schulz et al. (2000), the experts proposed the statements 21 to 24, which deal with the status aspect and uniqueness of cars as opposed to the notion of a car as being "just a tool". In the automotive industry, a principal distinction is made between comfortable (e.g., luxury car) and sporty (e.g., sports car) driving characteristics. Statements 25 and 26 deal with the driver's respective preferences. The items in Table 2.2 were used for developing the driver typology.

Since the study was designed in Germany but was also conducted in the USA, the whole questionnaire was also translated into English by professional translators and native speakers of English.

2.2.4 Statistical methods

2.2.4.1 Exploratory factor analysis

Exploratory factor analysis (EFA) is a method for finding the underlying factors of a given set of variables with a certain number of observations. The aim is to reduce the complexity and dimensionality of the dataset and make it easier to interpret. For an extensive overview of EFA see Field (2009). For this study the factor extraction was done using *principal component analysis* (PCA), and the subsequent rotation was performed according to the *Varimax*-criterion, which yields orthogonal factors[4]. The appropriateness of the dataset for EFA was checked with the Kaiser-Meyer-Olkin *coefficient* (KMO) for the whole dataset, and the related *measure of sampling adequacy* (MSA) for each variable (also called KMO in Field, 2009). For determining

[4]Using this common criterion and thus assuming orthogonality of factors is not free from criticism. However, comparison of the final result (see Section 2.3.2) with the outcome of factor rotation according to the (oblique) *Promax*- and *Oblimin*-criteria showed similar factor structures.

the number of factors to be extracted, the KAISER-GUTTMAN *criterion* (Guttman, 1954) and the *scree test* by CATTELL (Cattell, 1966) were applied. As the choice of the number of factors is also a question of how well the result matches with theory (Finch, 2013, p. 174), several EFAs were carried out and compared in terms of interpretability. CRONBACH's α was used as a measure of the internal consistency of the factors (see Field, 2009).

2.2.4.2 Cluster analysis

Cluster analysis is a method for finding groups of observations that are similar with respect to certain variables (see Rapkin and Luke, 1993). The procedure applied in this study is a combination of a *hierarchical method* (to find the number of clusters) with an *iterative method* (to refine the partitioning), as recommended in Rapkin and Luke (1993, p. 269). For the "hierarchical pre-clustering", an agglomerative approach was chosen with the squared Euclidean distance measure, using WARD*'s method* (Ward Jr., 1963). The correct number of clusters can be determined from an "elbow" in the plot of the change in the error sum of squares (ESS) over iteration steps (Timm, 2002, p. 531). As the results of this method depend on the order of observations, the clustering procedure was performed with several randomized orders. For the "iterative post-clustering", the *k-means*-algorithm was applied. Since the results of this algorithm heavily depend on the initial partitioning, all partitionings found by the hierarchical pre-clustering (arising from the different orders) were fed into the post-clustering as initial partitionings.

The reliability of the found partitioning was checked by the procedure described in Morey et al. (1983). However, COHEN's κ (Cohen, 1960) was used as a measure for the similarity of the yielded cluster partitionings.

2.2.4.3 Statistical tests

Multivariate analysis of variance (MANOVA) is a statistical test to determine whether there are mean differences between different groups with regard to several dependent variables, where the groups are usually defined by one or more categorical variables (see Field, 2009). WILKS' *lambda* was used as the test statistic for deciding on the significance of MANOVA. The *effect size* of each categorical variable is represented by its *partial* η^2, denoted as η_p^2. A value of $\eta_p^2 \approx .01$ was interpreted as a "small", $\eta_p^2 \approx .06$ as a "medium" and $\eta_p^2 \approx .14$ as a "large" effect (Richardson, 2011, p. 140). For all statistical analyses in this study, a significance level of $\alpha = 5\,\%$ was used. In order to avoid familywise error inflation in post hoc tests, the BONFERRONI *correction* was applied (Field, 2009).

Discriminant analysis was used to test for age differences between different groups of participants (see Field, 2009). PEARSON*'s chi-square test* was used to test for independence of two categorical variables (see Field, 2009).

2.3 Results

For the construction of the driver typology, groups (or clusters) of participants that responded in a similar way to the questions concerning opinion and personality were searched for by means of a cluster analysis; as a "preprocessing" of this, the corresponding items were first condensed into six representative factors of personality by means of an EFA (cf. Rapkin and Luke, 1993, p. 262). After an overview of the sample demographics the EFA, as well as the cluster analysis, will be described. In Sections 2.3.5 and 2.3.6 it will be reported how the participants' demographics and judgements of vehicle sounds were used to determine if the typology makes meaningful distinctions.

2.3.1 Sample description

On average, it took the participants less than 30 minutes to finish the questionnaire. Five participants had to be excluded because they missed too many answers. Another 24 participants were excluded because they finished the questionnaire in less than ten minutes. Given the number of items, it did not seem reasonable that they completed the questionnaire carefully and conscientiously.

A total of n = 1,837 participants were left for the analyses, 20.1 % of which were female and 79.9 % were male (see Table 2.3). The average age was 56.9 years. As stated above, the participants were asked to refer to the car they drive most often. In Table 2.3 those car makes that account for at least 1 % of the total sample are listed. For the five most frequent makes in the sample (i.e., Audi, BMW, Mercedes, Volkswagen, Porsche) the fractions in different car segments are listed. The five segments distinguished in the following are "compact cars", "mid-size cars and mid-size luxury cars" (abbreviated as *mid-size (luxury) cars*), "full-size luxury cars", "sports utility vehicles (SUV)", and "sports cars, coupés and convertibles".

2.3.2 Factor analysis of opinion and personality items

Prior to EFA, the scale level of the participants' responses to the items in Table 2.2 was analyzed. For this purpose the mean values across the observed ratings were compared to the scale values resulting from MSI, as done in Schütte et al. (2009). It was found that the mean absolute deviation between observed and predicted cumulative percentages from MSI lay between 0.24 % (Item 23) and 2.69 % (Item 6), and the correlation between mean values and scale values of the items was r^2 = 99.8 %. Therefore the observed ratings were further considered as having interval-scale level.

For the 26 items shown in Table 2.2 the KMO was .871, which according to Kaiser and Rice (1974) is "meritorious", and all MSA (see Table 2.2) lay between .697 (for Item 18) and .921

Table 2.3: Demographics for the German and the American subsample and the total sample regarding age, gender, make of car (> 1 % of total) and vehicle category. Percentages indicate the fraction of the respective (sub)sample in the column.

	Germany	USA	Total
No. of participants	1228	609	1837
Age Ø	55.4 yr	60.0 yr	56.9 yr
Female	17.0 %	26.3 %	20.1 %
Male	83.0 %	73.7 %	79.9 %
Audi	11.1 %	1.1 %	7.8 %
BMW	13.0 %	2.3 %	9.5 %
Mercedes	61.6 %	85.2 %	69.4 %
Porsche	1.2 %	1.5 %	1.3 %
Volkswagen	5.5 %	0.2 %	3.8 %
Compact car	17.4 %	0.2 %	11.7 %
Mid-size (luxury) car	37.9 %	36.3 %	37.4 %
Full-size luxury car	6.3 %	19.5 %	10.7 %
SUV	10.9 %	10.5 %	10.8 %
Sports car/ coupé/convertible	17.8 %	23.2 %	19.5 %

(for Item 15), which ranges from "middling" to "marvelous". The Kaiser-Guttman criterion yielded seven factors for the current dataset, which is usually considered an upper bound for the number of factors (Field, 2009, p. 641). In the scree plot (see Figure 2.1) no clear elbow was found, but at around four factors the plot flattened; between seven and eight factors there was another drop of the eigenvalues. Therefore several EFAs with four to seven factors were carried out and the results were compared in terms of interpretability. Starting from the solution with four factors, it was observed that the factor structure did not change entirely when increasing the number of factors, but rather single factors of the previous solution split up into new factors representing distinct aspects of the previous factor. This was the case for up to six factors. The solution with six factors thus offered the highest degree of differentiation, and was more plausible than the solution with seven factors that linked seemingly unrelated items and contained a new single-item factor. Therefore, the solution with six factors was chosen, which can be characterized as follows:

1. **"dynamic driving style & acoustical orientation"** embodies the disposition to drive in a dynamic and sporty manner on the one hand, but on the other hand the wish for a functional vehicle sound is also expressed, e.g., as a warning for other road users or as orientation for the driver.
2. **"noise sensitivity"** encompasses most items adopted from Notbohm (2010), and represents aspects of noise sensitivity and being easily disturbed by noise.

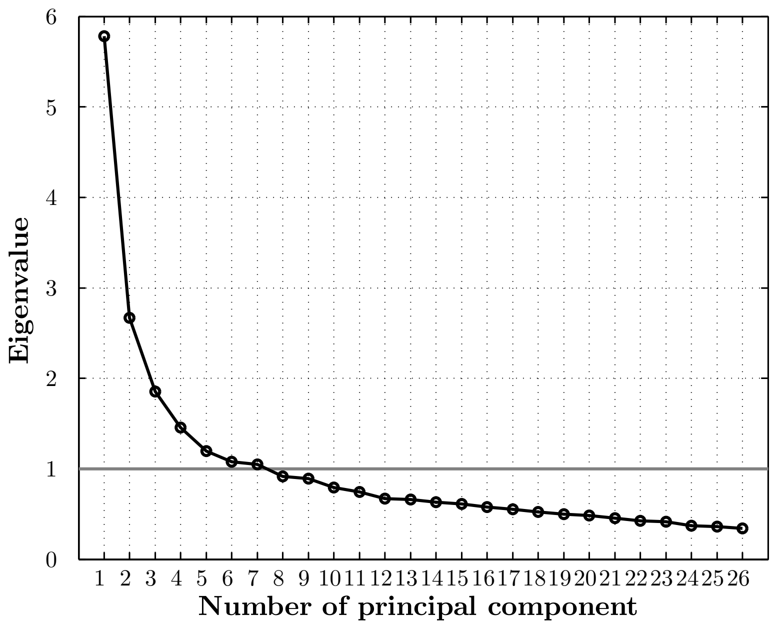


Figure 2.1: Scree plot for the factor analysis of opinion and personality items.

3. **"sound awareness"** embodies the awareness of the vehicle sounds, which are an important factor and under special attention.
4. **"status aspect"** encompasses the items describing cars as either being just a tool or being a status symbol.
5. **"reservations about cars"** expresses a feeling of health and environment being threatened by cars and traffic, stated by the items inspired by Guski (1999).
6. **"demand for comfort"** embodies the wish for comfortable and unstressed driving.

Cronbach's α was calculated for every factor by taking into account all items that have a loading with an absolute value greater than .5 on this factor. For the six factors, α ranged from .513 to .782. When excluding Items 18 and 19 the values for α of the corresponding factors increased; these two items were also the ones with the lowest MSA in the dataset (cf. Table 2.2), and after exclusion the KMO of the dataset rose slightly. When repeating the EFA with the 24 remaining items, the composition of the factors stayed the same as with all 26 items, with some deviations of the loadings, which is why this solution was further considered. The loadings of the 24 items on the six final factors are shown in Table 2.4. It was observed that Items 24, 4, 9 and 2 did not have any loadings greater than .5, and thus were not considered further.

By calculating a factor-wise mean raw score, a *scale value* per participant and factor can be obtained, which indicates how much the particular factor is pronounced for the particular participant. These scale values were used in a cluster analysis in order to determine whether groups of participants could be found that responded in a similar way, as will be described in the following section. In this article, the scale values were derived by taking into account the

Table 2.4: Factor loadings of the items for the solution with six factors, after exclusion of Items 18 and 19 for matters of factor consistency. Absolute values over .5 are printed in bold, absolute values below .35 are printed in gray. The final factor structure results from the bold numbers. Variables' communalities are given in the rightmost column, explained variance and CRONBACH's α in the two lowest rows.

	Factor						
Item	1	2	3	4	5	6	Com.
26	**.708**	.010	.159	-.160	-.212	-.150	.619
10	**.694**	-.098	-.111	.084	.061	.016	.515
20	**.664**	.060	.160	-.016	-.226	-.089	.530
15	**.640**	-.126	.424	-.063	-.197	-.053	.651
16	**.601**	-.209	.181	-.390	.026	-.109	.602
11	**.593**	-.158	.120	-.098	.173	-.013	.430
24	.441	.021	.300	-.434	-.095	.167	.509
5	-.154	**.777**	.018	-.063	.144	.076	.658
3	-.100	**.739**	-.002	-.143	.070	.027	.583
1	-.198	**.567**	-.122	.259	.358	.141	.591
6	-.040	**.535**	-.165	.162	.485	.074	.581
4	.190	.497	-.019	.382	.028	.178	.462
9	-.377	.389	-.153	.240	.363	.237	.563
12	.016	-.012	**-.741**	.187	.098	.219	.643
13	.230	-.020	**.714**	-.061	.016	.082	.575
14	.437	-.090	**.583**	.035	-.035	.136	.559
2	.098	-.098	.329	-.172	.170	.221	.235
22	.006	.028	-.046	**.688**	.052	.127	.495
23	.327	.024	.113	**-.687**	-.017	.194	.630
21	-.052	.014	-.465	**.539**	.265	.078	.587
7	-.074	.158	-.031	.081	**.763**	-.055	.624
8	-.038	.177	.027	.012	**.740**	.093	.590
25	-.165	.152	.078	-.003	-.054	**.769**	.650
17	-.034	.119	-.050	.058	.117	**.754**	.603
% var.	14.1	9.9	9.2	8.4	8.2	6.4	
α	.790	.745	.629	.596	.610	.513	

items that had loadings > .6 on one factor but loadings < .35 on all other factors. This was a stricter criterion than the one used in the analyses before, and was a pragmatic choice in order to obtain scale values that uniquely describe each factor on the one hand, and to have at least two items per factor for robustness on the other hand. Hence, Items 26, 10 and 20 were used to calculate the scale value for Factor 1, Items 5 and 3 for Factor 2, Items 12 and 13 for Factor 3, Items 22 and 23 for Factor 4, Items 7 and 8 for Factor 5, and Items 25 and 17 for Factor 6.

In Figure 2.2 the sample means of all scale values are shown. While most of these scale values were located around the middle of the scale, the demand for comfort had a high level throughout the sample. This might have been due to the sample consisting of owners of premium vehicles.

After determining these six scales concerning opinion and personality, the next step was to

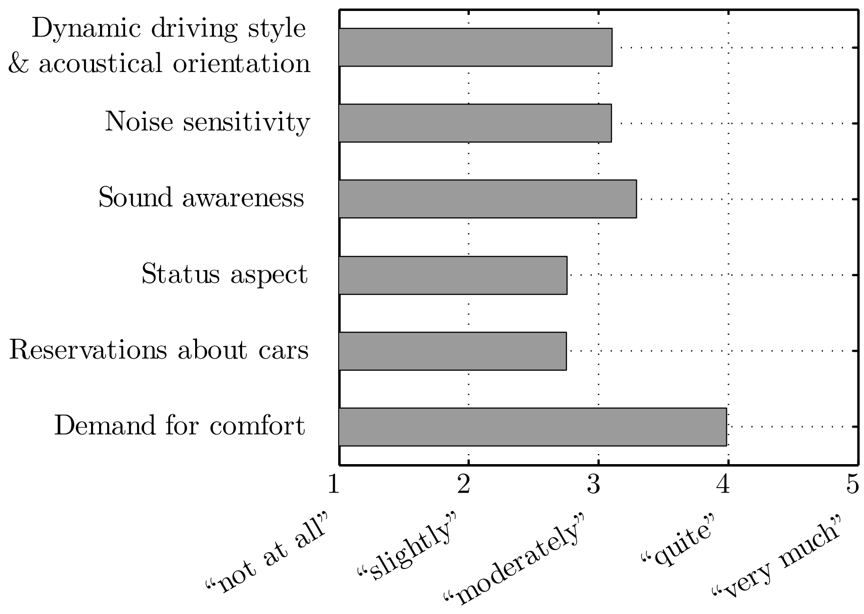


Figure 2.2: Sample means of the scale values on the six factors.

investigate whether these occur in random combinations across different drivers, or whether there are typical combinations that occur more frequently together in particular drivers. If such combinations exist, they relate to certain “driver types”. The steps performed in the course of a cluster analysis are described in the following section.

2.3.3 Clustering of participants

A cluster analysis was performed in order to find groups of participants that responded similarly to the questionnaire in terms of the six factors, represented by the corresponding scale values. Based on such clusters, it was then possible to determine a typology of the participants and to investigate how it relates to their judgement of sounds.

2.3.3.1 Hierarchical pre-clustering

The analysis started with one cluster for each participant and the algorithm proceeded by iteratively merging two clusters. In addition to the “normal” order of participants, denoted by O_0, the clustering procedure was repeated with six randomized orders, denoted by O_1 to O_6. In Figure 2.3, the change in ESS over the number of clusters is shown. When considering the trend of the mean for the original data it could be seen that the rise in ESS occurred roughly when merging $k = 4$ into 3 clusters. For two of the seven orders the rise occurred when merging $k = 3$ into 2 clusters. Hence, a number of $k = 3$ or $k = 4$ clusters seemed reasonable.

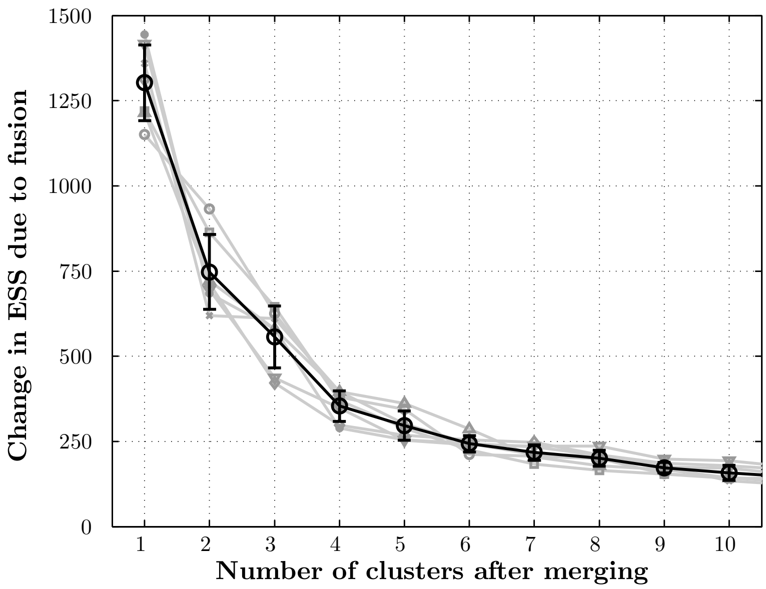


Figure 2.3: Change in error sum of squares over number of clusters left after merging in agglomerative clustering for different orders (in gray); means and standard deviations (over different orders) are printed in black.

2.3.3.2 Iterative post-clustering

As the results of the k-means-algorithm heavily depend on the initial partitioning, all partitionings found by the hierarchical pre-clustering (arising from the orders O_0 to O_6) were fed into the post-clustering as initial partitionings. For $k = 3$ clusters, all orders practically resulted in the same partitioning after the k-means-algorithm. For $k = 4$ clusters, however, the partitionings differed. This was due to order O_5 providing a partitioning that was different from the other ones. As all orders were considered equivalent, this was taken as evidence for just $k = 3$ clusters in the data. For $k = 3$ clusters, the values of Cohen's κ lay above .989 for all comparisons of orders, but for $k = 4$ clusters, order O_5 just reached values of $\kappa \leq .463$ in the comparisons with the other orders. Therefore, the final partitioning with $k = 3$ clusters was considered further, as this partitioning was yielded regardless of the used order.

2.3.3.3 Reliability of the clusters

When testing the reliability for order O_0 as described in Morey et al. (1983), again $k = 3$ to 4 clusters were implied by the plots of the change in ESS over the number of clusters (analog to Figure 2.3). When comparing the partitionings yielded from both groups, it showed that for $k = 3$ clusters the partitionings were the same ($\kappa \geq .974$), while for $k = 4$ clusters the partitionings differed ($\kappa \leq .414$). From this it was concluded that the clusters can indeed be reliably determined from the data.

