Sound Quality Label for E-Bikes

Andre Fiebig[1*], Andreas Herweg[2], Mirko Djukic[2]

[1] Technische Universität Berlin

[2] HEAD acoustics GmbH

[*] Corresponding author, andre.fiebig@tu-berlin.de

Abstract

The sound quality of products is becoming more and more important. This also applies for pedelecs, which are increasingly popular for older people but also for younger buyership as mobility behavior and environmental awareness change. As the electric engine of pedelecs emits audible noise and product noise belongs to one of the most important product features for purchase decisions, the information about the product sound quality is of high interest for potential buyers. Therefore, the paper presents an investigation of the noise from a variety of pedelecs. As e-mountainbikes (E-MTB) and e-city/e-trekking bikes (E-Trek) hold currently the largest market shares, 6 E-MTB and 8 E-Treks were measured on a test bench and were subject to extensive listening experiments to determine perception relevant noise characteristics. The jury tests provided comprehensive data to determine the link between signal properties and sound assessments. Based on the listening experiment data sound quality metrics were developed. It turned out that the sound quality of pedelecs can be well described by the psychoacoustic parameters loudness, sharpness, tonality and impulsiveness. Using the information about acceptable and unacceptable pedelec product sounds, a kind of sound quality label for providing product information about perceived sound quality to lay persons is proposed.

1 Introduction

1.1 Relevance of Pedelecs

Over the years, the number of sold e-bikes is increasing and the market share of newly sold e-bikes rises to 42% in Germany. Sales of e-bikes in Germany increased almost tenfold from 2010 to 2020, which now accounts for almost 9% of the total bicycle stock [3]. In 2020 alone, in Europe 5.1 million e-bikes were sold. Germany and the Netherlands continue to be by far the largest markets in Europe. More than 90% of the sold e-bikes belong to the model classes E-city bike, E-trekking-bike and E-mountainbike; the remaining less than 10 % split in e-cargo bikes, e-racing bicycle, s-bicycles or others [1]. In 2022, more than 10 million people own an e-bike in their household [4].

1.2 Use of pedelecs and customer requirements

According to [1] the usage behavior between a conventional bike without an electric engine and pedelecs is similar. In both cases the usage for leisure activities is dominant. However, it has to be mentioned the use is dependent on age; for older groups (older than 50 years) the covered distance is significantly larger with pedelecs than with standard bicycles [1]. It seems that the speed of e-bike driver is slightly higher than for conventional bicycles. Twisk et al. observed in average 2.3 km/h faster speeds in urban context and around 4 km/h higher speeds outside of cities [6]). The observed average speeds were slightly over 20 km/h. Friedrich summarized the range of speed and power for untrained cyclist in average between 15 and 20 km/h and 50 to 100 Watt [9].
According to Bourne [7] in his literature review the most indicated benefits of pedelecs are the ability to ride longer distances, faster journeys, ability to ride hilly terrain and new routes with new destinations, time saving, ability to carrier heavy goods and lower environmental impact. Although one of the benefits on the individual level is the reduced perspiration in comparison to conventional cycling due to the support from the electric engine and at the same time as an disadvantage that less physical efforts, studies indicate that the physical activity is similar between pedelecs and conventional cycling due to the faster driving, longer distances and driving times. Thus, Stenner et al. [8] observed that due to a sufficient heart rate increase in e-bikes, pedelecs even offer a more active form of transportation to enhance physical activity. Based on the literature four typical use cases were derived which guide the acoustic measurements and the listening experiments (table 1).