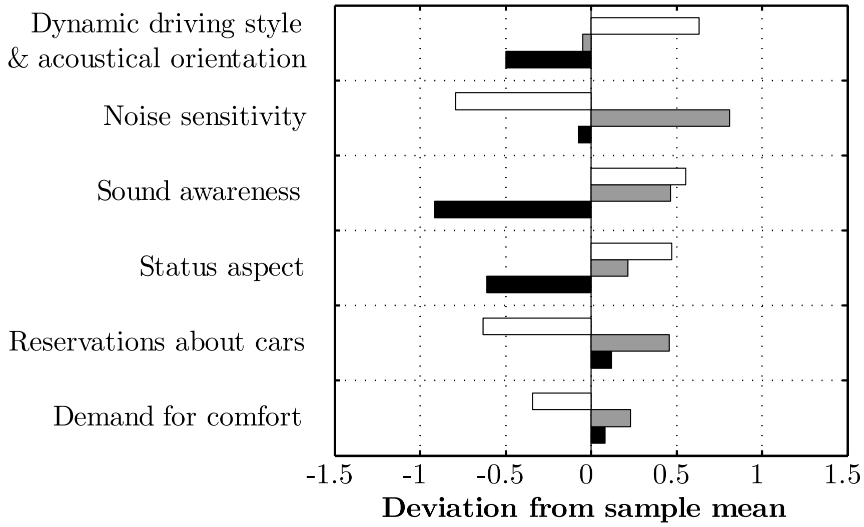


Figure 2.4: Mean scale values on the six factors relative to the sample mean for the three driver types, i.e., the *sound-oriented fun-driver* (white), the *noise-sensitive aesthete* (gray) and the *sound-uninterested pragmatist* (black).

2.3.4 Premium car driver typology

The three clusters found in the previous section shall be interpreted here as three "premium car driver types". The mean scale values on the six factors relative to the sample mean (cf. Figure 2.2) for the three driver types are shown in Figure 2.4. The driver types can thus be characterized as follows:

Type 1 (562 participants) has a high value on the factors "dynamic driving style & acoustical orientation", "sound awareness" and "status aspect", so this type of driver seems to care about the sounds his/her vehicle emits, drives in a sporty manner and to him/her a car is also a status symbol. Especially the factors "noise sensitivity" and "reservations about cars" are less pronounced for this type. Type 1 shall therefore be called the ***sound-oriented fun-driver***.

Type 2 (609 participants) has a value on the factor "sound awareness" that is as high as for Type 1[5], but in contrast Type 2 has also high values on the factors "noise sensitivity" and "reservations about cars", which makes him/her seem more critical. It seems that they both care about their vehicle sounds in equal measure, but for different reasons: Type 1 does so due to his/her sporty driving style, and Type 2 due to his/her high noise sensitivity. Type 2 shall therefore be called the ***noise-sensitive aesthete***.

Type 3 (649 participants) has substandard values on the factors "dynamic driving style & acoustical orientation", "sound awareness" and "status aspect", with the rest of the scale

[5]This is the only case for which the difference between two driver types is not statistically significant (one-way MANOVA ($\Lambda = .18$, $F(12, 3624) = 415.46$, $p < .05$, $\eta_p^2 = .58$), followed by post hoc tests).

values being close to the sample average. Type 3 sees his/her car rather as being a tool, and does not care much about vehicle sounds. Type 3 shall therefore be called the ***sound-uninterested pragmatist***.

The presented driver typology emerged from items concerning the participants' opinion and personality. However, there is not necessarily an effect from dividing the participants into those driver types on their assessment of vehicle sounds. If there *is* indeed an effect, it might be worth considering the driver typology in future studies about vehicle sound perception. In the following sections, the driver types will be characterized in terms of demographics as well as their assessment of vehicle sounds.

2.3.5 Characterization of the driver types

In Table 2.5, different subsamples are compared in terms of the distribution of the three driver types within the respective subsample. For the continuous variable age, a discriminant analysis was used to test for differences between the three driver types; with a chi-square test, the discrete counts underlying the remaining row categories in Table 2.5 were tested for independence from the driver type, i.e., the country of residence, gender, car make and car segment. All of these tests yielded significant results[6], meaning that there is indeed a significant dependency between the driver type and these five categories (Age: $\Lambda = 0.966$, $\chi^2(2) = 62.111$, $p < .01$, canonical $R^2 = .03$; country: $\chi^2(2) = 76.86$, $p < .01$; gender: $\chi^2(2) = 14.22$, $p < .01$; make: $\chi^2(8) = 26.88$, $p < .01$; segment: $\chi^2(8) = 86.40$, $p < .01$). Post hoc tests showed that the *sound-oriented fun-drivers* were about 5 years younger compared to the other two driver types. From Table 2.5 it was concluded that there are more *sound-uninterested pragmatists* and fewer *noise-sensitive aesthetes* in Germany than expected, whereas in the USA it is the other way round. This might be due to the generally worse state of the road surface in the USA than in Germany, making the participants more aware of noisy sounds. Among the females there are very little *sound-oriented fun-drivers*. The drivers of cars from BMW and Porsche are more likely to be *sound-oriented fun-drivers* and less likely to be *noise-sensitive aesthetes*, and Porsche drivers are very unlikely to be *sound-uninterested pragmatists*. This seemed reasonable since both makes have an emphasis on sporty driving characteristics, which is why one can expect their drivers to have a dynamic driving style, which in turn is particularly pronounced for *sound-oriented fun-drivers*. In contrast, the majority of Volkswagen drivers are *sound-uninterested pragmatists*. In the segment of compact cars there are very few *sound-oriented fun-drivers*, but more *sound-uninterested pragmatists* and in the segment of sports cars, coupés and convertibles, it is the other way round. This seemed reasonable, as these segments represent "practical" cars or, respectively, cars with sporty driving characteristics. The segment of full-size luxury

[6] The Bonferroni correction was applied, i.e., the significance level was lowered to $\alpha = .01$ due to the five comparisons.

Table 2.5: Average age and percentages of driver types among different subsamples (rows add up to 100 %). Deviations from the percentages of the total sample exceeding 5 % are printed in **bold**.

	Type 1	Type 2	Type 3
Age Ø	53.5 yr	57.9 yr	58.9 yr
Total sample	30.9 %	33.5 %	35.7 %
Germany	31.0 %	**27.4 %**	**41.6 %**
USA	30.6 %	**45.7 %**	**23.7 %**
Female	**22.7 %**	37.0 %	40.3 %
Male	32.9 %	32.6 %	34.5 %
Audi	34.5 %	29.6 %	35.9 %
BMW	**39.5 %**	**26.7 %**	33.7 %
Mercedes	28.9 %	35.6 %	35.4 %
Porsche	**58.3 %**	**25.0 %**	**16.7 %**
Volkswagen	**24.6 %**	**24.6 %**	**50.7 %**
Compact car	**19.5 %**	29.8 %	**50.7 %**
Mid-size (luxury) car	29.4 %	33.7 %	36.9 %
Full-size luxury car	**24.2 %**	**45.9 %**	**29.9 %**
SUV	27.9 %	32.5 %	39.6 %
Sports car/ coupé/convertible	**46.9 %**	30.7 %	**22.4 %**

cars is dominated by *noise-sensitive aesthetes* and lacks *sound-oriented fun-drivers* and *sound-uninterested pragmatists*, which might be due to the low sound pressure levels inside those cars, making them especially interesting for noise-sensitive persons. All in all, the division into the three driver types allows for meaningful distinctions with respect to sociodemographic as well as car segment and make distributions.

2.3.6 Assessment of vehicle sounds

In this section, the influence of the driver type on the assessment of the sound of the own vehicle as well as the ideal follow-up vehicle is considered. Finally, the differences between the driver types will be illustrated using the scales *comfortable – sporty* and *unpleasant – pleasant*.

The results of the semantic differential were obtained using a seven-point ordinal scale. As for the data used in Sections 2.3.2 and 2.3.3, these results were tested for having interval-scale level using MSI. It was found that the mean absolute deviation between observed and predicted cumulative percentages lay between 0.37 % (*high quality – low quality*) and 2.40 % (*unpleasant – pleasant*). The correlation between mean values and scale values of the items was $r^2 = 99.7\,\%$, which is why the observed ratings of the semantic differential were further considered as having interval-scale level.

2.3.6.1 Own vehicle

In Part 2 of the questionnaire the participants were asked to provide a general description of their vehicle's driving sounds. For this task, the participants were provided with a semantic differential, using the adjective pairs 1 to 13 from Table 2.1 with a seven-point scale; pair 14, *familiar – innovative*, was not used as the vehicle sound of the own vehicle was assumed to be highly "familiar" to the participants.

To test whether there is an effect of the driver type on the vehicle sound assessment, a two-way MANOVA was performed on the results of these 13 items. Besides the driver type the car segment was also considered as an independent factor, because it strongly affects the emitted vehicle sounds themselves, e.g., the sounds of a sedan are usually quite different from the sounds of a sports car. Hence, an influence on the sound assessment appeared obvious. While both factors, i.e., the car segment ($\Lambda = .82$, $F(52, 5590.83) = 5.56$, $p < .05$, $\eta_{\mathrm{p}}^2 = .05$) as well as the driver type ($\Lambda = .93$, $F(26, 2886) = 4.33$, $p < .05$, $\eta_{\mathrm{p}}^2 = .04$), became significant with a small effect, their interaction did not ($\Lambda = .92$, $F(104, 9947.90) = 1.17$, $p > .05$). To be specific, the driver type has an effect on all pairs except for *bright – dark*, *howling – not howling*, *unpleasant – pleasant* and *don't like it at all – like it a lot* ($p > .05$). For the adjective pairs *loud – quiet*, *calming – exciting*, *aggressive – defensive*, *comfortable – sporty*, *restrained & elegant – striking & bold*, *beefy – mellow* and *high quality – low quality*, a small effect was indicated by η_{p}^2. However, for the remaining adjective pairs, i.e., *rough – smooth* and *untypical of* `make` *– typical of* `make`, η_{p}^2 was below .01. Post hoc tests showed that the *sound-oriented fun-drivers* rated their own cars as being *louder*, more *exciting*, more *aggressive*, *sportier*, more *striking & bolder* and *beefier* than the other two types. In addition, the *sound-uninterested pragmatists* rated their own cars significantly *lower in quality*, more *defensive* and more *untypical of* `make` than the other two types. The *noise-sensitive aesthetes* rated their own cars significantly *smoother* than the other two types.

At this point it was concluded that the driver type does have an effect on the participants' ratings of their own vehicle's sounds. The effect size is comparable to that of the car segment, and the driver type affects all but the evaluative scales *unpleasant – pleasant* and *don't like it at all – like it a lot*.

2.3.6.2 Ideal follow-up vehicle

Analog to the rating of the own vehicle, the participants were asked to describe the sound of the (hypothetical) ideal follow-up vehicle of their current vehicle. As it did not seem reasonable to ask for the evaluation of an ideal vehicle (which by definition was assumed to be evaluated best), the respective adjective pairs were not used in this part of the questionnaire. However, the adjective pair *familiar – innovative* was added to the semantic differential. Thus, the adjective pairs 1 to 8 and 14 from Table 2.1 were used.

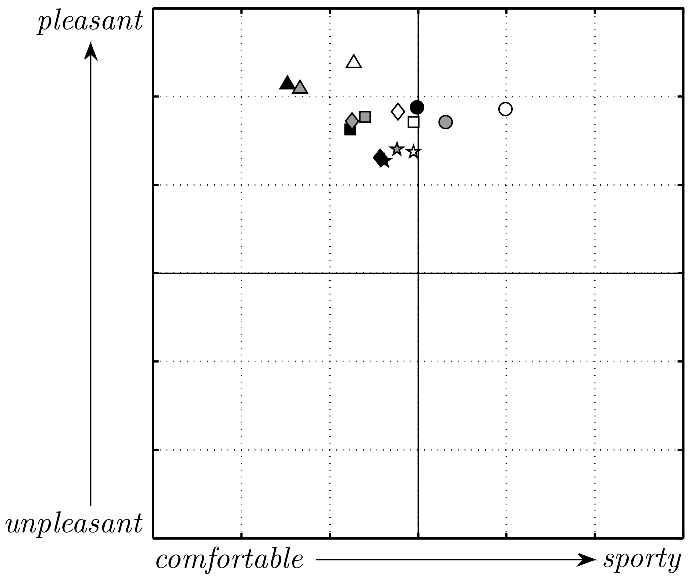
In a two-way MANOVA with these nine adjective pairs as independent variables, both the car segment ($\Lambda = .86$, $F(36, 5600.45) = 6.21$, $p < .05$, $\eta^2_\mathrm{p} = .04$) and the driver type ($\Lambda = .84$, $F(18, 2988) = 15.72$, $p < .05$, $\eta^2_\mathrm{p} = .09$) became significant, while no significant interaction between them was found ($\Lambda = .94$, $F(72, 9095.23) = 1.23$, $p > .05$). Hence, the car segment has a small and the driver type has a medium effect on the rating of the ideal vehicle sounds. More precisely, the driver type has a significant medium effect on all adjective pairs except for *bright – dark* and *familiar – innovative*, for which $p < .05$ but $\eta^2_\mathrm{p} \leq .01$. Post hoc tests showed that the *sound-oriented fun-drivers* wished the sound of their ideal follow-up vehicle to be *louder*, more *exciting*, more *aggressive*, *rougher*, *sportier*, more *striking & bolder* and *beefier* than the *noise-sensitive aesthetes* or the *sound-uninterested pragmatists*. The *sound-uninterested pragmatists* tended to the opposite side of these scales in comparison to the other two driver types (except for *rough – smooth*, where *noise-sensitive aesthetes* and *sound-uninterested pragmatists* did not differ).

To sum up, the driver type has an even larger effect on the preferred sound characteristics of the ideal vehicle than the car segment has. This influence affects all requested scales. It was concluded that particularly the wishes for the vehicle sounds, even more than the assessment of actual sounds, are influenced by the driver type.

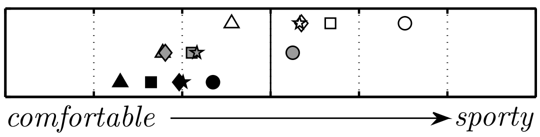
In order to illustrate these findings, the two scales *comfortable – sporty* and *unpleasant – pleasant* were chosen as examples. These scales are well in line with the findings of Bisping (1994), who stated that "acceptance (pleasant – unpleasant) and sportiness (sporty – not sporty)" are the "basic factors of the emotional evaluation space" of vehicle sounds[7]; however, the opposite poles for the adjective *sporty* differ. In Figure 2.5 the participants' mean ratings on these two scales are shown, distinguished between the three driver types and the five segments. As *unpleasant – pleasant* was not requested for the ideal follow-up vehicle, just the scale *comfortable – sporty* is depicted in Figure 2.5b. Interestingly, all ratings of the own vehicle lay in the half-plane corresponding to *pleasant*, which might have been due to the cars being mainly premium vehicles. Little differentiation was observed on this scale between the driver types. On the scale *comfortable – sporty*, a somewhat broader range was occupied by the different segments and driver types. As discussed earlier, the driver type has a small effect on the rating of the own vehicle in terms of *comfortable – sporty*, but no effect in terms of *unpleasant – pleasant*.

For the rating of the ideal follow-up vehicle (Figure 2.5b), the driver type has a medium effect in terms of *comfortable – sporty*. To be specific, the ratings of all three driver types

[7] See Bisping (1997) for an English reference. Note that Bisping uses the term "powerfulness (powerful – weak)" there instead of sportiness; however, Zeitler and Zeller state that "weak – powerful" also describes the factor sportiness (Zeitler and Zeller, 2006).



(a) Own vehicle.



(b) Ideal follow-up vehicle.

Figure 2.5: Average ratings of the vehicle sounds on the scales *comfortable – sporty* and *unpleasant – pleasant* for the own vehicle (a), and *comfortable – sporty* for the ideal vehicle (b), respectively. Segments: compact cars (☆), mid-size (luxury) cars (□), full-size luxury cars (△), SUVs (◊) and sports cars/coupés/convertibles (○); driver types: *sound-oriented fun-drivers* (white), *noise-sensitive aesthetes* (gray) and *sound-uninterested pragmatists* (black).

differed significantly, with the *sound-oriented fun-driver* wishing the sportiest and the *sound-uninterested pragmatist* wishing the most comfortable driving sound. This was due to the *sound-uninterested pragmatist* wishing the driving sounds of all segments to be *comfortable*, while the *sound-oriented fun-driver*'s and the *noise-sensitive aesthete*'s requirements covered a broader range, particularly separating the sports cars from the other segments. However, the interaction between the driver type and car segment did not become significant.

2.4 Discussion

As argued in the introduction, it is important to understand the influence of personality traits on sound quality evaluation when considering automotive applications such as sound design or the development of sound quality metrics. The driver typology presented in this article is an approach that offers a reasonable solution to this challenge. Several aspects of opinion and personality are mapped to a single nominal scale, which assigns drivers to the three categories: *sound-oriented fun-drivers*, *noise-sensitive aesthetes* and *sound-uninterested pragmatists*. These three driver types represent prototypical real-world configurations of the considered personality traits, and therefore it is to be expected that these can account for assessments made by most individual drivers.

Along with the car segment, the driver type has shown to be a significant source of variation in the rating of and wishes for the vehicle sounds. When considering the rating of the participants' own vehicles, however, the driver type has no effect on the evaluative scales *unpleasant – pleasant* and *don't like it at all – like it a lot*. Furthermore, the corresponding ratings are biased towards *pleasant* and *like it a lot*, respectively. This might be a form of justification bias, i.e., the participants give better ratings because they *want* to be satisfied with their own vehicle. In contrast, subjects in Notbohm et al. (2013) tended to judge negatively, on average, pass-by sounds of vehicles unrelated to them, and these ratings were correlated to a reported feeling of unpleasantness. Furthermore, a higher noise sensitivity reinforced these responses, while in the current study no influence of the driver types (who incorporate noise sensitivity) on pleasantness was found. As another example, Altinsoy (2015) shows in a listening experiment that being either a "sports car enthusiast" or a "luxury car enthusiast" affects the rating of pleasantness, but not the rating of sportiness for exterior vehicle engine start sounds: sportiness is rated consistently across all participants, but the sports car enthusiasts rate sporty sounds as being pleasant, while the luxury car enthusiasts rate these as being unpleasant. Hence, it seems advisable to collect data from subjects that are not emotionally involved, e.g., in listening experiments with "unrelated" vehicle sounds. Also the wishes for the ideal follow-up vehicle are affected by the driver type, indicated by a medium effect on almost all requested scales. A significant interaction between driver type and car segment was not found, but some dependencies could be supposed from the mean ratings, i.e., that both *noise-sensitive aesthetes* and *sound-oriented fun-drivers* request more differentiation between the sounds of different car segments than the *sound-uninterested pragmatists* do.

These findings have implications for the optimization of vehicle sounds, for which it seems useful to consider these different driver types separately. Nowadays cars already offer user-selectable "driving programs" to adjust a car's handling and road performance characteristics to the driver's individual requirements. In the course of individualization, these driving programs could be combined with certain sound characteristics of the car, each suitable for a particular driver type. Also metrics for vehicle sound quality should be adjustable to evaluate sound quality "through the ears of" a specific driver type to ensure ecological validity of the results.

As the driver typology developed in this exploratory study is based on a large but specific sample, it will be interesting to see to which extent a different sample will affect the influence of the driver types presented here.

Acknowledgement

Special thanks go to Ewald Strasser from the University of Oldenburg for his support in terms of discussion, to Dr. Martin Tischler from Daimler AG for his assistance in designing the questionnaire and interpreting the results, and Dr. Helga Sukowski, formerly the University of Oldenburg, for her support in designing the questionnaire.

3 Contextual aspects in subjective vehicle sound assessment[1]

In vehicle acoustics, the quantities comfort, sportiness, type conformity and make conformity are facets of overall sound quality. However, little attention is usually paid to contextual influences such as the vehicle's type or make, the driving condition one is engaged in or the driver type itself. These contextual aspects are expected to have a significant influence on sound quality assessment. In this study, the results of an exploratory field study are used to determine the interactions between the aforementioned acoustic quantities on the one hand and contextual aspects on the other hand when assessing sound quality. Based on field test results of 85 subjects that participated in six car sound conditions, a regression model was determined that predicts overall sound quality from comfort, sportiness and type conformity. The predictive power of these three quantities varies depending on the considered car and driver type. Further analyses showed that for comfort and sportiness, much predictive power was obtained from specific driving conditions. Comfort is determined mainly in "static" driving conditions such as driving with constant speed, whereas sportiness manifests in a broader range of situations, with an emphasis on dynamic driving. Type conformity was observed as being closely related to comfort for non-sporty cars and to sportiness for sporty cars, respectively. These effects show that besides acoustic factors, context effects influence how acoustic properties are evaluated by subjects.

[1]This chapter is based on Maiberger et al. (2018b), a manuscript currently under revision:
D. Maiberger, E. Strasser, U. Letens, S. van de Par: Contextual aspects in subjective vehicle sound assessment. Acta Acustica united with ACUSTICA, 2018, currently under revision.

3.1 Introduction

Blauert and Jekosch (1997) distinguish three levels of acoustics for product sounds, which were transferred into the domain of vehicle sound by Letens (2010):

Physical acoustics, describing "sound events" in terms of, e.g., level, spectral or temporal properties of the sound wave.

Psychoacoustics, describing "auditory events" in terms of, e.g., loudness, sharpness or roughness. These are usually determined in listening experiments, where subjects are asked to make "unbiased" judgements, disregarding any further contextual aspects or informational content of the sounds (Blauert and Jekosch, 1997).

Psychological acoustics, describing the evaluation of sounds by subjects in terms of, e.g., pleasantness, annoyance or sound quality. These are subject to contextual aspects, e.g., intersensory effects, situation and expectation (Letens, 2010), and involve psychological processes such as information reduction and the use of different frames of reference (Blauert and Jekosch, 1997).

Due to the context dependency of sound quality (which falls in the category of "psychological acoustics"), it cannot be defined universally, but rather with respect to a particular context, e.g., source, situation and listener (Blauert and Jekosch, 1997, Letens, 2010, Jekosch, 1999, Västfjäll, 2004, Fiebig, 2015). In the domain of vehicle sounds, the source can possibly be characterized by the *vehicle type* (i.e., segment) and *make*, the situation can possibly be characterized by the *driving condition*, and the listener can possibly be characterized by the *driver type*, which in combination might imply a certain expectation with regard to the corresponding sounds. Studies concerning vehicle sounds, however, rarely aim at giving an integrated view of the influence of contextual aspects on sound quality, yet acknowledging its existence. In a previous study, Maiberger et al. (2018a) indeed found opinion and personality of the driver, aggregated to a discrete *driver typology*, to be a significant predictor of vehicle sound assessment (see Chapter 2). Particularly the wishes regarding vehicle sounds differ considerably between the three identified driver types, labelled as *sound-oriented fun-driver*, *noise-sensitive aesthete* and *sound-uninterested pragmatist*.

In general, it seems to be agreed that vehicle sound quality depends on the perceptual quantities comfort, sportiness, type conformity and make conformity. For example, Letens (2010) proposed the distinction between three factors for an integrated description of the overall "sound performance" of a car: 1) *perceived value* ("Wertanmutung"), 2) *emotionality* and 3) *conformity* with respect to market positioning of the car as well as the make. He describes *perceived value* as freedom from disturbing noise; *emotionality* is associated with sporty, dynamic driving characteristics, and an "aesthetic" sound; *conformity* is described as the extent to which the sound matches the car, and the extent to which a brand sound ("corporate sound") is realized, respectively. It can thus be concluded that sound quality considerations must include vehicle type and vehicle make in order to be able to determine conformity.

Köber et al. (2016) state that "customers' requirements with regard to the sound impression of the exhaust system are mainly dependent on the market positioning of the vehicle", and thus influenced by the non-acoustic factor of market positioning. According to them, those requirements manifest themselves as the contrasting ideals of "comfort" and "sportiness", which implies that both quantities are not necessarily realizable at the same time.

Bisping (1997) states that the perceptual factors "pleasantness" and "powerfulness" account for more than 60 % of the variance of car interior sound quality in standard driving conditions. Based on prior research he derives a map with the corresponding axes *weak–powerful* and *unpleasant–pleasant*, and shows based on the results of relative scaling experiments that cars of different types (i.e., small, middle, luxury, sport, truck) occupy different quadrants of the map. These locations can serve "as templates for target sound design or respectively as guidelines for optimization". In this way, the vehicle type is implicitly introduced as a context variable; moreover, conformity can be assumed to depend on pleasantness and powerfulness. In the diagram, Bisping finds "a general trend indicating that the pleasantness of the sounds decreases while their powerfulness increases", due to the "trade-off between the technical realization of both pleasantness and powerfulness", which again indicates a certain dependency between those two quantities.

Zeitler and Zeller (2006) define vehicle sound quality as "the extent to which the sound character matches the overall vehicle character and brand values", bringing both vehicle type and vehicle make into play. They depict the requirements for the sound character of different vehicle types in a coordinate system, with axes composed of the *driving noise at constant speed* on the one axis and the *engine level difference between cruise and acceleration* as the other axis. These axes were found to be related to the perceptual dimensions of "comfort/loudness" and "sportiness", respectively[2]. In the diagram no vehicle type is found in the region that corresponds to high values for "comfort/loudness" *and* "sportiness", which seems to exclude the idea of both quantities being completely independent. Furthermore, as the vehicle types occupy different regions with respect to comfort/loudness and sportiness within the diagram, the earlier assumed dependency of conformity on comfort and sportiness is reinforced.

In summary, vehicle sound quality encompasses the following quantities:

1. *comfort*,
2. *sportiness*,
3. *type conformity*, i.e., the match between comfort, sportiness and the vehicle type, and

[2]These were determined in factor analysis, together with "harshness" and "timbre", and "are assumed to reflect the perceptual space underlying the [sound] evaluations" (Zeitler and Zeller, 2006).

4. *make conformity*, i.e., the match between comfort, sportiness and the vehicle make.

It is not clear how exactly these four quantities act and interact to form the overall sound quality; however, comfort and sportiness must not be assumed to be perfectly orthogonal (Köber et al., 2016, Bisping, 1997, Zeitler and Zeller, 2006), and also a certain dependency of the conformities on comfort and sportiness is indicated (Bisping, 1997, Zeitler and Zeller, 2006). However, for sound quality evaluation also contextual aspects must be taken into account (Blauert and Jekosch, 1997, Letens, 2010). Type conformity and make conformity already incorporate the context as they are intrinsically linked to the non-acoustic factors vehicle type and vehicle make, respectively, but also the driving condition and the driver type should be considered for a valid understanding of vehicle sound assessment.

The aim of this study is to find out to what extent the evaluation of vehicle sound quality depends on non-acoustic factors, or more specifically the context variables vehicle type and make, driving condition and driver type. To this end, subjects were placed in a real driving context, i.e., they drove a car themselves in normal road traffic on a predetermined route. In the course of this study, the subjects were asked to rate the acoustic quantities comfort, sportiness, type conformity, make conformity and sound quality. In addition, personality traits were assessed via questionnaires. The results are used to explore the nature of vehicle sound assessment by subjects, by means of statistical methods. The ultimate goal is to construct a linear model which shows the dependency of overall sound quality on the four quantities and especially the connection to the contextual aspects. Subsequently, also submodels for the main predictors of sound quality are constructed from the data, to get deeper insight into their relation to one another as well as their context-dependency. The leading questions in this process are whether the quantities found in literature, i.e. comfort, sportiness, type conformity and make conformity, indeed do influence sound quality, and if so, whether they exert an effect “on their own” (as a *main effect*) or just in combination with other quantities (as an *interaction*), especially with the abovementioned contextual aspects. If such interactions with contextual aspects are found, this would emphasize and specify the context-sensitivity of the vehicle sound assessment process.

3.2 Data acquisition

In this section, the acquisition of the subjective, contextualized data will be described.

3.2.1 Procedure

The study was designed as a field test because this allows for contextualized data acquisition, offering a high ecological validity. For the exploratory approach followed in this study, this design seemed adequate. The field test was conducted at two locations: Sindelfingen, Germany (*Location A*), with 3 cars of the same make but different types, namely a compact car, a mid-size car (sedan) and a roadster; and Oldenburg, Germany (*Location B*), with a single mid-size car (sedan) of the same make as in Location A, but three different engine sound conditions, namely "original", "booming/howling" and "rough/sporty". The sound in Location B was modified via electric sound synthesis and reproduction over the car's audio system (cf. Section 3.2.3), aiming at explicitly altering the type conformity of the car's sound.

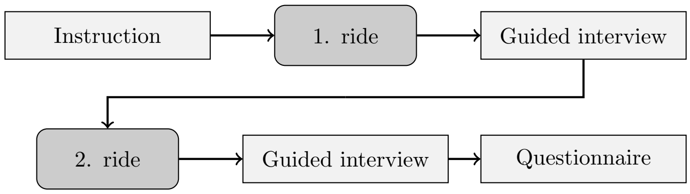


Figure 3.1: Schematic of the experimental procedure.

The procedure, which was inspired by Sellerbeck (2014) and Schulte-Fortkamp et al. (2014), was the same in both locations, and is depicted in Figure 3.1. The subjects were instructed to assess the driving sound of the respective car in a test ride by driving themselves. The term "driving sound" was introduced to the subjects as the sounds coming from the engine, the tires, the wind and the chassis. For the time of the trip, the subjects were asked to "think aloud", i.e., to tell the investigator everything that comes to their mind concerning the driving sound during the assessment process, e.g., giving a description of the sound, or naming positive, neutral as well as negative aspects of the sound. This was intended for focusing the participants on the topic matter of the study, while not priming them with regard to the aspects comfort, sportiness, type conformity and make conformity. The subjects then drove on a predetermined route, which contained city traffic, country road and highway. The investigator, posed as the co-driver, guided along the route, took notes about the subjects' comments and reminded them to think aloud if necessary.

After the first half of the route, which usually took around half an hour, there was a pause on a parking area for a guided interview. The interview started with a question for the overall sound quality ("Which rating do you give for the driving sound of this car?", using Scale (a) from Figure 3.2) as well as aspects of the driving sound that were most striking to the subjects (as an open question). After that, the focus was set on the (acoustic) aspects comfort, sportiness, type conformity and make conformity of the driving sound, and the subjects had to rate the extent to which each of these quantities was met ("How comfortable/sporty is the overall driving sound of this car to you?", "How well does the driving sound fit the vehicle type/make?", using Scale (b)). Furthermore, the subjects rated the importance of comfort and sportiness, respectively, for the respective vehicle type as well as for the vehicle make ("How important is it to you for the driving sound of this vehicle type/make to be comfortable/sporty?", using Scale (c)). After answering

(a)	very bad	1	2	3	4	5	6	7	8	9	10	very good
(b)	not at all	1	2	3	4	5	6	7	8	9	10	very much
(c)	unimportant	1	2	3	4	5						important

(d)	not at all	slightly	moderately	quite	very much
(e)	does not apply at all	does rather not apply	neither nor	rather applies	totally applies

Figure 3.2: Rating scales used in the study.

an open question for important driving conditions ("In general, when you do a test drive, in which driving conditions do you attach importance to the driving sound?"), the subjects were asked to indicate the importance of nine preselected driving conditions (see Section 3.2.4) for the overall impression of comfort and sportiness ("How important is this situation to you for the judgement of how comfortable/sporty the driving sound is?", using Scale (c)).

The second half of the route was identical to the first half. The subjects were again asked to think aloud when driving, and to give a rating of the comfort and sportiness, respectively, for the nine preselected driving conditions once they occurred ("How comfortable/sporty do you find the driving sound?", using Scale (b)); in Location B, these ratings were acquired directly after the ride.

After the ride, some concluding questions were asked concerning the importance of comfort, sportiness, type conformity and make conformity for the overall sound quality of the driving sound ("How important for your overall rating was how comfortable/sporty the driving sound is?", "How important for your overall rating was whether the driving sound fits the vehicle type/make?", using Scale (c)). After that the participants completed a combined questionnaire, with items to determine their *driver type* (Maiberger et al., 2018a), their noise sensitivity (by means of the *questionnaire on the individual attitude towards the acoustical environment,* Notbohm, 2010) and the *Big Five dimensions of personality* (Rammstedt et al., 2013). For the items concerning the driver type and noise sensitivity, Scale (d) was used, while for items concerning the Big Five, Scale (e) was used, following Maiberger et al. (2018a), Notbohm (2010) and Rammstedt et al. (2013). For determining the driver type, the six scales "dynamic driving style & acoustical orientation", "noise sensitivity", "sound awareness", "status aspect", "reservations about cars" and "demand for comfort" were calculated according to Maiberger et al. (2018a); from the resulting profile each participant was assigned to one of the driver types *sound-oriented fun-driver*, *noise-sensitive aesthete* or *sound-uninterested pragmatist*, according to their prototypical profiles in Maiberger et al. (2018a). In contrast to *sound-uninterested pragmatists*, who consider a car as "just a tool", *sound-oriented fun-drivers* and *noise-sensitive aesthetes* do care about vehicle sounds: the former because of their dynamic driving style, the latter because of their noise sensitivity. For the analyses, the four scales of noise sensitivity and the five scales of the Big Five were further calculated according to Notbohm (2010) and Rammstedt et al. (2013), respectively. Finally, the subjects did a computer-based *go/no-go association task* (GNAT) (Nosek and Banaji, 2001) that dealt with the implicit attitude towards status symbols, where "implicit" means that the attitude is measured without directly asking participants. This technique is especially useful when potentially socially sensitive domains should be assessed, as in these cases self-report measures are often distorted by social desirability (for an overview over implicit measures, see Reis and Judd, 2014). The resulting GNAT score is derived from the participants' response time and assignment of selected words to categories, and indicates the implicit preference for status symbols.

The conversation during the whole ride was recorded with a microphone inside the car cabin, as well as the engine speed, driving velocity, acceleration and accelerator pedal position. However, these data are not the focus of this study but might be used for future analyses.

3.2.2 Sample

In Location A, 61 participants between 22 and 69 years (Ø = 49, $\sigma = 14$) took part, 21 of which were female. 20 participants were invited for the compact car, 21 for the mid-size car and 20 for the roadster. The acquisition was done via a panel consisting of vehicle drivers, who are willing to take part in various car-related studies. The panel members drive vehicles of premium makes, i.e., makes with a high perceived quality and value, and the participants were required to be experienced with a car of the corresponding segment.

In Location B, 24 university members between 20 and 54 years (Ø = 37, $\sigma = 9$) took part, 10 of which were female. As opposed to Location A, every participant did a ride in each of the three sound conditions, but with a separation interval of one to two weeks between successive rides. The participants were not told about the sound modification, but were instructed to pay as much attention with respect to the driving sound as in the ride(s) before, as something *could* have changed. The order of the three sound conditions was randomized for all participants.

No statistical evidence was found to support a learning effect of the participants in Location B, which is why the respective results were further considered as (the more conservative type of) between-subject data. Furthermore, no effect of location (A and B) on the dependent variables was found, which is why all data were treated as one sample.

As stated before, the driver type (Maiberger et al., 2018a) of every participant was determined. The distribution of participants over driver types is shown in Table 3.1. It can be seen that there is a lack of *sound-oriented fun-drivers* in Location B as compared to the other two driver types. This might originate from the different backgrounds of the subsamples in the two locations, i.e., premium vehicle drivers vs. university members. It seems that university members are less likely to be *sound-oriented fun-drivers*, which means that they are less interested in dynamic driving and more critical against cars and noise than the premium vehicle drivers.

Table 3.1: Distribution of participants over the three driver types *sound-oriented fun-driver* (SF), *noise-sensitive aesthete* (NA) and *sound-uninterested pragmatist* (SP).

	SF	NA	SP	Total
Total	25	31	29	85
Location A	23	20	18	61
Location B	2	11	11	24

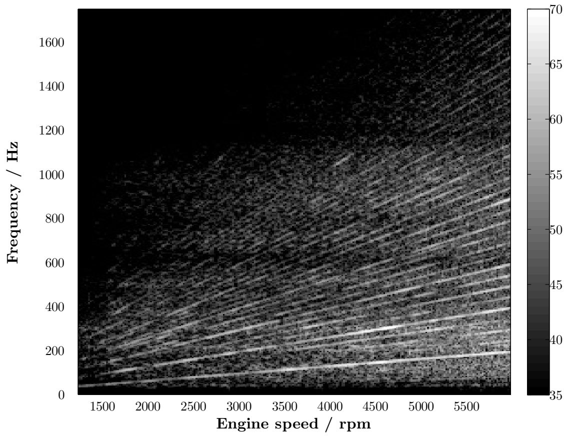
3.2.3 Sound conditions

In Location B, the same car was used in three different sound conditions. Along with the *original condition*, i.e., original sound of the car, two modifications were made by generating sinusoidal components corresponding to certain engine orders (EO) over the car's audio system, with levels depending on engine speed and load. The CAMPBELL *diagrams*, i.e., the depictions of the frequency distribution with respect to engine speed, of the three sound conditions are shown in Figure 3.3, as measured with wide-open-throttle (WOT) in third gear. For the *booming/howling condition*, the 2. EO is emphasized up to around 3500 rpm (for the booming part), and the 16. EO is emphasized between 2500 and 4500 rpm (for the howling part) (see Figure 3.3b). For the *rough/sporty condition*, the EOs 1.5 to 8 (in steps of 0.5 EO) are emphasized, which gives the sound a rough and thus sporty character (see Figure 3.3c).

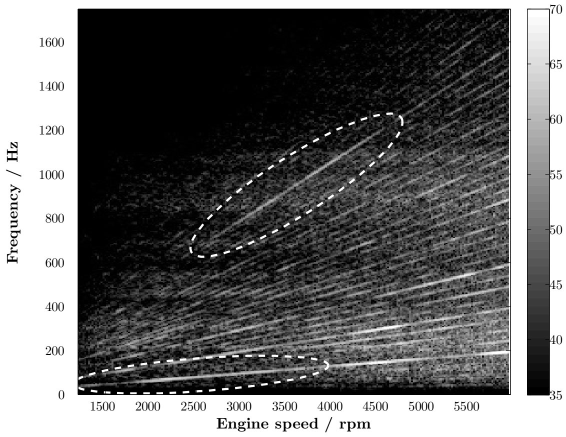
3.2.4 Selection of driving conditions

In the course of the experiment, the subjects were provided with a printed list of nine driving conditions and were asked to give ratings regarding various aspects. The selection of driving conditions on this list was composed prior to the experiment, based on the driving conditions that are frequently considered in studies dealing with the process of vehicle sound assessment (such as Bisping, 1997, Letens, 2000, Schulte-Fortkamp and Genuit, 2005, Altinsoy et al., 2010, Jung and Grützmacher, 2011, Sukowski et al., 2013, Busse, 2014), with the optimization of the subjectively rated sound character of a car (such as Ishii et al., 2001, Fuhrmann and Garcia, 2001) or with the development of metrics for vehicle sound quality (such as Zeitler and Zeller, 2006, Fiebig and Kamp, 2015, Alenius et al., 2016). The list gained from these studies was further adjusted and complemented, and finally consisted of nine driving conditions:

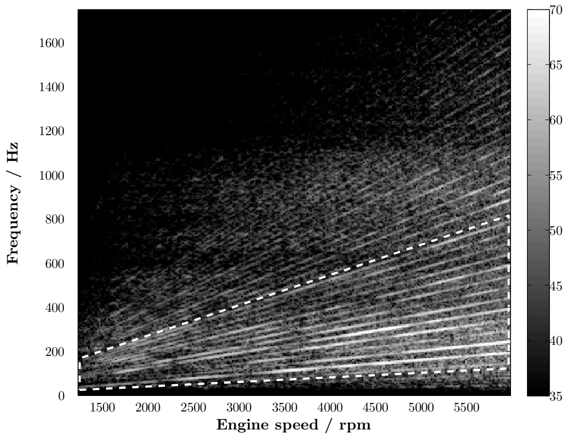
1. engine start;
2. engine idle;
3. driving off from standstill;
4. overtaking;
5. accelerating;



(a) Original condition.



(b) Booming/howling condition.



(c) Rough/sporty condition.

Figure 3.3: CAMPBELL diagrams of the three sound conditions used in Location B, measured with WOT in third gear. The modified regions are marked with white dashed lines.

6. coast down;
7. constant speed of 50 km/h (inner-city);
8. constant speed of 80 km/h (rural road);
9. constant speed of 140 km/h (highway).

3.3 Analysis of driving conditions

In contrast to the driver type and the car, the context variable “driving condition” is continuously changing during the ride, and could thus be called the “short-term context”. Due to this property, also the differentiation between the driving conditions is more intricate. In the course of the experiment several questions were asked concerning comfort, sportiness and importance of the nine preselected driving conditions. In this section the corresponding findings will be grouped together to make them handable, for the linear model that will be fitted in Section 3.4.

It is expected that not all driving conditions contribute equally to the sound assessment. Some might occur more frequently than others, and some might have a greater effect than others on certain aspects of the vehicle sound. In order to separate the more important driving conditions from the less important, several approaches were chosen: an open question directly asking for the most important driving conditions, a rating of the importance of nine preselected driving conditions, and a factor analysis of the ratings of comfort and sportiness of the nine preselected driving conditions.