Table 1: Defined use cases for acoustic measurements and sound quality assessment

Use case	**Power in W**	**Pedaling speed in c/min**	**Bike speed in km/h**
relaxed drive	45	40	15
business/errand drive	85	60	20
leisure activity drive	210	80	15
sports drive	375	80 and 100	15 and 20

The defined use cases cover a broad range of power, similar to the power setting used in ([11]) In Becker et al. were set as power levels of 100, 200 and 250 Watt to determine noise emissions and sound power levels of different e-bikes [11]. According to Friedrich [9] for non-professional drivers the performance is usually significantly less than 300 W, whereas professional athletes even achieve a performance of over 400 W. There are also differences between professional and hobby riders when it comes to cycling cadence, the pedaling rate as a measure of angular speed. In performance-oriented cycling, high cadences of 100 to 110 rpm are ridden, whereas sporty hobby cyclists often cadence of 80 to 90 rpm. During recreational rides, the cycling cadence is usually much lower, around 60 rpm according to [9] Thus, the defined use cases include most of the typical conditions where bicycles and pedelecs are used.

2 Acoustic measurements

14 pedelecs were measured on a test bench which was specifically built-up for the investigation. The test bench utilized an indoor cycling trainer (Garmin, Tacx Flux) which was combined with pedal sensors (Favero, Speed 7 Cadence). The cycling trainer was acoustically treated to minimize its acoustic influence. Data from cycling trainer and pedal sensors was collected via a cycling computer (Garmin, Edge) and then merged and synchronized with the remaining data via HDF API (HEAD acoustics). Binaural signals were recorded at driver position (HEAD acoustics, BHS) and at bystander position (HMS IV, HEAD acoustics GmbH). Mono signals were recorded at both sides of the electrical engine.

Figure 1: Selection of measured pedelecs (E-Trekking bikes and E-Mountainbikes)

For the subsequent listening experiments, recordings were edited and selected to consider steady-state conditions only with a constant support of the electric engine for at least 10 s. Moreover, recordings which exhibit an clearly audible share of disturbing noise patterns from the chain and the gear rings, were excluded for the listening tests.

3 Listening experiments

3.1 Participants

In total, 37 participants (26 male and 11 female) took part in the experiments. Therefrom, 22 took part in Herzogenrath and 15 participants carried out the listening experiment in Berlin. The majority of the participants belong to the age class ranging from 30 to 49. 24% of the participants were younger than 30 years and 11%

were 50 years or older. The majority of the participants were *moderately* to *rather* sensitive to product sound and showed a *rather* to *very high* level of environmental awareness. 17% of the participants owned a e-bike and almost halve of the group of participants indicated to could imagine to buy a pedelec within the next few years. Consequently, 51% of the participants declared as the reason for refraining from buying a pedelec the relatively high price.

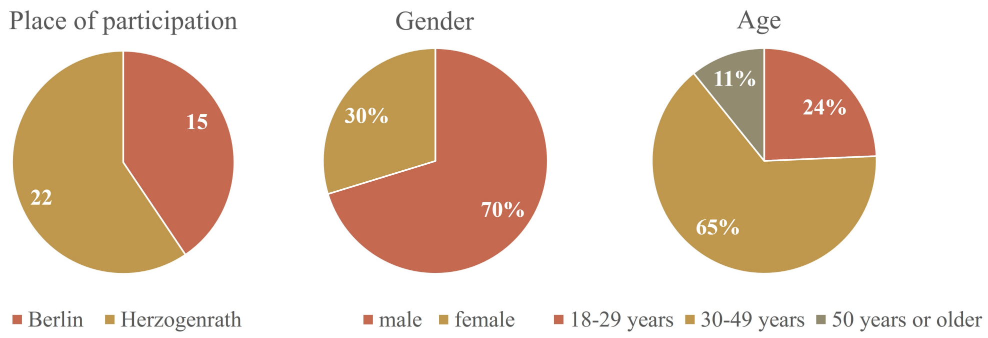


Figure 2: Distribution of participants, gender and age

3.2 Stimuli

All binaural sounds were edited to a lenght of 10s. Only stead-state signals were chosen for the listening experiments. In total, 75 different sound stimuli were presented, 36 E-MTB and 39 E-Trek recordings. For each of the defined conditions,. i.e., relaxed drive, business/errand drive, leisure activity drive, sports drive, for the model classes E-MTB and E-Trek 7 to 12 signals were selected.