3.3.1 Open question

In an open question, the participants were asked directly for the driving conditions to which they attach importance with regard to the driving sound during a test ride. The answers were sorted into nine categories corresponding to the driving conditions introduced in Section 3.2.4. The relative frequencies are depicted in Figure 3.4. It is observed that “constant speed (highway)” is named most often (43 % of the participants), followed by “accelerating” (22 %) and “constant speed (inner-city)” (15 %).

3.3.2 Rated importance

During the guided interview, a table was handed to the participants that contained the nine abovementioned driving conditions. The task of the participants was to indicate the importance of each driving condition for the rating of the overall acoustic comfort on the one hand, and for the rating of the overall acoustic sportiness on the other hand. For this purpose, a five-point scale ranging from “unimportant” to “important” was given to the participants. The means and

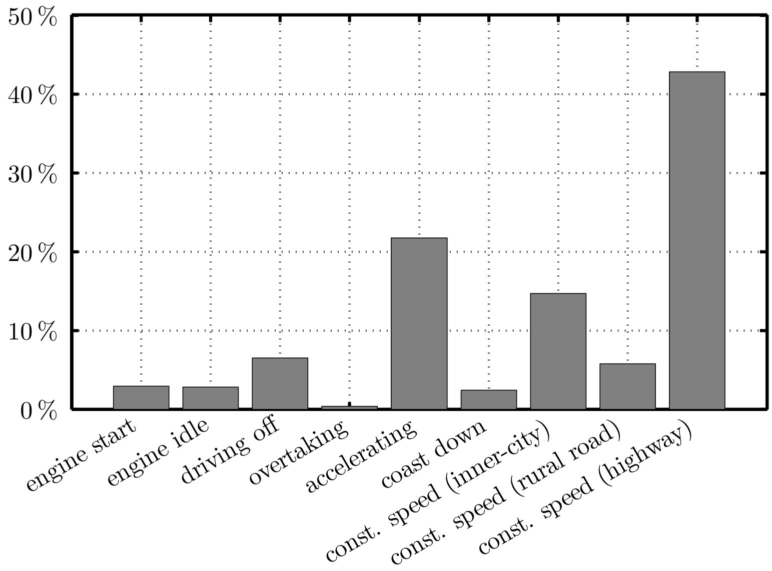


Figure 3.4: Relative frequencies of driving conditions named by the participants as being important for sound assessment.

95 %-confidence intervals (CIs) across participants are depicted in Figure 3.5. For the overall acoustic comfort, driving with constant speed is rated “important”, especially for high velocities. This is well in line with the answers to the open question presented in Section 3.3.1. In post hoc tests (with Bonferroni correction, Field, 2009), the three constant driving conditions differ significantly from the rest of the situations. In contrast, for overall acoustic sportiness the dynamic driving conditions accelerating and overtaking are rated most important. In post hoc tests, these driving conditions show to be significantly more important than the other driving conditions; furthermore, accelerating is rated significantly more important than overtaking. The importance of accelerating matches the results from the open question (Section 3.3.1); however, the high importance of overtaking was not observed. A possible explanation is that overtaking, as a combination of driving with constant speed and accelerating, is not regarded as a separate driving condition by the participants unless they are specifically asked to rate this maneuver.

3.3.3 Factor analysis of ratings

After the guided interview the participants were asked to rate the actual extent to which comfort and sportiness were met in each of the driving conditions, using a ten-point scale ranging from “not at all” to “very much”. These ratings of comfort and sportiness in the driving conditions seem to be important for the modeling in Section 3.4, because in contrast to the overall ratings they allow for differentiated consideration of the driving conditions. However, this means a lot of additional variables, which is why the driving conditions are grouped together beforehand

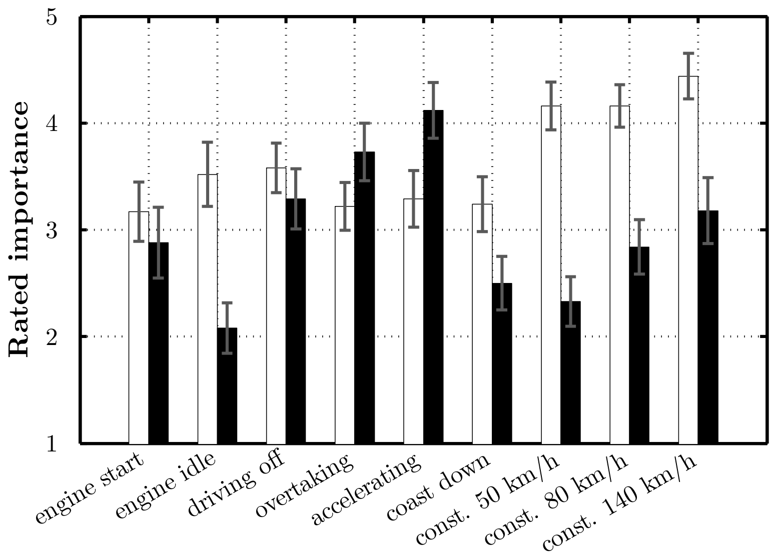


Figure 3.5: Means and 95 %-CIs of the rated importance of the driving conditions for the overall acoustic comfort (white) and the overall acoustic sportiness (black), respectively.

depending on the correlation between their ratings, by an *exploratory factor analysis* (EFA) (see Field, 2009). As each resulting factor defines an independent construct, a *scale value* will be calculated for each factor by averaging the ratings of those driving conditions that have high factor loadings on the corresponding factor. These scale values will be further used in the model calculations.

EFA of the comfort ratings for all driving conditions revealed two factors. Although the *elbow criterion* from the *scree plot* was somewhat ambiguous, the Kaiser-Guttmann *criterion* pointed at this solution. Further, the rotated factor solution showed nice and clear-cut loadings for the majority of the driving conditions on the two factors (see Table 3.2, column "comfort"). More importantly, the factors could intuitively be named as 1) *dynamic comfort* and 2) *static comfort*. At this point dynamic comfort consisted of driving off, overtaking and acceleration, and static comfort consisted of constant speed of 50 km/h, 80 km/h and 140 km/h. Engine start, engine idle and coast down showed ambiguous loadings and were therefore excluded. The reliability was further analyzed to see whether any of the driving conditions had to be excluded to guarantee that the yielded comfort scales are reliable. It showed that driving off (for dynamic comfort) and constant speed of 140 km/h (for static comfort) lead to considerably lower reliability, and thus were excluded from the respective scales.

EFA of the sportiness ratings for all driving conditions revealed only one factor (see Table 3.2, column "sportiness"). Both the elbow criterion from the scree plot as well as the Kaiser-Guttmann criterion pointed in this direction. The reliability analysis revealed no real outliers and therefore all driving conditions were included in the new scale.

To sum up, overall comfort splits into two parts, namely the static comfort and the dynamic

Table 3.2: Factor loadings of the two factor analyses: column "comfort" shows the rotated component matrix for the comfort ratings, column "sportiness" shows the component matrix for the sportiness ratings. Values below .4 are printed in gray; bold values mark the driving conditions that were used for calculating the scale values for each factor.

	Comfort		Sportiness
Factor no.	1	2	1
engine start	.664	.413	**.882**
engine idle	.538	.632	**.841**
driving off	.860	.303	**.881**
overtaking	**.835**	.299	**.824**
accelerating	**.904**	.153	**.764**
coast down	.520	.486	**.894**
const. 50 km/h	.216	**.935**	**.914**
const. 80 km/h	.225	**.922**	**.910**
const. 140 km/h	.386	.775	**.930**

comfort, which means that two distinct kinds of comfort are perceived by the participants. Overall sportiness, however, manifests in the ratings of all driving conditions in equal measure, which is contrary to the higher rated importance of the dynamic driving conditions for sportiness as observed in Section 3.3.2. There might be a difference between the importance rated by the participants and the "latent" importance as revealed by EFA.

For the further model calculations in Section 3.4 the following variables, reflecting the factors, were calculated: **dynamic comfort** as the average of comfort ratings during overtaking and accelerating, **static comfort** as the average of comfort ratings during constant speed of 50 km/h as well as 80 km/h, and **mean sportiness** as the average of sportiness ratings in *all* driving conditions.

3.4 Model for vehicle sound assessment

In this section, a regression model for vehicle sound assessment will be presented. Since it was an exploratory study, the idea was to make as little prior assumptions about dependencies as possible, in order not to mask any relations present in the data. Only in cases of statistical ambiguity theoretical considerations were used as decisive criterions. The modeling procedure started with the most general rating, the overall sound quality, and assumed it to be a dependent variable. All the remaining data were in principle considered as predictor variables for the rated sound quality, based on a linear regression model. The aim was then to find an "economic" model, i.e., a model that uses as little variables as possible while making the best possible prediction of sound quality. Such a model is assumed to give hints about the participants' vehicle sound assessment process.

The same modeling procedure was then undertaken in order to find regression models to predict each of the main predictors of sound quality (which will be called *submodels*); in this way, a deeper insight into the vehicle sound assessment process is possible.

3.4.1 Modeling procedure

A *multiple linear regression model* was taken as the basis, which aims at predicting sound quality (SQ) as a weighted sum of the predictor variables P_i:

$$\begin{aligned} SQ &= w_0 + w_1 \cdot P_1 + w_2 \cdot P_2 + \ldots + w_n \cdot P_n + \epsilon \\ &= w_0 + \sum_{i=1}^{n} w_i \cdot P_i + \epsilon, \end{aligned} \tag{3.1}$$

each weighted with a regression coefficient w_i. The constant term w_0 is also called the *intercept*. The error term ϵ accounts for the variance in SQ that is not explained by any of the predictors P_i. In addition, two-way *interaction* terms of the form $P_i \cdot P_j$ can be introduced that account for effects that depend on the simultaneous outcome of two variables:

$$SQ = w_0 + \sum_{i=1}^{n} w_i \cdot P_i + \sum_{i \neq j} w_{ij} \cdot P_i \cdot P_j + \epsilon. \tag{3.2}$$

From their mathematical formulation as products it becomes evident that interaction terms contribute all the more to SQ as the (absolute) values of P_i *and* P_j get higher. In a similar manner, higher order interactions (three-way, four-way etc.) are possible. As opposed to interaction terms, the "standard" terms $w_i \cdot P_i$ are called *main effects*.

As it does not seem reasonable and insightful to just use all the collected variables to explain sound quality, a procedure was first applied to identify the most important predictors as will be explained below. With these predictor variables several models complying with Equation (3.2) were built successively and compared with regard to their predictive power. The final model providing the best fit will be described in Sections 3.4.2 to 3.4.5. Prior to that, the modeling procedure will be described in more detail.

Potential predictor variables The following variables were considered as potential predictor variables for the rated sound quality: rated comfort; rated sportiness; rated type conformity; rated make conformity; the rated importances of comfort, of sportiness, of type conformity and of make conformity, respectively, for sound quality; the rated importances of comfort and of sportiness, respectively, for type conformity and for make conformity, respectively; dynamic comfort; static comfort; mean sportiness; the driven car; the location (A or B); gender; age; the driver type (Maiberger et al., 2018a); the four scales of noise sensitivity (Notbohm, 2010), i.e., critical attitude towards (environmental) noise, activation by music, noise sensitivity and disturbance by noise, appreciation of silence and natural sounds; the Big Five dimensions of personality (Rammstedt et al., 2013), i.e., extraversion, neuroticism, openness to experience, conscientiousness, agreeableness; the GNAT score.

Reduction of predictors The basic idea for reducing the number of predictors is to compare the predictive power of each variable P_i to a certain reference. This reference is chosen as the coarsest model, the "intercept only"-model, that predicts sound quality just as the collective average $\overline{SQ}$ of sound quality ratings. For the comparison, one predictor variable P_i is added to the model, and the predictive power of the resulting model is compared to the "intercept only"-model in terms of their *-2 log-likelihood*[3], which is a measure of the goodness of fit of a model (Field, 2009). If the model including P_i has a significantly lower -2 log-likelihood, it is regarded as a better model and the corresponding variable P_i is kept for further analyses; otherwise P_i is excluded. This is repeated with every potential predictor P_i, comparing its predictive power to the "intercept only"-model. As a result, many variables can be excluded from the sound quality model and its submodels specifically, keeping only the valuable predictors. The corresponding results will be presented in Sections 3.4.2 to 3.4.5.

Relationship between predictors All predictor variables P_i that have not been excluded yet are used as main effects in a model corresponding to Equation (3.1), i.e., disregarding any possible interactions. The result indicates whether each variable's contribution in explaining SQ is still significant in the presence of the other (maybe better) predictors. For this purpose variables are compared by their *probability value p*, which represents the probability for the corresponding regression coefficient w_i to actually be zero, given its observed value. Hence, variables with high *p*-values are not significant and can in principal be excluded (Twisk, 2006, p. 82). Although such an insignificant main effect is a hint that a variable is not useful for the model, the possibility remains that the variable is involved in an interaction that is of predictive value in the model (corresponding to Equation (3.2), or even higher order interactions). Hence, these kinds of models were tested thereafter. Since it would be very laborious to test all possible interactions and because such an approach would also cause a lot of results becoming significant by chance, not all interactions were tested systematically. Instead, the main focus was set on those variables that do have a main effect, and nonsignificant variables were kept as interaction partners when they made sense from a theoretical standpoint. In order to get an economic but comprehensive model, variables were excluded if they were only considered to be meaningful as a main effect but did not get significant. When there were pairs of redundant variables (in the sense that they only got significant if one or the other was considered in the model), the

[3]To be specific, the *Mixed Linear Model* (MLM) in SPSS (IBM, 2013) was used to analyze the data at this point.

Table 3.3: Contributions to the best-fit model for the rated sound quality; $R^2 = .825$, $R^2_{\text{adj}} = .742$.

Predictor variable	p-value	η^2_{partial}
Intercept	.000	.187
car $*$ driver type $*$ comfort	.001	.382
type conformity	.000	.219
car $*$ driver type $*$ sportiness	.001	.397

one with the lower predictive power was also excluded. The predictive power of a variable or interaction was assessed by η^2_{partial}, which is a measure for the variance of SQ explained by this variable.

In this way a variety of models was built, and these models were compared by their *adjusted* R^2 (denoted by R^2_{adj}) (Gordon, 2015, p. 376), which, as compared to the unadjusted *coefficient of determination* R^2, "penalizes" a model for a higher number of predictor variables. As this step-by-step approach produces a lot of data which would go beyond the scope of this chapter, only the resulting model with the highest R^2_{adj} will be presented, denoted as the *best-fit* model.

3.4.2 Model for sound quality

The initial reduction of predictor variables revealed that the following variables can potentially predict the rated overall sound quality: comfort; type conformity; make conformity; driven car; openness (Big Five); static comfort; sportiness; driver type.

In the next step, these variables were used as pure main effects. It was observed that make conformity, the driven car, openness, static comfort, sportiness and the driver type are not significant. Since the driver type and the driven car, as categorical variables, were very interesting for potential interaction, they were kept in the analysis. Sportiness was regarded to be underrepresented in the study because only one car could be considered a sports car and therefore was also kept in the analysis as an interesting interaction partner.

As stated in the introduction, the conformities have an "intermediate" role as they depend on comfort and sportiness to some extent; therefore, the leading question for the development of the model was if and how the hypothetic predictors comfort, sportiness and the conformities predict sound quality, and especially whether conformity (in this case only type conformity) is tied to comfort and sportiness in interaction terms or if it exerts a main effect on sound quality. For both cases, many models were possible but those who employed conformity as a separate main effect had a higher R^2_{adj}. Table 3.3 shows the model that finally emerged as the best-fit model. While comfort and sportiness interact with the driven car and the driver type, type conformity unfolds its greatest predictive power when it is used as a pure main effect. A depiction of the dependencies between the constituents of the presented model for sound quality is given in Figure 3.6.

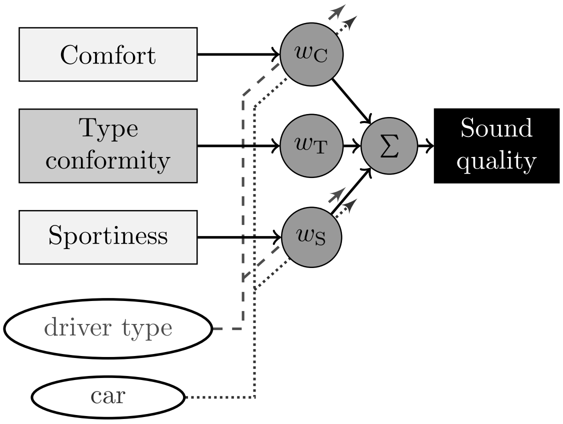


Figure 3.6: Model for sound quality, derived from the experimental data. Weights w_i vary depending on driver type and car.

In the following, the regression coefficients w_i associated with the best-fit model will be discussed. It is noteworthy that the w_i of comfort and sportiness, which occur as interaction terms with the categorical variables driver type and car, vary depending on the specific combination of driver type and car. As the explorative data underlying these estimates are unbalanced, only the sign of the significant w_i will be considered. For all cars except the roadster (A), higher comfort leads to higher sound quality. For *sound-oriented fun-drivers* and *noise-sensitive aesthetes*[4] in the roadster (A), higher sportiness leads to higher sound quality; for *noise-sensitive aesthetes* in the compact car (A), the mid-size car (A) and the original condition (B), higher sportiness leads to lower sound quality; for *noise-sensitive aesthetes* in the rough/sporty condition (B), higher sportiness leads to higher sound quality. The higher the type conformity, the higher the sound quality.

A general distinction is found between the non-sporty cars on the one hand (i.e., the compact car and the sedans in all sound conditions) and the sporty car on the other hand (i.e., the roadster); the sound quality of the former is mainly influenced by comfort, and the sound quality of the latter is mainly influenced by sportiness. In addition, especially for the influence of the sportiness, the driver types come into play. In particular the *noise-sensitive aesthetes* make distinctions with regard to the car; for the compact car (A), the mid-size car (A) and the original condition (B), sportiness affects their sound quality judgement negatively, as it might be considered inappropriate and noisy. In contrast, for the roadster (A) and the rough/sporty condition (B), where in fact a sportier sound is present, it positively affects their sound quality ratings. It seems that the *noise-sensitive aesthete* is able to perceive sportiness, but does not demand it in general, because of his/her noise sensitivity; when a sporty sound is present, however, he/she can appreciate it. In contrast, the *sound-uninterested pragmatist* does not care at all about sportiness, and may even be unable to perceive it, which is well in line with

[4]For *noise-sensitive aesthetes* this regression coefficient only slightly misses significance ($p = .054$).

the disinterest of this driver type in vehicle sounds, indicated already by the choice of name. Type conformity has a significant positive main effect on sound quality, but its η^2_{partial} is the lowest among the model's constituents. As it was itself expected to be influenced by comfort and sportiness, it remains unclear how it effectively affects sound quality. This "entanglement" will be further considered in the submodel for type conformity (see Section 3.4.5).

After having determined the best-fit model for sound quality emerging from the data, submodels for the three main predictor variables comfort, sportiness and type conformity (gained in a similar manner) will be presented in the following sections.

3.4.3 Submodel for comfort

In the previous section, comfort, sportiness and type conformity were found to form the basis of sound quality. The next step is now to apply the same approach for modeling comfort instead of sound quality, obtaining a "submodel" to the sound quality model. The same variables as before (cf. Section 3.4.1) are considered as potential predictors for comfort.

The initial reduction of predictor variables revealed that the following variables can potentially predict the rated overall comfort: driven car; extraversion (from Big Five); type conformity; make conformity; dynamic comfort; static comfort; importance of comfort for sound quality.

In the next step, these variables were used as pure main effects. It was observed that the driven car, extraversion, make conformity, dynamic comfort and importance of comfort for sound quality are not significant. However, all of these variables were still considered as being potentially relevant for comfort; after testing the model configurations that seemed reasonable from what was stated in the introduction, it became clear that dynamic comfort, importance of comfort for sound quality and make conformity could indeed be excluded as they did not increase the model's predictive power. From the remaining variables, the model with the best R^2_{adj} is shown in Table 3.4. The model consists of three two-way predictors each including the driven car, and is visualized in Figure 3.7.

The significant regression coefficients w_i imply that for all sedans (i.e., the mid-size car (A) and all conditions in Location B), higher type conformity leads to higher comfort. For the compact car (A), the roadster (A), the booming/howling condition (B) and the rough/sporty

Table 3.4: Contributions to the best-fit model for the rated acoustic comfort; $R^2 = .716, R^2_{\text{adj}} = .660$.

Predictor variable	p-value	η^2_{partial}
Intercept	.384	.008
car ∗ type conformity	.000	.334
car ∗ static comfort	.000	.256
car ∗ extraversion	.032	.138

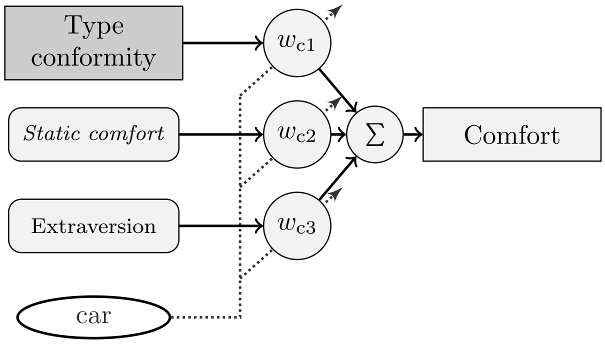


Figure 3.7: Submodel for comfort, derived from the experimental data. Weights w_i vary depending on the car.

condition (B), higher static comfort leads to higher comfort. For the roadster (A), higher values of extraversion lead to higher comfort.

Again type conformity comes into play and affirms the finding from the model for sound quality, namely that type conformity of non-sporty cars is connected to comfort; the reversed influence observed in this model (i.e., type conformity influences comfort) results from the choice of dependent and independent variables. This connection conforms to the depictions in Bisping (1997) and Zeitler and Zeller (2006), where luxury cars are positioned at the regions corresponding to high pleasantness or comfort. It seems that the sedans with conform sound (i.e., mid-size car (A) and original condition (B)) do not need the static comfort to be rated as comfortable, whereas the other cars (i.e., the compact car (A), the sporty roadster (A) and the non-conform booming/howling condition (B) and rough/sporty condition (B)) do need the static comfort. It is possible that the participants already have their predetermined comfort judgement in mind, and stick to it as long as the experienced acoustic comfort does not differ too much from it. The influence of extraversion, which is one of the Big Five personality traits, for the roadster might be a form of justification if the driver cares about his/her fellows: his/her car is not just sporty and fun-oriented, but also comfortable, which might socially be more accepted (for "socially desirable responding", see e.g. Paulhus, 1991). Another possible explanation is that extroverted drivers are especially attracted by sports cars, and thus give in general better ratings for those as a kind of "halo effect" (cf. Thorndike, 1920).

The static driving conditions (in the form of the variable "static comfort") have shown to be important for the rated comfort. This affirms the findings of Section 3.3 where the high importance of static driving conditions was found, in particular for comfort. The participants' conscious as well as subconscious ratings seem to match up.

Table 3.5: Contributions to the best-fit model for the rated acoustic sportiness; $R^2 = .829$, $R^2_{\text{adj}} = .797$.

Predictor variable	p-value	η^2_{partial}
Intercept	.709	.002
car ∗ importance of sportiness for type conformity ∗ type conformity	.000	.456
car ∗ mean sportiness	.000	.414
driver type ∗ importance of make conformity for sound quality	.000	.290

3.4.4 Submodel for sportiness

The initial reduction of predictor variables revealed that the following variables can potentially predict the overall rated sportiness: the driven car; the driver type; importance of sportiness for sound quality; importance of make conformity for sound quality; type conformity; importance of sportiness for type conformity; importance of sportiness for make conformity; GNAT score; mean sportiness.

In the next step, these variables were used as pure main effects. It was observed that type conformity, the importance of make conformity for sound quality, the importance of sportiness for sound quality, the importance of sportiness for make conformity and the GNAT score are not significant. The latter three of these were excluded from further analyses because they did not contribute significantly to any of the model configurations that seemed reasonable from what was said in Section 3.1. Table 3.5 shows the model that emerged as the best-fit model. The model consists of a three-way predictor and two two-way predictors. The variables driven car, type conformity and mean sportiness are more related to the actual car sound, whereas importance of sportiness for type conformity, importance of make conformity for sound quality and the driver type are more related to the subject. The model is visualized in Figure 3.8.

The significant regression coefficients w_i imply that for the roadster (A) and the mid-size car (A), the simultaneous occurrence of high rated importance of sportiness for type conformity and high type conformity leads to higher sportiness. For the compact car (A) and all conditions in Location B, higher mean sportiness leads to higher sportiness. For the *sound-uninterested pragmatist*, higher rated importance of make conformity for sound quality leads to higher sportiness.

The influence of type conformity on sportiness is moderated by the importance of sportiness in the roadster (A) and the mid-size car (A). That means that for those participants who find sportiness important, type conformity is directly correlated to sportiness, which seems a reasonable outcome. However, the indicated direction of the effect is again attributed to

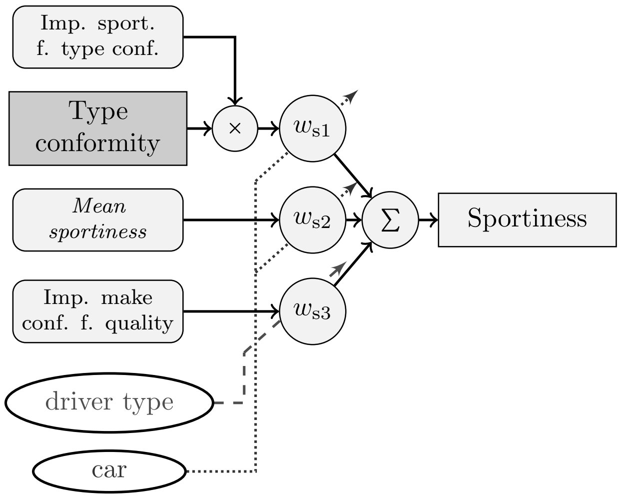


Figure 3.8: Submodel for sportiness, derived from the experimental data. Weights w_i vary depending on driver type or car.

the chosen roles for the variables (independent vs. dependent), as in the comfort submodel in the previous section. The influence of the mean sportiness can be interpreted similarly to the influence of static comfort in the comfort model (see Section 3.4.3): participants might have a predetermined sportiness rating in mind, which they dismiss only if the experienced acoustic sportiness differs considerably. As the affected cars are not influenced by aspects of type conformity in the model, the dependency on the mean sportiness might be due to the lack of a prospect of the sportiness of these cars. Those *sound-uninterested pragmatists* who attach more importance to a make conform vehicle sound tend to give higher sportiness ratings. As there was only one single make tested in this study, which belongs to the premium segment, it is supposed that again a halo effect is observed: the cars get rated higher due to the assumed “premiumness” of the make by the *sound-uninterested pragmatists* who, as discussed in Section 3.4.2, might be unable to perceive sportiness.

3.4.5 Submodel for type conformity

The initial reduction of predictor variables revealed that the following variables can potentially predict the overall rated type conformity: openness (Big Five); sportiness; comfort; make conformity; static comfort; driven car.

In the next step, these variables were used as pure main effects. It was observed that sportiness, static comfort and openness are not significant. The latter two were excluded from further analyses, as static comfort was regarded as redundant because rated comfort was already in the model. Openness, which is one of the Big Five personality traits, was only assumed to be meaningful as a main effect. The best-fit model can be seen in Table 3.6. It consists of three two-way predictors, each of which includes the driven car, and is visualized in Figure 3.9.

Table 3.6: Contributions to the best-fit model for the rated type conformity; $R^2 = .853, R^2_{\text{adj}} = .825$.

Predictor variable	p-value	η^2_{partial}
Intercept	.720	.001
car ∗ make conformity	.000	.493
car ∗ comfort	.000	.339
car ∗ sportiness	.011	.159

The significant regression coefficients w_i imply that for all cars except the original condition (B), higher make conformity leads to higher type conformity. For the mid-size car (A) and the original condition (B), higher comfort leads to higher type conformity. For the compact car (A), higher sportiness leads to higher type conformity.

Make conformity has the strongest influence on type conformity. However, no variability was provided with regard to the cars' makes as they all were of the same make. The significant correlation between the ratings of the two conformities might result from an "integral conformity" that the participants effectively rated. Thus, also an alternative model without make conformity as a predictor is considered. The alternative model just consists of the interaction between the driven car and the rated comfort and the interaction between the driven car and the rated sportiness (see Table 3.7).

In this alternative model, the significant regression coefficients w_i imply that for all cars except the roadster (A), higher comfort leads to higher type conformity. For the compact car (A) and the roadster (B), higher sportiness leads to higher type conformity.

This submodel confirms the findings of the previous models: type conformity of non-sporty cars is dominated by comfort, type conformity of sporty cars is dominated by sportiness. The exception is the compact car (A), where sportiness has a positive effect on type conformity, too. It is possible that in the context of a compact car, the participants interpret loudness or unpleasantness as sportiness, or maybe it is pointless to assess the sportiness of such a (rather

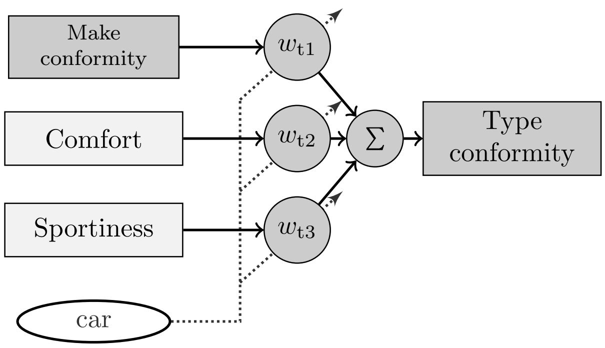


Figure 3.9: Submodel for type conformity, derived from the experimental data. Weights w_i vary depending on the car.

Table 3.7: Contributions to the alternative model for the rated type conformity; R^2 = .710, R^2_{adj} = .675.

Predictor variable	p-value	η^2_{partial}
Intercept	.889	.000
car ∗ comfort	.000	.656
car ∗ sportiness	.000	.455

pragmatic) car, which is why the rating might follow the overall impression of type conformity.

This submodel recaps the earlier observed correlation between comfort and sportiness on the one hand and type conformity on the other hand (Sections 3.4.3 and 3.4.4). However, the supposable *causal* relationship is reflected in the present submodel, namely comfort and sportiness determining type conformity (and not vice versa). It is noteworthy that apart from the driven car there are no further predictors in the alternative model for type conformity, especially not the driver type.

As make conformity was not encompassed in the sound quality model (see Section 3.4.2) and was highly correlated with type conformity for the data at hand, no separate model for make conformity was fitted.

3.4.6 Final model for sound quality

The final model for overall sound quality assessment can be visually assembled from the submodels discussed in Sections 3.4.2 to 3.4.5, and it is depicted in Figure 3.10. It summarizes the dependencies found in the previous sections. First and foremost, it confirms that sound quality depends on comfort, sportiness and conformity. While type conformity seems to be a general requirement for sound quality, the influence of comfort and sportiness depends on the driven car as well as the driver type. This influence boils down to the following: if a non-sporty car is considered, comfort is required; for a sporty car, sportiness is required; for *noise-sensitive aesthetes*, sportiness can decrease sound quality if it is not adequate for the car; for *sound-uninterested pragmatists* sportiness does not play a role at all. Although the drivers have a different notion of how a car should sound, they agree on whether a given sound matches a certain vehicle type, as the driver type is not directly involved in type conformity.

The car itself has a huge impact on the whole assessment process, as can be seen from its moderating influence on almost all of the weights w_i. Only w_{T} and w_{s3} are independent of the car, indicating the paramount role of type conformity for sound quality, and the *sound-uninterested pragmatists*' strategy of assessing sportiness from make attributes, respectively.

Overall comfort and sportiness can be traced back to the driving conditions: static driving conditions, such as driving with constant speed, are particularly important for comfort, while sportiness manifests in all considered driving conditions. For comfort ratings in the sporty car,

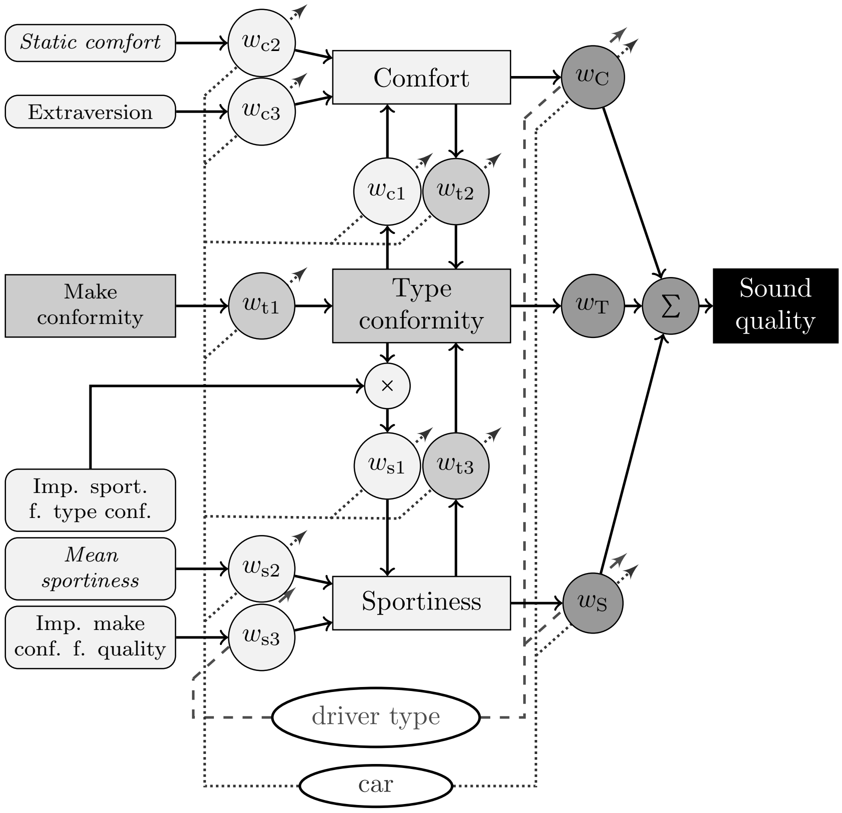


Figure 3.10: Final model for overall sound quality, assembled from the previous submodels. Weights w_i vary depending on driver type and/or car.

also the personality trait extraversion plays a role.

A striking observation is the "dependency-loop" between comfort and sportiness on the one hand and type conformity on the other hand, i.e., type conformity depends on comfort and sportiness and vice versa. This is considered to be an artifact from the modeling approach and the visual assembly in Figure 3.10: due to the successive but independent construction of the submodels, the aforementioned quantities take turns being predictors or dependent variables, respectively, which implies a certain direction of causality. However, *correlation* rather than causality can be determined in linear regression models. Sportiness, comfort and type conformity are indeed valuable predictors for each other, as there seems to exist a tight relation between these, but from what was said in the introduction it seems more likely that comfort and sportiness cause the outcome of type conformity rather than vice versa. That is to say, in Figure 3.10 the paths via w_{t2} and w_{t3} rather than w_{c1} and w_{s1} are to be used in future model calculations.

To further test this "intermediate" role of type conformity, a *mediation analysis* was performed on the respective part of the model. To be precise, comfort and sportiness were taken as the two independent variables, sound quality was taken as the dependent variable, and type conformity was put in between as a mediator. In this model, both comfort and sportiness have an influence on type conformity (moderated by the car) and on sound quality (moderated by the interaction between car and driver type). In addition, also type conformity has an (unmoderated) effect on sound quality. In mediation analysis, it is tested whether the *indirect effect* via type conformity, characterized by the product of the respective regression coefficients, is significantly different from zero (Preacher et al., 2007). This was realized using *bootstrapping* with $k = 10^6$ random samples. It was observed that the indirect effect originating from comfort was significantly positive for all cars except the roadster (A) ($p = .122$), while the indirect effect of sportiness was only significant (and positive) for the compact car (A) ($p = .015$) and the roadster (A) ($p = .001$). This corroborates the earlier interpretation.

Since multicollinearity, i.e., linear dependencies between the predictor variables, can substantially lower the model estimation accuracy, the *variance inflation factor* (VIF) was calculated for all predictors. A VIF greater than 10 indicates serious multicollinearity between the predictors, which leads to inaccuracy of the estimated weights w_i (Field, 2009). In the set of all predictors used in the final model, the highest VIFs are those of type conformity (4.97) and make conformity (4.87). The high correlation between these two was already observed and discussed in Section 3.4.5, and without make conformity the VIFs are all smaller than 3.99, which is the value for sportiness.

3.5 Discussion

As discussed in the beginning, vehicle sound quality is assumed to depend on (acoustic) comfort, sportiness, type conformity and make conformity. The aim of this study was to specify the relations between these quantities, as well as the influence of contextual aspects on vehicle sound assessment from data of an exploratory study. In the developed model, the driven car and the driver are included as moderating factors, and different combinations of driving conditions account for the averaged comfort and sportiness terms. Since unnecessary predictors were excluded during the modeling procedure, it can be concluded that the model's predictive power indeed benefits from the consideration of these contextual aspects. In the final model, especially comfort and sportiness determine sound quality, in a way that depends on the driven car and driver type: sporty cars are basically determined by sportiness, the other vehicle types are basically determined by comfort. It showed that *sound-uninterested pragmatists* are not sensitive to sportiness, while *sound-oriented fun-drivers* and *noise-sensitive aesthetes* are; however, *noise-sensitive aesthetes* in most cases do not require sportiness, or find it even annoying, as opposed to *sound-oriented fun-drivers*, who enjoy sporty sounding cars. As expected, type conformity plays an important role but is entangled with comfort and sportiness. Although hypothesized from literature, make conformity itself was not found to be a significant predictor for sound quality in this study. However, just cars of a single make were used, and thus no variance was provided in this dimension. It rather seems that the participants rated an "overall conformity" for the particular car, as seen by the strong interrelatedness of make conformity and type conformity. As customer expectations are particularly high for premium makes, the absence of make conformity in the sound quality model is considered to be a modeling artifact. Hence, it will be interesting to see if this finding will be confirmed with data that contain cars of different makes.

Furthermore, comfort showed to be mainly dominated by "static driving conditions", such as constant driving at different speeds, and to some extent by the personality trait of extraversion. Thus, when optimizing the sound of a non-sporty car, acoustic engineers can first focus on static driving conditions before making the dynamic situations more comfortable. Sportiness seems to depend on all driving conditions, but is especially required by the participants in dynamic situations such as accelerating and overtaking. As there was only one real sports car among the tested vehicles, it just might not have been possible to resolve the individual role of the driving conditions in the sportiness model. Type conformity depends on comfort and sportiness, but their influence differs depending on the vehicle type: sporty cars need sportiness to be type conform, while other cars need comfort to be type conform. It seems to be of great importance to understand the nature of type conformity (and possibly make conformity) in future research as it showed to be related to *all* other quantities, and thus plays a central role in the whole vehicle sound assessment process.

The presented model was obtained from highly contextualized data of an exploratory study, which allows for a high ecological validity. As a field study, however, many of the parameters were not controllable as it would be the case in a pure listening experiment. A future challenge will be to investigate the differences between judgements made in a car context and such made in a listening experiment, and especially to compare the results in the light of the model described above.

Acknowledgement

The authors would like to thank Reinhard Weber, formerly University of Oldenburg, for his support in designing the study and valuable discussions.

4 Field versus lab: situational influences on vehicle sound assessment[1]

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 Maiberger et al. (2018b) 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 Maiberger et al. (2018b). 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 Maiberger et al. (2018a), makes meaningful distinctions with respect to the evaluation of vehicle sounds.

[1]This chapter is based on Maiberger et al. (2018c), a manuscript currently under revision:
D. Maiberger, E. Strasser, U. Letens, S. van de Par: Field versus lab: situational influences on vehicle sound assessment. Acta Acustica united with ACUSTICA, 2018, currently under revision.

4.1 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 (Bisping, 1997, Västfjäll et al., 2002, Sköld et al., 2005, Zeitler and Zeller, 2006). 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 Blauert and Jekosch (1997) 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 (Blauert and Jekosch, 1997). The "cognitive" situation encompasses aspects such as prior knowledge, expectation, experience, taste or social and cultural influences (Blauert and Jekosch, 1997, Bednarzyk, 1999). Several authors point out the role of the spatial, time, semantic and response context of an experiment on the rated sound quality (Zeitler et al., 2006, Guski and Blauert, 2009, Fiebig, 2015). The semantic context, for example, is related to the meaning of the sound (Jekosch, 1999). Additional visual information about the stimuli given in the instruction (Susini et al., 2009) or during the experiment (Fiebig, 2015) 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 (Zeitler et al., 2006, Guski and Blauert, 2009). 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 (Bednarzyk, 1999). 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 (Steffens, 2013, Fiebig, 2015). 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 (Västfjäll, 2004). Sound quality evaluation can be influenced by "input from non-auditory modalities" (Blauert and Jekosch, 1997). For example, the simultaneous presentation of auditory and visual (Fiebig, 2015) or vibrotactile (Ho et al., 2007) stimuli can influence subjective reactions as compared 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 (Krebber, 2000). 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. (2018b) developed a comprehensive model for vehicle sound quality evaluation from data of an exploratory, contextualized field study (see Chapter 3). The model takes the acoustic factors *comfort*, *sportiness*, *type conformity* and *make conformity*, which are commonly treated in literature (Bisping, 1997, Zeitler and Zeller, 2006, Letens, 2010, Köber et al., 2016), 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* (Maiberger et al., 2018a) moderate the strength of the relationship between input and output variables. According to Maiberger et al. (2018a), 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 (Maiberger et al., 2018b). As this model is based on data from a field study, it is assumed to be ecologically valid.