3.3 Apparatus

The sounds were presented via digital equalizers (labP2, HEAD acoustics GmbH) and Sennheiser headphones HD 650 in an acoustically treated listening studio at the premises of the HEAD acoustics GmbH in Herzogenrath, Germany as well as in a quiet office room at the TU Berlin. The playback system was calibrated and equalized. All instruction texts and interview questions were presented on computer screens via the software SQuala (HEAD acoustics GmbH). The software controlled the run of entire experiment. The sounds were played back with free-field equalization. The same software and hardware components were applied at both sites.

3.4 Procedure

The participants were informed about the aim of the study, the investigation of sound quality of pedelecs. The experiment started with an instruction: "In the following you will listen to pedelec sounds with a duration of 10 s. Please judge the overall sound quality after listening to the entire sound on an 10-point rating scale, which ranges from very poor to excellent." The 10pt rating scale with verbal labels at each category. Before first ratings were requested the participants had to listening to several sounds first in order to get familiar with the different e-bike sounds and to reduce scaling effects, such as floor and ceiling effects. All sets were randomized for each participant and the sounds within each set were randomized as well. During the playback of a sound, the input was blocked in order to force the participant to listen to the full sound stimulus before providing an assessment. It was not possible to repeat the playback of a sound stimulus. For some sets the semantic differential methods was additionally used and participants had to judge the sound stimuli on the 7-pt. rating scales regarding *loud-quiet, pleasant-disturbing, steady-unsteady, powerful-powerless, high-low, high quality-low quality, tonal-not tonal, and weak-strong.* For this evaluation task it was possible to repeat once the playback of a sound stimulus.

The participants could take as much time as needed to rate the stimulus; no time limit was given. After providing an assessment, the next sound stimulus was automatically played back. The duration of the entire experiment including the questionnaires which had to be filled out after the assessment of all sounds, was around 50 to 60 minutes. After the first part of the experiments the participants had to take a break of 5 minutes. The participants could take additionally a break at any time. After the judgment of all sound stimuli, the participants filled out questionnaires to provide information about the person, bike use, sound sensitivity and environmental awareness.

3.5 Results

The data from the two sub-groups (Berlin and Herzogenrath) were compared in order to detect potential differences relevant for further analyses. The comparison of the sound assessments as well as the interview data did not yield any systematic differences between the subgroups. Therefore, in the following the data from both groups were merged and analyzed.

3.5.1 Interview data

As figure 3) (left) shows the participants acknowledge sound and comfort aspect as one of the relevant product features when thinking about purchasing an e-bike. When asked about the most unwanted noise properties of pedelecs the majority of participants indicated loud, unsteady, sharp, and tonal noise is critical. Consequently, as preferred sound characteristics mostly low frequency content, quiet and steady were mentioned. Moreover, the term powerful (*kraftvoll*) was frequently mentioned as well. Only rarely the participants referred to a recognition of the electric engine and its mode due to the sound.

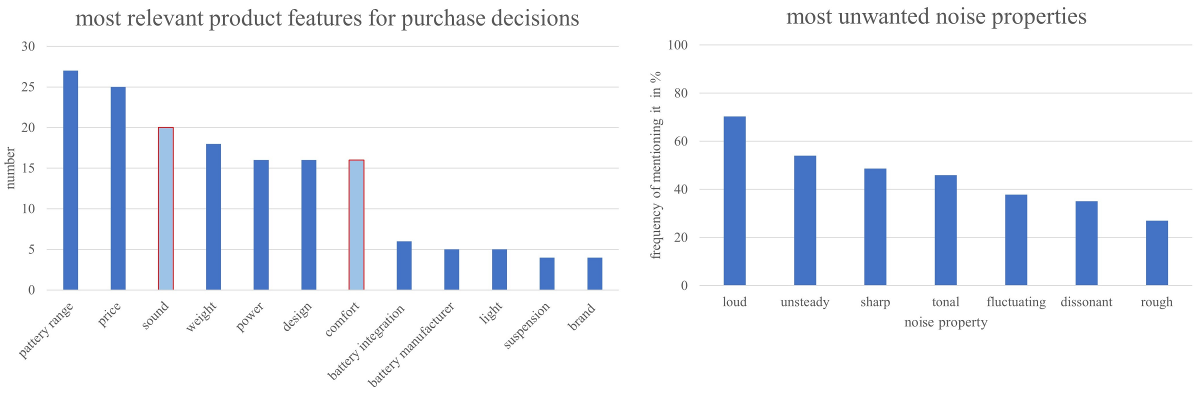


Figure 3: Total number of selected most relevant product features for purchase decisions (left) and selected most unwanted noise properties (frequency of occurrence in %). Multi-selection was possible.

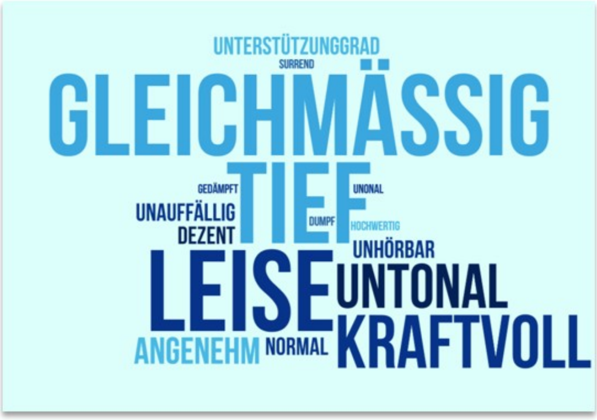


Figure 4: Wordcloud of most preferred noise characteristics of pedelecs (German)

The gained information from the interviews regarding the desired sound characteristics and the noise patterns which likely adversely influence sound quality can guide the selection of acoustic indicators for the prediction of sound quality as shown in chapter 4

3.5.2 Sound assessment data

Figure 5) shows that the pedelecs were judged differently depending on the condition, whereas the model class did not play a significant role. In average, E-Trekking and E-Mountainbike were judged similar, though the E-Treks were judged slightly better than the E-MTB.
Figure 6) illustrates the result of a cluster analysis of the data from the semantic differential test. It can be seen that the more positive items are separated from the more or less negative items. This results provides further

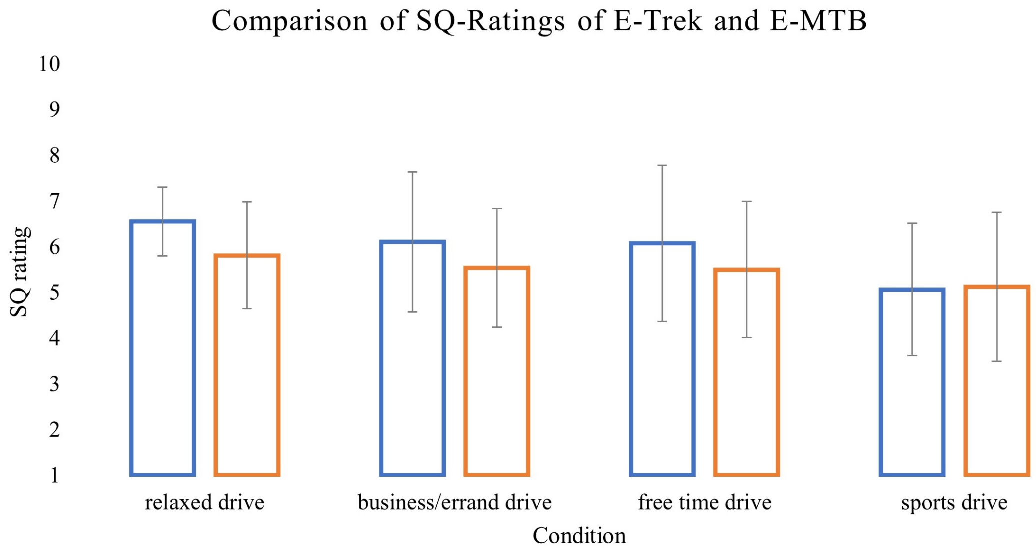


Figure 5: Comparison of sound quality ratings of E-Trek (blue bars) and E-MTB sounds (orange bars) over conditions. Average means and standard deviations are displayed.

information about relevant psychoacoustic parameter to predict sound quality. The signals should possess a steady-state character, should not be loud, tonal or sharp (i.e., high).