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

1. Can the model of Maiberger et al. (2018b), 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 Maiberger et al. (2018b) 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.

4.2 Methods

4.2.1 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 Maiberger et al. (2018b). Table 4.1 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 (MathWorks, 2014). 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 4.1). 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 graying 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.

Table 4.1: 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

13 *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 Mercedes) and car segments (i.e., compact car, mid-size (luxury) car, full-size luxury car, roadster, coupé,

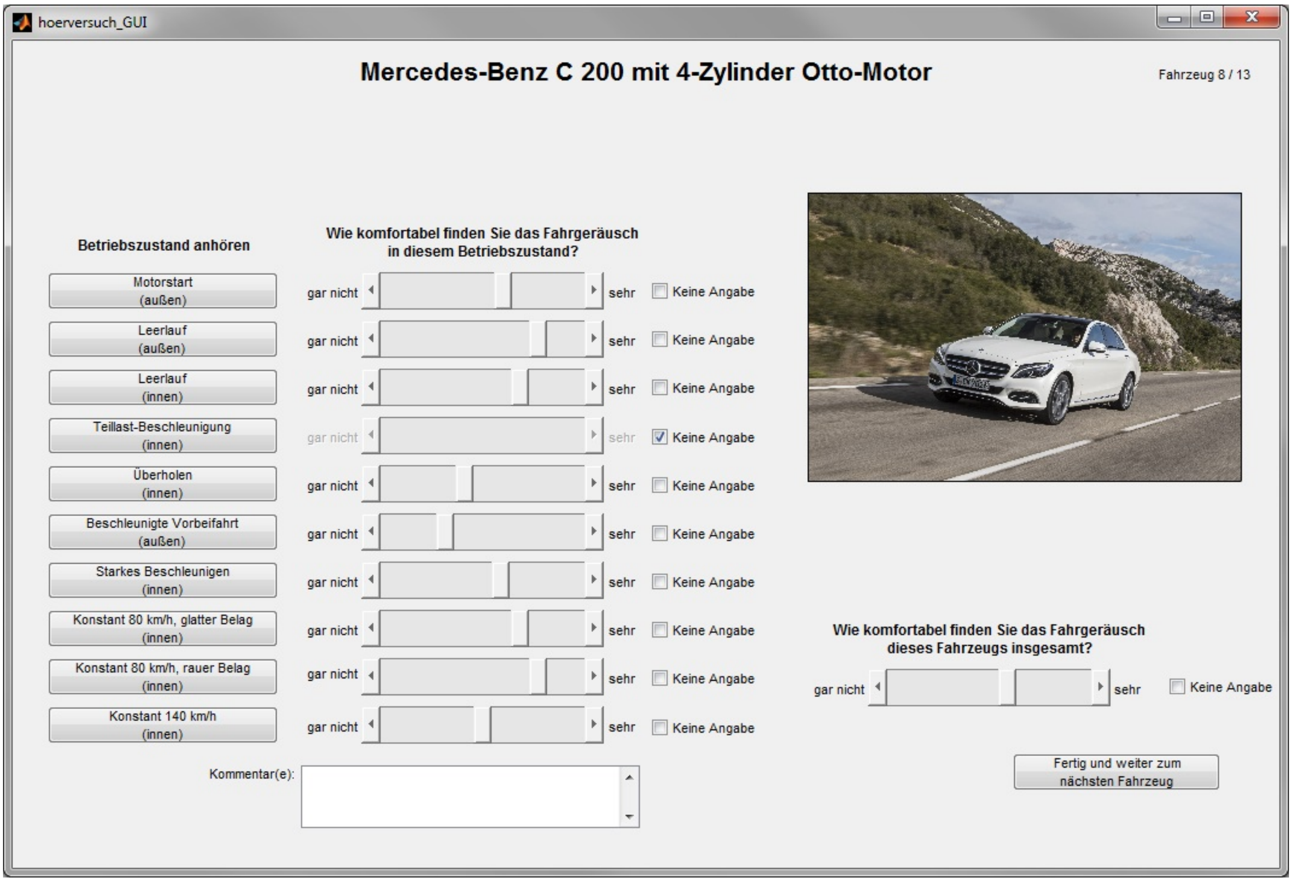


Figure 4.1: Layout of the graphical user interface used for the listening experiment. The interface for Pass 2, rating of comfort, is shown (cf. Section 4.2.2).

convertible, SUV), see Table 4.2. Since for nine cars the recordings of the constant driving conditions (8–10 in Table 4.1) 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 4.3); 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

Table 4.2: 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

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 4.3.

Table 4.3: Combinations forming the mismatched cars in the listening experiment; the numbers in the second and third column refer to the original cars from Table 4.2.

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

4.2.2 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.

4.2.3 Sample

The participants were recruited via notice boards at the University of Oldenburg. Since it was found in a previous study (Maiberger et al., 2018b) that *sound-oriented fun-drivers* are rare amongst the university members, additional participants were recruited from a vintage

Table 4.4: 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

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. Maiberger et al., 2018a). As can be seen from Table 4.4, 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 (Maiberger et al., 2018a).

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.

4.2.4 Statistical methods

4.2.4.1 General linear models

For the statistical analyses in this study, 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 study, 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" (Field, 2009). 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 Field, 2009, 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.

4.2.4.2 Pearson's correlation coefficient

Pearson*'s correlation coefficient* R measures the amount of linear association between two variables (see Field, 2009). For significance testing in this study, 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.

4.2.4.3 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 Field, 2009). 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.

4.2.4.4 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 Field, 2009). EFA will be used in Section 4.3.3 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.

4.3 Results

4.3.1 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.

4.3.1.1 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 Frigge et al., 1989). 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 4.3.2. This participant is a *sound-uninterested pragmatist*.

4.3.1.2 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 ($\Lambda = .82$, $F(3, 312) = 23.16$, $p < .05$, $\eta_\mathrm{p}^2 = .18$). This effect affects sportiness ($F(1, 314) = 69.56$, $p < .02$, $\eta_\mathrm{p}^2 = .18$) but neither sound quality ($F(1, 314) = 1.05$, $p > .02$) nor 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.

4.3.2 Influence of car identification

As described in Section 4.2.1, 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 4.2 and 4.3).

The overall effects of both type conform identification ($\Lambda = .94$, $F(5, 600) = 8.12$, $p < .05$, $\eta_\text{p}^2 = .06$) and make conform identification ($\Lambda = .94$, $F(5, 600) = 7.07$, $p < .05$, $\eta_\text{p}^2 = .06$) were found to be significant. Follow-up analyses showed that the type conform identification affects sportiness ($F(1, 604) = 7.10$, $p < .05$, $\eta_\text{p}^2 = .01$) and type conformity ($F(1, 604) = 18.11$, $p < .05$, $\eta_\text{p}^2 = .03$), 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 ($F(1, 604) = 14.64$, $p < .05$, $\eta_\text{p}^2 = .02$), on type conformity ($F(1, 604) = 10.20$, $p < .05$, $\eta_\text{p}^2 = .02$) and on make conformity ($F(1, 604) = 5.74$, $p < .05$, $\eta_\text{p}^2 = .01$). 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

Table 4.5: 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 gray; 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	

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.

4.3.3 Factor analysis of driving conditions

In Maiberger et al. (2018b), 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.

4.3.3.1 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 4.5. 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,

Table 4.6: 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

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.

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, β = .236, p < .05), *static comfort* (b = 0.340, β = .288, p < .05) and *dynamic comfort* (b = 0.495, β = .506, p < .05) 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 (Maiberger et al., 2018b). 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 Maiberger et al. (2018b). However, the N_5 diffuse field loudness[2] (DIN 45631/A1, 2010) of the driving conditions comprising dynamic comfort is higher than for the driving conditions comprising static comfort, as can be seen from Table 4.6. 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 (Steffens, 2013).

4.3.3.2 Sportiness

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

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

Table 4.7: 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 gray; 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	

(Field, 2009, 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 4.7. 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* ($b = 0.299$, $\beta = .211$, $p < .05$) and *dynamic sportiness* ($b = 0.794$, $\beta = .675$, $p < .05$) were found to be statistically significant, with $R^2 = .617$. It is concluded that overall sportiness is mainly determined by *dynamic sportiness*.

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 (Maiberger et al., 2018b). As the significance of the dynamic driving conditions for sportiness was concluded from the participants' answers in Maiberger et al. (2018b), 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.

4.3.4 Validation of sound quality model

In Maiberger et al. (2018b) 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 4.2. 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 modelling 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.

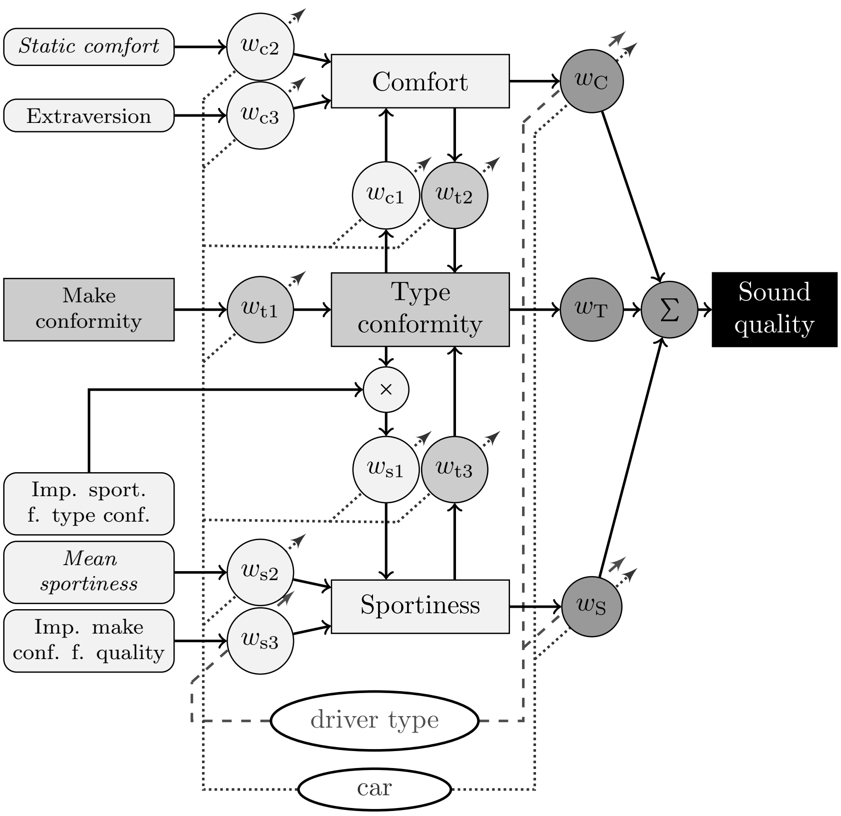


Figure 4.2: Model for sound quality assessment, taken from Maiberger et al. (2018b).

In the current study only the rightmost part of Figure 4.2 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.

4.3.4.1 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 4.8). Compared to the field data from Maiberger et al. (2018b) 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 4.2.1) 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 4.2) 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,

Table 4.8: Contributions to the sound quality model with the individual cars; $R^2 = .806$, $R^2_{\text{adj}} = .741$.

Predictor variable	p-value	η^2_{p}
Intercept	.103	.006
car $*$ driver type $*$ comfort	.000	.571
car $*$ driver type $*$ sportiness	.000	.260
type conformity	.000	.113

Table 4.9: Contributions to the sound quality model with five car segments; $R^2 = .757, R^2_{\text{adj}} = .744$.

Predictor variable	p-value	η_{p}^2
Intercept	.061	.006
segment ∗ driver type ∗ comfort	.000	.521
segment ∗ driver type ∗ sportiness	.000	.169
type conformity	.000	.149

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 4.2.1, 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 Maiberger et al. (2018a). 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 4.9.

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.

4.3.4.2 Equality of sound quality models

Up to now the principal structure of the sound quality model (as found in Maiberger et al., 2018b) 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 Maiberger et al. (2018b) 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 Maiberger et al. (2018b).

The effect sizes η_{p}^2 of both sound quality models are given in Table 4.10. 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.

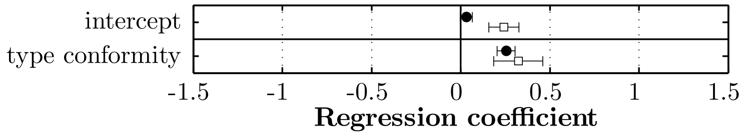
The regression coefficients b_i, which give the amount of change in predicted sound quality per unit change of the predictors (i.e., comfort, sportiness or type conformity) for each combination

Table 4.10: Effect sizes η_{p}^2 of sound quality model applied to field test data (Maiberger et al., 2018b) and lab-test data (this study), respectively.

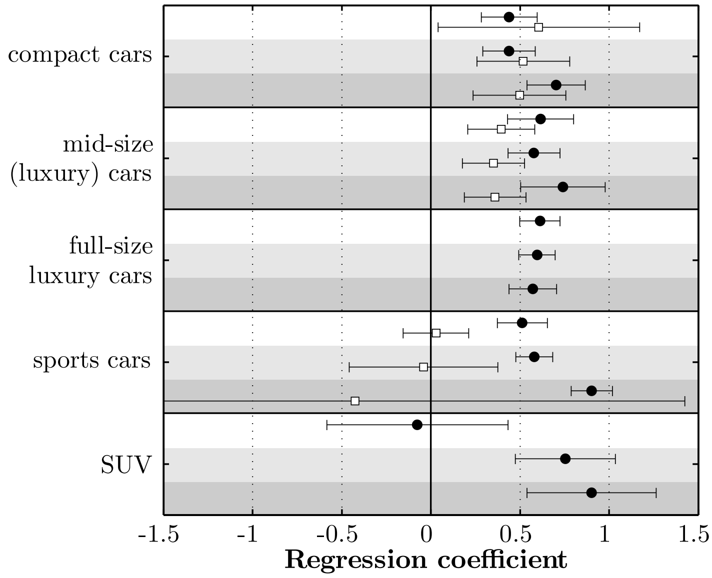
Predictor variable	η_{p}^2 Lab	η_{p}^2 Field
Intercept	.006	.270
segment ∗ driver type ∗ comfort	.521	.249
segment ∗ driver type ∗ sportiness	.169	.242
type conformity	.149	.195

of segment and driver type, are shown in Figure 4.3 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.

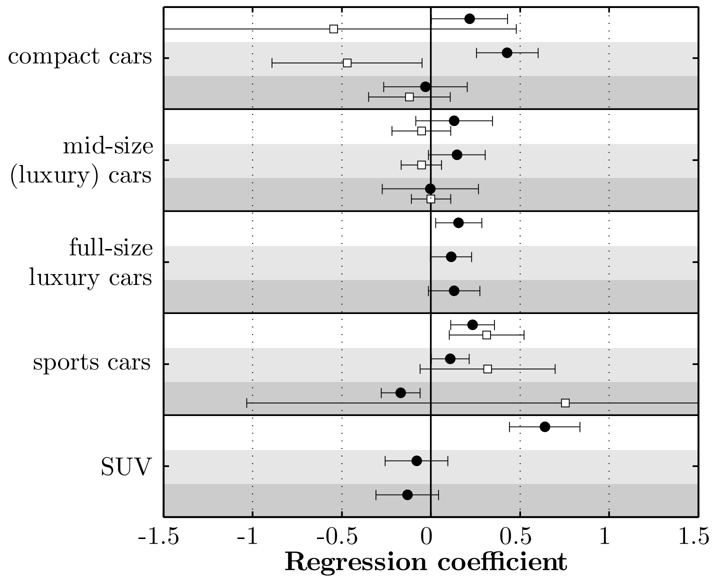
It is observed that for almost all coefficients the confidence bounds between studies overlap, implying that the corresponding values do not differ between lab and field, statistically speaking. The only exceptions are the intercept, the regression coefficients of comfort for *sound-oriented fun-drivers* 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 4.3.1.2.



(a) Intercept and regression coefficient b_i for type conformity.



(b) Regression coefficients b_i for comfort.



(c) Regression coefficients b_i for sportiness.

Figure 4.3: 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 gray regions correspond to *noise-sensitive aesthetes* and dark gray regions correspond to *sound-uninterested pragmatists*.

4.4 Discussion

This study investigated to what extent a model developed to predict vehicle sound quality using field tests (cf. Maiberger et al., 2018b) 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 (Sanders and Poeppel, 2007). 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 (Maiberger et al., 2018b), 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 (2013) found, 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*, the latter of which had already been assumed in Maiberger et al. (2018b). 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 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 Maiberger et al. (2018b) 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 Steffens (2013) 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 Maiberger et al. (2018b), the driver typology from Maiberger et al. (2018a) 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 Maiberger et al. (2018b); *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 chapter 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 4.2) 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 study with other makes in future research.

5 Discussion

The aim of this work was to establish a model that retraces the subjective evaluation process of vehicle sounds as experienced by the driver. The focus was on the integration of non-acoustic contextual aspects, such as the driver type, the driven car and the situational context. The paramount role of comfort and sportiness for sound quality, as supposed in earlier works, was confirmed in the experiments. The dependency could even be extended, by allowing moderation of their influence by personal aspects (driver type) in combination with the specific car, as well as by considering type conformity as a car-dependent mediator. Although the situational context was found to be influential on the strengths of certain of the model's relationships, the principal model structure was found to be valid in both field and laboratory contexts.

5.1 Summary

In order to operationalize the factors that lie within the subject, a driver typology was developed in Chapter 2 from data of an online survey among drivers of premium vehicles. By means of a cluster analysis three prototypical driver types were determined, which were named *sound-oriented fun-driver*, *noise-sensitive aesthete* and *sound-uninterested pragmatist*. These driver types basically differ in their attitude towards cars and sounds, the attitude being conceptualized by six factors revealed in exploratory factor analysis (EFA). These factors reflect (i) the extent to which the drivers exhibit a dynamic driving style combined with acoustical orientation, i.e., the wish for a functional car sound, (ii) their noise sensitivity, (iii) their awareness for vehicle sounds, (iv) to what extent they see cars as a status symbol, (v) their reservations about cars and (vi) their demand for comfort. While the *sound-oriented fun-drivers* can be characterized by their dynamic driving style and acoustical orientation, the *noise-sensitive aesthetes* have a high noise sensitivity, whereas the *sound-uninterested pragmatists* hardly care about vehicle sounds. Furthermore, the three driver types showed significant differences with regard to their assessment of vehicle sounds and their wishes for future vehicle sounds, as deduced from the results of a semantic differential for rating their own cars' sounds. It was found that these differences were particularly present in the sound-describing dimensions, such as *loud – quiet*, *aggressive – defensive* and *comfortable – sporty*, rather than evaluative scales, such as *unpleasant – pleasant* and *don't like it at all – like it a lot*. For the two evaluative scales, all of the participants' cars scored high, trending towards *pleasant* and *like it a lot*. It was supposed that this result was due to the emotional involvement in the rated cars, as these were the raters' own cars. It was observed that the *sound-oriented fun-drivers*, in contrast to the other two driver types, tended to the sides of the scales that are classically connoted negative or unpleasant, such as *loud*, *aggressive* or *sporty*. It was concluded that the driver typology is capable of dividing car drivers in a meaningful way, making reasonable distinctions with regard to the perception and evaluation of vehicle sounds.

After establishing the typology of car drivers, a regression model for the sound quality assessment was developed in Chapter 3 in order to gain insight into the evaluation process as well as the influence of contextual variables on this process. The data for the modeling procedure were gathered in a field experiment, which in principal offers a high degree of ecological validity. The model resulting from an extensive statistical modeling directly relates sound quality to the acoustic aspects comfort, sportiness and type conformity, which along with make conformity have previously been hypothesized to be influential in the assessment of vehicle sound quality throughout literature. It was found that type conformity is a general requirement for sound quality, while the direct influence of comfort and sportiness on sound quality depends on the combination of the car and the driver type. Most notably, sportiness does not seem to play any role for the sound assessment by *sound-uninterested pragmatists*. For *sound-oriented fun-drivers* and *noise-sensitive aesthetes*, sportiness increases the sound quality of sports cars. However, *noise-sensitive aesthetes* perceive it as annoying in all other car types. Comfort basically has a positive effect on sound quality for all non-sporty cars. In EFA it was found that comfort itself is determined mainly in static driving conditions, i.e., driving with constant speed. Sportiness, on the other hand, manifested in all considered driving conditions, but it was reported by the participants that the dynamic driving conditions (e.g., overtaking and accelerating) play a major role for sportiness. This discrepancy was supposed to be due to the sports cars being underrepresented in the experimental design, making it more difficult to resolve the driving conditions in EFA. Type conformity was found to have a mediating role, i.e., it is determined by comfort and sportiness while it influences sound quality. This "indirect effect" is moderated by the car. To be precise, type conformity is high for sports cars that exhibit a high acoustic sportiness, while for all other cars type conformity is high when these are sounding comfortable.

As the model was developed in the field context, the potential transfer of results gained in a lab context to the field situation was the central question in Chapter 4. From the results of a listening experiment it could be concluded that the principal structure of the model also holds in the laboratory context. However, it was found that the strengths of some relations within the model were different. Most importantly, comfort gains greater importance for sound quality as compared to the field context. This was attributed to an increased sensitivity to loudness in the lab, as loudness can be assumed to be inversely related to comfort (Bisping, 1997, Zeitler and Zeller, 2006). In addition, less distraction is experienced in the lab as compared to the field situation, where sensory input from other modalities is also present. Furthermore, the composition of both comfort and sportiness ratings from the individual driving conditions differed from the field situation. In the lab, the comfort ratings in standstill (e.g., engine start and engine idle) and especially in dynamic driving conditions (e.g., overtaking and accelerating) gained increased importance. Along with the overestimation of loudness, this was attributed to an overrepresentation of these driving conditions in the lab context as compared to the field. For sportiness a higher importance of the dynamic driving conditions was found, as already hypothesized from the findings in Chapter 3. However, as most of the model parameters were found to be of comparable magnitude between the two studies, it was concluded that the proposed model structure is robust for both field and lab-test situations. In order to be able to generalize the results, a working categorization of cars into the five groups of *compact cars*, *mid-size (luxury) cars*, *full-size luxury cars*, *sports cars* and *SUV* was successfully introduced. The division of subjects into the three driver types also proved to make meaningful distinctions in the lab context. In particular, it was observed that the driver types have different evaluation strategies for vehicle sounds, as seen from the differing regression coefficients of comfort and sportiness, and that particularly *sound-oriented fun-drivers* were able to recognize fake cars, whereas some *sound-uninterested pragmatists* failed to give consistent ratings.

5.2 Driver typology

The driver typology developed in Chapter 2 of this work has proven successful in resolving evaluation differences that are due to characteristic personality traits. Accordingly, in order to determine the driver type of a particular person the steps described in the following have to be made. The person has to indicate the degree of agreement with 13 statements from Table 2.2, using the five-point scale ranging from "not at all" to "very much" (see Figure 3.2(d)). For the follow-up calculations, consecutive integers from 1 ("not at all") to 5 ("very much") have to be assigned to the levels of the answering scale. The scale values for the six factors listed in Table 5.1 can then be calculated by averaging the person's responses to Items 10, 20 and 26 (for Factor 1), Items 3 and 5 (for Factor 2), Items 12* and 13 (for Factor 3), Items 22* and 23 (for Factor 4)[1], Items 7 and 8 (for Factor 5), and Items 17 and 25 (for Factor 6). Items marked with an asterisk have to be reversed in polarity before calculating the average, i.e., the integer order has to be inverted, counting down from 5 ("not at all") to 1 ("very much"). Taking the six scale values as coordinates of the person, the driver type of the person is determined as the cluster to the center of which the person has the smallest Euclidean distance (see Table 5.1 for

[1] As compared to the factor loadings in Table 2.4, Factor 4 is defined with reversed polarity in order to be interpretable as the "status aspect".

Table 5.1: Coordinates of the cluster centers of the three driver types *sound-oriented fun-driver* (SF), *noise-sensitive aesthete* (NA) and *sound-uninterested pragmatist* (SP).

No.	Factor	SF	NA	SP
1	Dynamic driving style & acoustical orientation	3.737	3.055	2.605
2	Noise sensitivity	2.306	3.908	3.024
3	Sound awareness	3.846	3.758	2.378
4	Status aspect	3.229	2.972	2.149
5	Reservations about cars	2.122	3.209	2.871
6	Demand for comfort	3.643	4.216	4.068

the coordinates of the cluster centers).

In the future, the questionnaire would probably benefit from adding further items that allow for a broader and thus more robust coverage of the six factors, all of which are represented with at most three items.

Although not explicitly included in the driver typology, the personality trait of "extraversion" from the *Big Five dimensions of personality* was found to be influential for comfort in the field test. This influence was observed for the sports car, for which "static comfort" (i.e., the comfort experienced during driving with constant speed) did not play a role. This indicates that in some cases attitude rather than the actual sound determines subjective ratings, which is why acoustic engineers engaged in sound optimization also have to exhibit psychological skills. Though the driver typology has proven quite successful throughout this work, this result indicates that it might be worthwhile to extend the typology in the future with, e.g., the Big Five dimensions of personality.

The fourfold construct of noise sensitivity as found by Notbohm (2010) has been part of the development of the driver typology. The four distinct aspects distinguished by Notbohm (2010) are (i) "critical attitude towards (environmental) noise", (ii) "activation by music", (iii) "noise sensitivity and disturbance by noise" and (iv) "appreciation of silence and natural sounds". In the course of the development of the driver typology, however, the first and the third of these aspects regathered in the factor "noise sensitivity", while the other two aspects showed ambiguous loadings and were thus not further considered. This discrepancy can probably be attributed to the sparse coverage of these four aspects in the current work, each represented by two items at most. In addition, several other personality traits were tested within the same task, which is probably why the four factors found by Notbohm could not be fully resolved. However, none of the items belonging to these other traits had high loadings on the factor "noise sensitivity", and thus no mixing between these traits did occur within the six revealed factors.

5.3 Sound quality model

The sound quality model presented in Chapter 3 describes the relationship between comfort, sportiness and type conformity as the determining parameters of sound quality. Type conformity was found to play a central role in sound quality assessment, and it is a general requirement for high sound quality. Type conformity, acting as a mediator, is determined by comfort and sportiness depending on the car type. In this work, make conformity dropped out during the modeling process, which was probably due to an imbalanced test design. However, Zeitler et al. (2006) state that differences due to make are expected to be smaller than differences due to vehicle type. Nevertheless, as type conformity plays a central role in sound quality assessment, it must be assumed that a similar connection exists between make conformity and sound quality. In order to further investigate this hypothetic connection, field tests should be conducted that cover a broader range of vehicle makes. As field tests are generally more time-consuming than lab tests, it will be a challenging task to cover a broader range of segments *and* makes to a satisfactory extent.

The direct influence of comfort and sportiness on sound quality depends on the combination of driver type and car type. Interestingly, for some of the cars in the field study participants had a predetermined rating in mind, which was only reconsidered if the experienced sound deviated dramatically from their expectation. This was concluded from the influence of static comfort and mean sportiness on those cars that were not expected to sound particularly comfortable or sporty, respectively. This indicates that expectation has a certain significance in the evaluation process, as already pointed out in Blauert and Jekosch (1997), Bednarzyk (1999), Zeitler et al. (2006). It is therefore necessary to further examine these expectations in order to formulate suitable target sounds or to actively shape customers' expectations by adequate marketing.

The driving conditions influencing the evaluation of overall sportiness are the dynamic conditions, i.e., accelerating with full load or overtaking. For the evaluation of overall comfort in the field, especially "static comfort" is important, i.e., the comfort experienced while driving with constant speed of 50 km/h or 80 km/h. The significance of "static comfort" in the field was opposed in the lab by an equally high importance of the other two comfort factors "dynamic comfort" and "standstill comfort". This finding might be due to the route profile in the field test, which required driving with constant speed most of the time. In contrast, in the lab the ten presented driving conditions spread evenly across the three factors "standstill comfort", "static comfort" and "dynamic comfort". The significance of "standstill comfort" and "dynamic comfort" in the lab could be due to the mere presence of these driving conditions. This is related to the line of reasoning put forward by Zeitler et al. (2006) who point at the stimulus context of sound evaluation, i.e., the influence of the specific set of stimuli that are tested together. As pointed out earlier, also the extent to which the rating context resembles the real usage context can influence the sound quality assessment (Steffens, 2013, Fiebig, 2015).

In Chapter 4, a first approach to validate the sound quality part of the model was presented. Further model validation should be conducted in the future, including the submodels for comfort, sportiness and type conformity. As the proposed model can also be viewed as a *structural equation model* (SEM) (see Bollen, 1989), the accuracy of the model could be checked by analyzing a new set of experimental data with the methodology of SEM in which care is taken of all dependencies simultaneously.

Although developed from exploratory data, the model in its current form seems to be “handable” for applications, e.g., in sound quality metrics. The model only needs a manageable set of context parameters, i.e., one out of three driver types, one out of five car segments, and consideration of at most three comfort factors plus two sportiness factors aggregated from the driving conditions.

5.4 Outlook

Type conformity plays a central role in vehicle sound assessment: it is related to many of the other quantities and serves as a mediator between comfort and sportiness on the one hand and sound quality on the other hand. Thus, a thorough understanding of type conformity (and probably make conformity) is an important aspect for future research. Its dependency on comfort and sportiness (as already formulated in Bisping, 1997, Zeitler and Zeller, 2006) could be confirmed in this work. However, first inspections also showed that the variance explained by the linear approach for type conformity, depending on the interactions between car type and comfort as well as between car type and sportiness, was considerably lower in the lab than in the field (not reported in this work). Perhaps a non-linear approach better suits type conformity, as for example a “conformity map” like the depictions in Figures 1.2 and 1.3 that allows for arbitrary dependencies of type conformity on comfort and sportiness.

The findings of this work should be implemented in the vehicle development process to allow for purposeful sound design. In an early stage, meaningful target sounds can be specified with respect to the targeted customer using the revealed wishes and preferences of the driver types. Depending on the market positioning of the particular car model either comfort or sportiness (or perhaps both) can then be considered in more detail, and acoustic improvements can then selectively be made in the driving conditions that have the greatest impact on these aspects, namely static driving conditions for comfort and dynamic driving conditions for sportiness.

Furthermore, the development of sound metrics should be directed in order to establish the connection between the “objective” sound produced by the car, auditory sensations as modeled by psychoacoustic algorithms, and the psychological quantities which have been investigated in this work. In particular, working metrics for acoustic comfort and acoustic sportiness need to be established that consider the driving conditions relevant for the evaluation of both. This will be the next step in bridging the gap from the physical measurement of sound to its subjective evaluation in the context of product usage. The connection of the model proposed in this work with such sound metrics would constitute a more sophisticated tool than the ones currently used, helping sound engineers to achieve the target sound really desired by the customers. However, the implementation of such metrics must be reviewed critically as there is a crucial difference between field evaluation and lab evaluation. In field experiments, subjects have to simultaneously drive the car while evaluating the sound, which makes it more of a "driving context", in contrast to the lab as a pure "rating context". Which of these contexts should be described by a sound metric is a decision made in advance when the engineer chooses an experimental design for data collection. The type of training data for the development of the specific sound metric determines its scope. From a customer-oriented point of view, it is favorable to have a metric that reflects the field context as this is the context in which the user will experience the car.

Currently electromobility is gaining importance. For example, the Chinese government forces that by 2019 10 % (12 % by 2020) of the newly produced or imported vehicles are electric vehicles (Tian et al., 2017). Also many European countries encourage their citizens to buy electric vehicles by offering lower taxes or scrappage bonuses (Association des Constructeurs Européens d'Automobiles (ACEA), 2018). The model presented in this work was developed without regard to electrified power trains, and within this framework it is totally unclear how the sounds of electric cars will be perceived and evaluated. It will be interesting to see to what extent the model can also be applied to the sound of electric vehicles, or whether (and which) modifications will be necessary to adapt the model to this type of power train. In particular, some sound characteristics classically perceived as being "sporty" are not present in the sound of electric vehicles, such as roughness produced by engine orders (Oetjen, 2017, Maiberger et al., 2017). On the contrary, electric vehicle sounds typically contain tonal components generated by the inverter and magnetostriction in the electric motor. Nowadays, customers' expectations with regard to the sounds of electric vehicles are not clear as only few customers are experienced with electric vehicles (Schulte-Fortkamp et al., 2014). Design methodologies and concepts have to be further investigated (Küppers, 2013).

List of abbreviations

-2 LL	–	-2 log-likelihood
ANCOVA	–	analysis of covariance
ANOVA	–	analysis of variance
CI	–	confidence interval
df	–	degree(s) of freedom
EFA	–	exploratory factor analysis
EO	–	engine order
ESS	–	error sum of squares
GNAT	–	go/no-go association task
GUI	–	graphical user interface
KMO	–	KAISER-MEYER-OLKIN coefficient
MANOVA	–	multivariate analysis of variance
MSA	–	measure of sampling adequacy
MSI	–	method of successive intervals
NA	–	noise-sensitive aesthete
NVH	–	noise, vibration, harshness
rmMANOVA	–	repeated measure multivariate analysis of variance
SEM	–	structural equation model
SF	–	sound-oriented fun-driver
SII	–	speech intelligibility index
SP	–	sound-uninterested pragmatist
SUV	–	sports utility vehicle
SQ	–	sound quality
VIF	–	variance inflation factor
WOT	–	wide-open-throttle

References

P. Alenius, M. Olsson, and T. Lindbom. Engine Presence in Diesel Engine Passenger Cars. *SAE Technical Paper*, 2016-01-1786, 2016. doi:10.4271/2016-01-1786.

M. E. Altinsoy. A model for the sportiness perception of exterior vehicle engine start sounds. *International Journal of Vehicle Noise and Vibration*, 11(3/4):313–337, 2015. doi:10.1504/IJVNV.2015.075170.

M. E. Altinsoy, M. Ferling, and U. Jekosch. The Semantic Space of Vehicle Sounds — An Approach to Develop a Semantic Differential in View of Customer Perceptions. In *Proceedings of 3rd International Workshop on Perceptual Quality of Systems*, pages 47–51, 2010.

ANSI S3.5. Methods for Calculation of the Speech Intelligibility Index. ANSI standard, American National Standards Institute, Inc., March 1997.

Association des Constructeurs Européens d'Automobiles (ACEA). Overview on tax incentives for electric vehicles in the EU. ACEA reference document, April 2018. Retrieved from www.acea.be/uploads/publications/EV_incentives_overview_2018_v2.pdf on June 20, 2018.

M. Bednarzyk. *Qualitätsbeurteilung der Geräusche industrieller Produkte — Der Stand der Forschung, abgehandelt am Beispiel der Kfz-Innenraumgeräusche*. PhD thesis, VDI-Verlag, Düsseldorf, 1999.

C. V. Beidl and W. Stücklschwaiger. Application of the AVL-Annoyance Index for Engine Noise Quality Development. *Acta Acustica united with ACUSTICA*, 83(5):789–795, 1997.

R. Bisping. Psychologische und akustische Aspekte der Klangqualität von Fahrzeuggeräuschen. In Q.-H. Vo, editor, *Sound-Engineering — Kundenbezogene Akustikentwicklung in der Fahrzeugtechnik*, chapter 3, pages 34–47. Expert Verlag, Renningen-Malmsheim, 1994.

R. Bisping. Emotional effect of car interior sounds: Pleasantness and power and their relation to acoustic key features. *SAE Technical Paper*, 951284, 1995. doi:10.4271/951284.

R. Bisping. Car Interior Sound Quality: Experimental Analysis by Synthesis. *Acta Acustica united with ACUSTICA*, 83(5):813–818, 1997.

J. Blauert and U. Jekosch. Sound-Quality Evaluation — A Multi-Layered Problem. *Acta Acustica united with ACUSTICA*, 83(5):747–753, 1997.

K. A. Bollen. *Structural Equations with Latent Variables.* John Wiley & Sons, first edition, 1989. ISBN 0-471-01171-1.

J. Bortz and N. Döring. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Springer Medizin, Heidelberg, fourth edition, 2006. ISBN 978-3-540-33305-0. doi:10.1007/978-3-540-33306-7.

M. Busse. Psychoakustische und nichtakustische Dimensionen bei der Bewertung von Fahrzeuginnengeräuschen. Master's thesis, Carl von Ossietzky University, Oldenburg, 2014.

R. B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2): 245–276, 1966. doi:10.1207/s15327906mbr0102_10.

J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, April 1960. doi:10.1177/001316446002000104.

G. W. Diederich, S. J. Messick, and L. R. Tucker. A general least squares solution for successive intervals. *Psychometrika*, 22(2):159–173, June 1957. doi:10.1007/BF02289051.

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., March 2010.

DIN 45681 Berichtigung 2. Akustik — Bestimmung der Tonhaltigkeit von Geräuschen und Ermittlung eines Tonzuschlages für die Beurteilung von Geräuschimmissionen, Berichtigungen zu DIN 45681:2005-03, mit CD. DIN standard, Deutsches Institut für Normung e.V., August 2006.

DIN 45692. Messtechnische Simulation der Hörempfindung Schärfe. DIN standard, Deutsches Institut für Normung e.V., August 2009.

B. Ekehammar and S. Dornic. Weinstein's noise sensitivity scale: Reliability and construct validity. *Perceptual and Motor Skills*, 70(1):129–130, March 1990. doi:10.2466/pms.1990.70.1.129.

W. Ellermeier, M. Eigenstetter, and K. Zimmer. Psychoacoustic correlates of individual noise sensitivity. *The Journal of the Acoustical Society of America*, 109(4):1464–1473, April 2001. doi:10.1121/1.1350402.

K. Engel, B. Snitil, K. Wolff, R. Handel, J. Krüger, F. Gauterin, M. Helfer, and O. Brass. Wesentliche Geräuschquellen im Fahrzeug und deren Charakterisierung. In K. Genuit, editor, *Sound-Engineering im Automobilbereich — Methoden zur Messung und Auswertung von Geräuschen und Schwingungen*, chapter 6, pages 205–316. Springer, first edition, 2010. ISBN 978-3-642-01415-4. doi:10.1007/978-3-642-01415-4_6.

H. Fastl and E. Zwicker. *Psychoacoustics — Facts and Models*. Springer Series in Information Sciences. Springer, third edition, 2007. ISBN 978-3-540-68888-4. doi:10.1007/978-3-540-68888-4.

A. Fiebig. Influence of context effects on sound quality assessments. In *Proceedings of Euronoise*, pages 2555–2560, 2015.

A. Fiebig and F. Kamp. Development of metrics for characterizing product sound quality. In *Proceedings of Aachen Acoustics Colloquium 2015*, pages 123–134, 2015.

A. Field. *Discovering Statistics using SPSS (and sex and drugs and rock 'n' roll)*. Sage Publications, third edition, 2009. ISBN 978-1-84787-907-3.

W. H. Finch. Exploratory Factor Analysis. In T. Teo, editor, *Handbook of Quantitative Methods for Educational Research*, chapter 8, pages 167–186. Sense Publishers, first edition, 2013. ISBN 978-94-6209-404-8. doi:10.1007/978-94-6209-404-8.

D. Y. T. Fong, N. Takemura, P.-H. Chau, S. L. Y. Wan, and J. Y. H. Wong. Measurement properties of the chinese Weinstein noise sensitivity scale. *Noise Health [serial online]*, 19 (89):193–199, 2017. doi:10.4103/nah.NAH_110_16.

M. Frigge, D. C. Hoaglin, and B. Iglewicz. Some Implementations of the Boxplot. *The American Statistician*, 43(1):50–54, February 1989. doi:10.2307/2685173.

B. Fuhrmann and P. Garcia. Grundsatzuntersuchungen zum Sportsound durch Abgaskrümmer- und Vorrohranpassung. *MTZ*, 62(5):356–366, May 2001. doi:10.1007/BF03227336.

R. A. Gordon. *Regression Analysis for the Social Sciences*. Routledge, second edition, 2015. ISBN 113881251X.

C. Guastavino, B. F. G. Katz, J.-D. Polack, D. J. Levitin, and D. Dubois. Ecological Validity of Soundscape Reproduction. *Acta Acustica united with ACUSTICA*, 91(2):333–341, March/April 2005.

R. Guski. Psychological Methods for Evaluating Sound Quality and Assessing Acoustic Information. *Acta Acustica united with ACUSTICA*, 83(5):765–774, September/October 1997.

R. Guski. Personal and social variables as co-determinants of noise annoyance. *Noise Health [serial online]*, 1(3):45–56, 1999. PMID: 12689499.