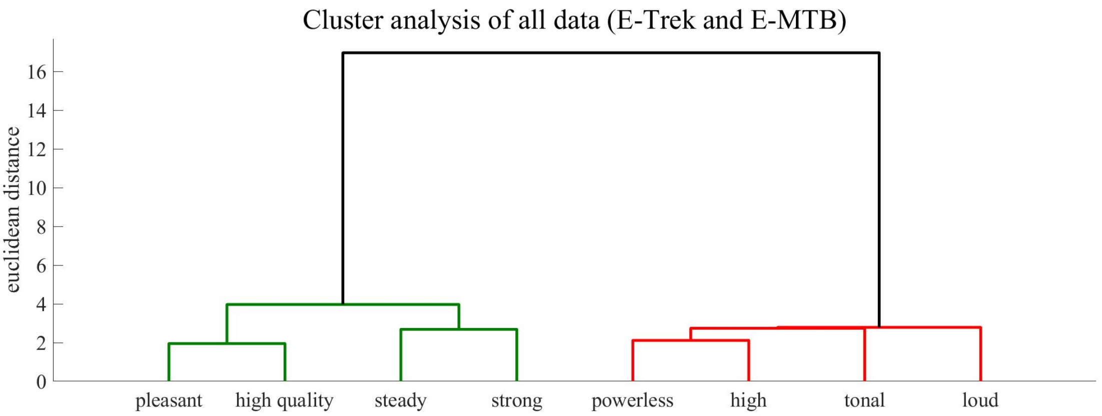


Figure 6: Dendrogram based on the cluster analysis (Ward method) of semantic differential data. Data from E-MTB and E-Trek sounds was combined before analyzed.

4 Development of sound quality metrics

In order to develop meaningful prediction models a variety of correlations between sound quality ratings and acoustic indicators were calculated. In addition, correlations of potential suppressor variables with the prediction error were determined.

Later, a non-exhaustive cross validation method was used to investigate the robustness of models more in detail. In particular, 3 cross validation folds with 20 repetitions were performed, i.e., 60 combinations of a potential metric with training and validation were performed to study the performance of indicators when the indicators are repeatedly applied to unknown data. The use of cross validation methods is frequently used in the context of the development of sound quality metrics ([5])

However, the prediction quality is not sufficient for individual conditions based on loudness alone, because the cross-validation shows considerable uncertainty for some conditions (relaxed drive (E-Trek), business/errands drive (E-MTB) and leisure activity drive (E-MTB)). At least for those conditions further (psycho)-acoustic parameters must be taken into account.

It turned out that the most relevant predictor is loudness according to the ISO 532-1 [10]. This parameter showed in most cases over pedelec model and condition, the highest correlations with the sound quality ratings (figure 7) even consistently better than typical level indicators. This is particular true when data sets of E-MTB and E-Trek were merged together to consider more data for correlation analyses as shown in figure 8.