R. Guski and J. Blauert. Psychoacoustics Without Psychology? In *Fortschritte der Akustik — NAG/DAGA 2009*, pages 1550–1551, 2009.

L. Guttman. Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2): 149–161, June 1954. doi:10.1007/BF02289162.

T. Hempel. *Untersuchung zur Korrelation auditiver und instrumenteller Messergebnisse für die Bewertung von Fahrzeuginnenraumgeräuschen als Grundlage eines Beitrags zur Klassifikation von Hörereignissen*. PhD thesis, Herbert Utz Verlag, München, 2001.

C. Ho, N. Reed, and C. Spence. Multisensory In-Car Warning Signals for Collision Avoidance. *Human Factors*, 49(6):1107–1114, December 2007. doi:10.1518/001872007X249965.

D. M. Hossain. Qualitative Research Process. *Postmodern Openings*, 2(7):143–156, September 2011.

IBM. SPSS Statistics for Windows, version 22, 2013.

R. Ingham, N. Otto, and T. McCollum. Sound Quality Metric for Diesel Engines. *SAE Technical Paper*, 1999-01-1819, 1999. doi:10.4271/1999-01-1819.

Y. Ishii, F. Brandl, and S. Thomann. Optimising the Interior Sound of Sports Cars. *ATZautotechnology*, 1(1):82–88, January 2001. doi:10.1007/BF03246584.

U. Jekosch. Meaning in the Context of Sound Quality Assessment. *Acta Acustica united with ACUSTICA*, 85(5):681–684, September/October 1999.

U. Jekosch and J. Blauert. Schritte zu einer vereinheitlichten Theorie der Qualität von Geräuschen. *Zeitschrift für Lärmbekämpfung*, 52(3):78–81, 2005.

R. F. S. Job. Noise sensitivity as a factor influencing human reaction to noise. *Noise Health [serial online]*, 1(3):57–68, 1999. PMID: 12689500.

O. Jung and V. Grützmacher. Analysis of Determining Parameters of Acoustical Comfort Inside Vehicles. *SAE Technical Paper*, 2011-01-1686, 2011. doi:10.4271/2011-01-1686.

H. F. Kaiser and J. Rice. Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34(1):111–117, April 1974. doi:10.1177/001316447403400115.

T. Keppens and K. van de Rostyne. Study for a combustion knocking metric for standard engine bench measurements. In *Proceedings of Aachen Acoustics Colloquium 2015*, pages 33–44, 2015.

W. Krebber. Interactive Vehicle Sound Simulation. In *Proceedings of Internoise*, pages 4056–4063, 2000.

C. Köber, D. Bönnen, C. Huth, and J. Bienert. Charakterisierung der Empfindungsgrößen Sportlichkeit und Komfort basierend auf der subjektiven Beurteilung von Fahrzeuggeräuschen. In *Fortschritte der Akustik — DAGA 2016*, pages 1202–1205, 2016.

T. Küppers. *Zielgeräuschentwicklung und Klangsynthesegestaltung im Innenraum von Elektrofahrzeugen*. PhD thesis, Schriftenreihe des Instituts für Kraftfahrzeuge (ika/fka), Aachen, 2013.

U. Letens. Exemplarische Darstellung einer subjektiven Beurteilung von PKW-Geräuschen. In *Fortschritte der Akustik — DAGA 2000*, pages 138–139, 2000.

U. Letens. Von der Psychoakustik über die psychologische Akustik zur ganzheitlichen Geräuschbewertung in der Fahrzeugakustik. In *Fortschritte der Akustik — DAGA 2010*, pages 433–434, 2010.

D. Maiberger, U. Letens, and S. van de Par. Development of a sportiness metric for vehicle sounds. In *Proceedings of Aachen Acoustics Colloquium 2017*, pages 163–174, 2017.

D. Maiberger, U. Letens, R. Weber, and 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 Acustica united with ACUSTICA*, 104(3):509–520, May/June 2018a. doi:10.3813/AAA.919193.

D. Maiberger, E. Strasser, U. Letens, and S. van de Par. Contextual aspects in subjective vehicle sound assessment. *Acta Acustica united with ACUSTICA*, 2018b. Currently under revision.

D. Maiberger, E. Strasser, U. Letens, and S. van de Par. Field versus lab: situational influences on vehicle sound assessment. *Acta Acustica united with ACUSTICA*, 2018c. Currently under revision.

MathWorks. MATLAB R2014a, 2014.

P. Mayring. Qualitative Inhaltsanalyse. In G. Mey and K. Mruck, editors, *Handbuch Qualitative Forschung in der Psychologie*, pages 601–613. VS Verlag für Sozialwissenschaften, Wiesbaden, first edition, 2010. ISBN 978-3-531-16726-8. doi:10.1007/978-3-531-92052-8_42.

R. R. McCrae and P. T. Costa Jr. Brief Versions of the NEO–PI–3. *Journal of Individual Differences*, 28(3):116–128, 2007. doi:10.1027/1614-0001.28.3.116.

R. R. McCrae, P. T. Costa Jr., and T. A. Martin. The NEO–PI–3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, 84(3):261–270, 2005. doi:10.1207/s15327752jpa8403_05.

D. Menzel, H. Fastl, R. Graf, and J. Hellbrück. Influence of vehicle color on loudness judgments. *The Journal of the Acoustical Society of America*, 123(5):2477–2479, May 2008. doi:10.1121/1.2890747.

Merriam-Webster. zeitgeist [Dictionary entry]. Website. Retrieved from www.merriam-webster.com/dictionary/zeitgeist on October 8, 2017.

H. M. E. Miedema and H. Vos. Noise sensitivity and reactions to noise and other environmental conditions. *The Journal of the Acoustical Society of America*, 113(3):1492–1504, March 2003. doi:10.1121/1.1547437.

A. Miśkiewicz and T. Letowski. Psychoacoustics in the Automotive Industry. *Acta Acustica united with ACUSTICA*, 85(5):646–649, September/October 1999.

T. Monahan and J. A. Fisher. Benefits of "Observer Effects": Lessons from the Field. *Qualitative Research*, 10(3):357–376, June 2010. doi:10.1177/1468794110362874.

L. C. Morey, R. K. Blashfield, and H. A. Skinner. A Comparison of Cluster Analysis Techniques Within a Sequential Validation Framework. *Multivariate Behavioral Research*, 18(3):309–329, 1983. doi:10.1207/s15327906mbr1803_4.

K. Mutalik, A. A. Gaikwad, N. V. Karanth, and S. Chivate. Sound Quality Rating of Passenger Car Diesel Powertrains. *SAE Technical Paper*, 2017-26-0189, 2017. doi:10.4271/2017-26-0189.

B. A. Nosek and M. R. Banaji. The go/no-go association task. *Social Cognition*, 19(6):625–664, December 2001. doi:10.1521/soco.19.6.625.20886.

G. O. Notbohm. *Personenspezifische Einstellungen und Wertungen in der Wahrnehmung der akustischen Umwelt*. PhD thesis, Shaker Verlag, Aachen, 2007.

G. O. Notbohm. A questionnaire on the individual attitude towards the acoustical environment. In *Proceedings of Internoise*, pages 5340–5349, 2010.

G. O. Notbohm, R. Schmook, S. Schwarze, and P. Angerer. Patterns of physiological and affective responses to vehicle pass-by noises. *Noise Health [serial online]*, 15(66):355–366, August 2013. doi:10.4103/1463-1741.116585.

K. Noumura and J. Yoshida. Perception Modeling and Quantification of Sound Quality in Cabin. *SAE Technical Paper*, 2003-01-1514, May 2003. doi:10.4271/2003-01-1514.

A. Oetjen. *Threshold and Suprathreshold Phenomena in Auditory Modulation Perception.* PhD thesis, Carl von Ossietzky University, Oldenburg, 2017.

A. Oetjen, S. van de Par, R. Weber, J. Verhey, J. Hots, and U. Letens. Verfahren zur Berechnung der Impulshaftigkeit von Motorgeräuschen. In *Fortschritte der Akustik — DAGA 2016*, 2016.

D. L. Paulhus. Measurement and control of response biases. In J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, editors, *Measures of personality and social psychological attitudes*, chapter 2, pages 17–59. Academic Press, San Diego, first edition, 1991. ISBN 9780080571102. doi:10.1016/B978-0-12-590241-0.50006-X.

B. Pfäfflin, H.-M. Gerhard, P. Ehinger, R. Herrmann, H. Bathelt, M. Scheinhardt, and J. Krüger. Geräuschgestaltung. In K. Genuit, editor, *Sound-Engineering im Automobilbereich — Methoden zur Messung und Auswertung von Geräuschen und Schwingungen*, chapter 9, pages 427–500. Springer, first edition, 2010. ISBN 978-3-642-01415-4. doi:10.1007/978-3-642-01415-4_9.

S. Pischinger, B. Lange, S. Heuer, and J. Hoppermanns. Objective evaluation of subjective noise impressions. *MTZ worldwide*, 67(11):28–31, November 2006. doi:10.1007/BF03227888.

B. Pletschen. Akustikgestaltung in der Fahrzeugentwicklung. In K. Genuit, editor, *Sound-Engineering im Automobilbereich — Methoden zur Messung und Auswertung von Geräuschen und Schwingungen*, chapter 3, pages 89–108. Springer, first edition, 2010. ISBN 978-3-642-01415-4. doi:10.1007/978-3-642-01415-4_3.

R. Porst. *Fragebogen — Ein Arbeitsbuch.* Studienskripten zur Soziologie. Springer VS, Wiesbaden, fourth edition, 2014. ISBN 978-3-658-02117-7. doi:10.1007/978-3-658-02118-4.

K. J. Preacher, D. D. Rucker, and A. F. Hayes. Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research*, 42(1):185–227, December 2007. doi:10.1080/00273170701341316.

B. Rammstedt, C. J. Kemper, M. C. Klein, C. Beierlein, and A. Kovaleva. A short scale for assessing the Big Five dimensions of personality. *methoden, daten, analysen*, 7(2):233–249, 2013. doi:10.12758/mda.2013.013.

B. D. Rapkin and D. A. Luke. Cluster analysis in community research: Epistemology and practice. *American Journal of Community Psychology*, 21(2):247–277, April 1993. doi:10.1007/BF00941623.

H. T. Reis and C. M. Judd. *Handbook of Research Methods in Social and Personality Psychology.* Cambridge University Press, second edition, 2014. ISBN 9780511996481. doi:10.1017/CBO9780511996481.

J. T. E. Richardson. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2):135–147, January 2011. doi:10.1016/j.edurev.2010.12.001.

B. Rohrmann. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9(3):222–245, 1978.

B. Rohrmann. Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data. Project report, University of Melbourne, Australia, January 2007. www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf.

L. D. Sanders and D. Poeppel. Local and global auditory processing: Behavioral and ERP evidence. *Neuropsychologia*, 45(6):1172–1186, 2007. doi:10.1016/j.neuropsychologia.2006.10.010.

B. Schulte-Fortkamp and K. Genuit. Exploration of Associated Imaginations on Sound Perception — A Subject-centered Method for Benchmarking of Vehicle. *SAE Technical Paper*, 2005-01-2263, 2005. doi:10.4271/2005-01-2263.

B. Schulte-Fortkamp, A. Fiebig, and K. Genuit. New Approach for the Development of Vehicle Target Sounds. In *Proceedings of Internoise*, pages 1302–1311, 2006.

B. Schulte-Fortkamp, A. Fiebig, M. Diechmann, and J. Kerkmann. E-mobility: Sound concepts and consumer reactions — Assessment and acceptance of driving sounds in the interior of electric vehicles. In *Proceedings of Aachen Acoustics Colloquium 2014*, pages 68–81, 2014.

A. Schulz and R. Fröming. Analysis of Driver Behaviour for Presentation of Adaptive Implementation Strategies of Assistance Systems. *ATZ worldwide*, 110(12):40–46, December 2008. doi:10.1007/BF03225052.

U. Schulz, H. Kerwien, and H. Gresch. Motivationale und emotionale Aspekte des Autofahrens. *Zeitschrift für Verkehrssicherheit*, 46(2):49–56, 2000.

M. Schütte, U. Müller, S. Sandrock, B. Griefahn, C. Lavandier, and B. Barbot. Perceived quality features of aircraft sounds: An analysis of the measurement characteristics of a newly created semantic differential. *Applied Acoustics*, 70(7):903–914, July 2009. doi:10.1016/j.apacoust.2009.01.004.

P. Sellerbeck. Eine innovative Methode zur Fahrzeugbewertung und Zielgeräuschdefinition. In *8. Magdeburger Symposium "Motor- und Aggregateakustik"*, pages 54–72, 2014.

V. P. Senese, F. Ruotolo, G. Ruggiero, and T. Iachini. The Italian Version of the Weinstein Noise Sensitivity Scale: Measurement Invariance Across Age, Sex, and Context. *European Journal of Psychological Assessment*, 28(2):118–124, 2012. doi:10.1027/1015-5759/a000099.

A. Sköld, D. Västfjäll, and M. Kleiner. Perceived Sound Character and Objective Properties of Powertrain Noise in Car Compartments. *Acta Acustica united with ACUSTICA*, 91(2): 349–355, March/April 2005.

J. Steffens. Realism and Ecological Validity of Sound Quality Experiments on Household Appliances. In *Fortschritte der Akustik — AIA-DAGA 2013*, pages 1473–1476, 2013.

S. S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, June 1946. doi:10.1126/science.103.2684.677.

E. Strasser, D. Maiberger, R. Weber, U. Letens, and S. van de Par. Untersuchung der Kontextvariablen bei der Bewertung von Fahrzeuggeräuschen. In *Fortschritte der Akustik — DAGA 2016*, pages 1210–1213, 2016.

H. Sukowski, R. M. Kühler, S. van de Par, and R. Weber. Perceived quality of the interior sounds in electric and conventional motor vehicles. In *Proceedings of Internoise*, pages 3523–3532, 2013.

P. Susini, O. Houix, N. Misdariis, B. Smith, and S. Langlois. Instruction's effect on semantic scale ratings of interior car sounds. *Applied Acoustics*, 70(3):389–403, March 2009. doi:10.1016/j.apacoust.2008.06.008.

O. Taubman-Ben-Ari, M. Mikulincer, and O. Gillath. The multidimensional driving style inventory — scale construct and validation. *Accident Analysis & Prevention*, 36(3):323–332, May 2004. doi:10.1016/S0001-4575(03)00010-1.

E. L. Thorndike. A constant error in psychological ratings. *Journal of Applied Psychology*, 4 (1):25–29, March 1920. doi:10.1037/h0071663.

Y. Tian, N. Sano, K. Buckland, A. Nussbaum, D. Welch, and E. Behrmann. China Gives Automakers More Time in World's Biggest EV Plan. Bloomberg online article, September 2017. Retrieved from www.bloomberg.com/news/articles/2017-09-28/china-to-start-new-energy-vehicle-production-quota-from-2019 on June 19, 2018.

N. H. Timm. *Applied Multivariate Analysis*. Springer, New York, first edition, 2002. ISBN 978-0-387-22771-9. doi:10.1007/b98963.

J. W. R. Twisk. *Applied multilevel analysis: a practical guide for medical researchers*. Cambridge University Press, 2006. ISBN 0521614988.

D. Västfjäll. Contextual Influences on Sound Quality Evaluation. *Acta Acustica united with ACUSTICA*, 90(6):1029–1036, November/December 2004.

D. Västfjäll, M.-A. Gulbol, M. Kleiner, and T. Gärling. Affective evaluations of and reactions to exterior and interior vehicle auditory quality. *Journal of Sound and Vibration*, 255(3): 501–518, August 2002. doi:10.1006/jsvi.2001.4166.

J. Wandmacher. *Einführung in die psychologische Methodenlehre*. Spektrum Akademischer Verlag, Heidelberg, Berlin, 2002. ISBN 978-3827413697.

J. H. Ward Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. doi:10.1080/01621459.1963.10500845.

N. D. Weinstein. Individual Differences in Reactions to Noise: A Longitudinal Study in a College Dormitory. *Journal of Applied Psychology*, 63(4):458–466, August 1978. doi:10.1037/0021-9010.63.4.458.

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

A. Zeitler, J. Hellbrück, W. Ellermeier, H. Fastl, G. Thoma, and P. Zeller. Methodological approaches to investigate the effects of meaning, expectations and context in listening experiments. In *Proceedings of Internoise*, pages 5621–5628, 2006.

K. Zimmer and W. Ellermeier. Eine deutsche Version der Lärmempfindlichkeitsskala von Weinstein. *Zeitschrift für Lärmbekämpfung*, 44(4):107–110, 1997.

Acknowledgements

I would like to thank the following people that contributed in different ways to this work:

Steven van de Par for encouraging me to do this thesis, for supervising my thesis, for always providing me with inspiring discussions and comments, and for his calm and relaxed way that is hard to find elsewhere.
Uwe Letens for the supervision of my thesis, for bringing me closer to vehicle acoustics, for all the discussions, practical suggestions and for letting me benefit from his experience, for introducing me to working in the industrial environment and, last but not least, for introducing me to the local jazz scene.
Ewald Strasser for all the valuable discussions, for the inspiring ideas, for the support in preparing and conducting the experiments, analyzing the data and endless proofreading, and for bringing me closer to psychology and statistics.
Reinhard Weber for encouraging me to do this thesis, for his support at the beginning of this work, for all the critical discussions and asking the right questions, and for letting me benefit from his scientific experience.
Ercan Altinsoy for his willingness to be the external referee for this thesis.
Volker Hohmann for agreeing to be a member of the examination committee.
Stephan Töpken for being a guide for proper scientific working, Arne Oetjen for introducing me to vehicle acoustics, and Helga Sukowski for providing the first insight into psychology, all three of which supported my work in various discussions, suggestions and feedback.
Wolfgang Kauke and the Daimler AG for offering me the possibility to work on this topic and for providing the infrastructure for doing the research that underlies this work.
Martin Tischler for helping me with the balancing act between requirements from both university and industry, for always providing practical advice, and together with Jan Hansmann and Linda Brandl for the substantial support in conducting the experiments.
Steffen Weber and Simon Stark for the inspiring discussions contributing to my thesis, Nicolas Pellé for providing me with the sound-tunings for the field study, and all further colleagues in my direct vicinity for creating a homelike climate and facilitating office work.
And, last but not least, Larissa Schmideder for helping me to organize myself, for her useful comments and practical support, and for the emotional assistance during all phases of this work.

Statement on own contributions

I hereby confirm, that David Maiberger contributed to the aforementioned studies as stated below:

Article:
David Maiberger, Uwe Letens, Reinhard Weber, Steven van de Par: *Individual influences on the evaluation of vehicle sounds: a typology of premium car drivers with regard to their attitude towards cars and sounds.* Acta Acustica united with ACUSTICA 104(3), 2018, 509–520. doi:10.3813/AAA.919193

Author Contributions:
David Maiberger formulated the research question, designed the study, carried out the experiments, analyzed and interpreted the data and wrote the final paper.
Uwe Letens formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.
Reinhard Weber formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.
Steven van de Par formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.

Article:
David Maiberger, Ewald Strasser, Uwe Letens, Steven van de Par: *Contextual aspects in subjective vehicle sound assessment.* Acta Acustica united with ACUSTICA, 2018, currently under revision.

Author Contributions:
David Maiberger formulated the research question, designed the study, carried out the experiments, analyzed and interpreted the data and wrote the final paper.
Ewald Strasser formulated the research question, participated in the study design, participated in carrying out the experiments, participated in data analysis and interpretation and participated in writing the final paper.
Uwe Letens formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.

Steven van de Par formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.

Article:
David Maiberger, Ewald Strasser, Uwe Letens, Steven van de Par: *Field versus lab: situational influences on vehicle sound assessment.* Acta Acustica united with ACUSTICA, 2018, currently under revision.

Author Contributions:
David Maiberger formulated the research question, designed the study, developed the technical framework for the experiments, analyzed and interpreted the data and wrote the final paper.
Ewald Strasser formulated the research question, participated in the study design, carried out the experiments, participated in data analysis and interpretation and participated in writing the final paper.
Uwe Letens formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.
Steven van de Par formulated the research question, participated in the study design, participated in data interpretation and participated in writing the final paper.

Date

Steven van de Par
Supervisor

I hereby confirm that I completed the work independently and used only the indicated resources. Further I confirm that I am aware of the guidelines of good scientific practice of the Carl von Ossietzky University Oldenburg and that I observed them.
Parts of this dissertation have already been published, as specified above. The contribution of the coauthors to the publications consisted in supervision of the work as well as corrections made to the manuscript. The dissertation has, neither as a whole nor in part, been submitted for assessment in a doctoral procedure at another university.

Date

David Maiberger