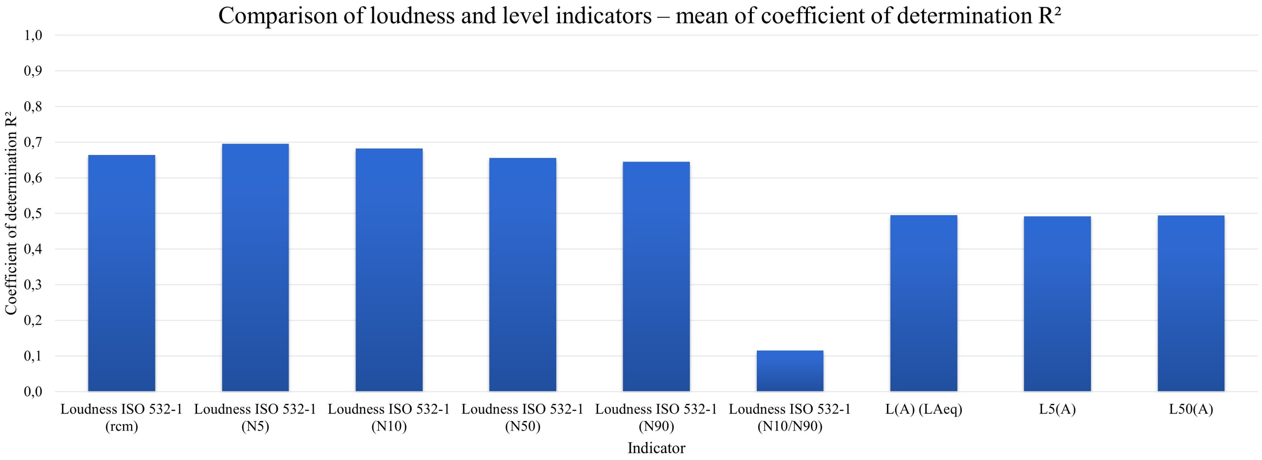


Figure 7: Coefficient of determination of loudness related indicators with respect to sound quality ratings of pedelecs. The coefficients of determination were determined for each condition and averaged over all conditions. Loudness indicators based on ISO 532-1 [10] and sound level indicators are shown.

By combining the most important predictor loudness (according to the ISO 532-1) with the strongest suppressor for each conditions, very good prediction results of the sound quality judgments were achieved. However, the prediction quality appear to be partly insufficient for individual conditions, as the cross-validation shows considerable uncertainties. Due to the small number of signals per condition, the risk of *overfitting* is likely.

Table 2: Prediction of SQ assessment by means of loudness N_5 and suppresor variable with highest correlation to residual error

E-Trek	**correlation r to N_5**	**supressor variable with highest correlation with residual error**	**correlation r of metric with N5 and supressor variable**	**coefficient of determination R^2 (averaged) for validation data**
relaxed drive	-0.43	Impulsiveness	0.91	< 0.5
business/errand drive	-0.93	Impulsiveness	0.96	< 0.5
leisure activity drive	-0.96	Relative Approach T	0.99	0.88
sports drive	-0.88	Relative Approach T	0.95	0.66
E-MTB				
relaxed drive	-0.90	Tonality	0.93	< 0.5
business/errand drive	-0.73	Roughness	0.96	0.67
leisure activity drive	-0.81	Tonality	0.94	0.56
sports drive	-0.92	Impulsiveness	0.95	0.79

However, as the correlations on the parameter of loudness were already high, it was clear that for only few signals, adding a suppressor variable can result in lower prediction accuracy for unknown data (i.e., validation data using a cross validation method). This problem obviously occurs for some conditions as table 2 displays. However, the suppressor variables illustrate already what kind of sound characteristics might be relevant for assessing the sound quality of pedelecs beyond the parameter loudness. As expected, tones (tonality), transient components (impulsiveness) or time patterns (Relative Approach T, roughness) can influence sound quality ratings as well. The soun d quality obviously depends on both the loudness of the product sound and the sound character. However, due to the low number of sounds per conditions, the relevance and weighting of those parameters is rather difficult to assess. Thus, merging the data together allows to develop more stable metrics for the prediction of sound quality ratings. Table 3 shows that the importance of the predictor loudness N_5 remains high when the data from E-Trek and E-MTB are combined for the respective driving scenario. Moreover, as table 3 indicates the prediction accuracy is far from perfect and the consideration of further parameters is reasonable.

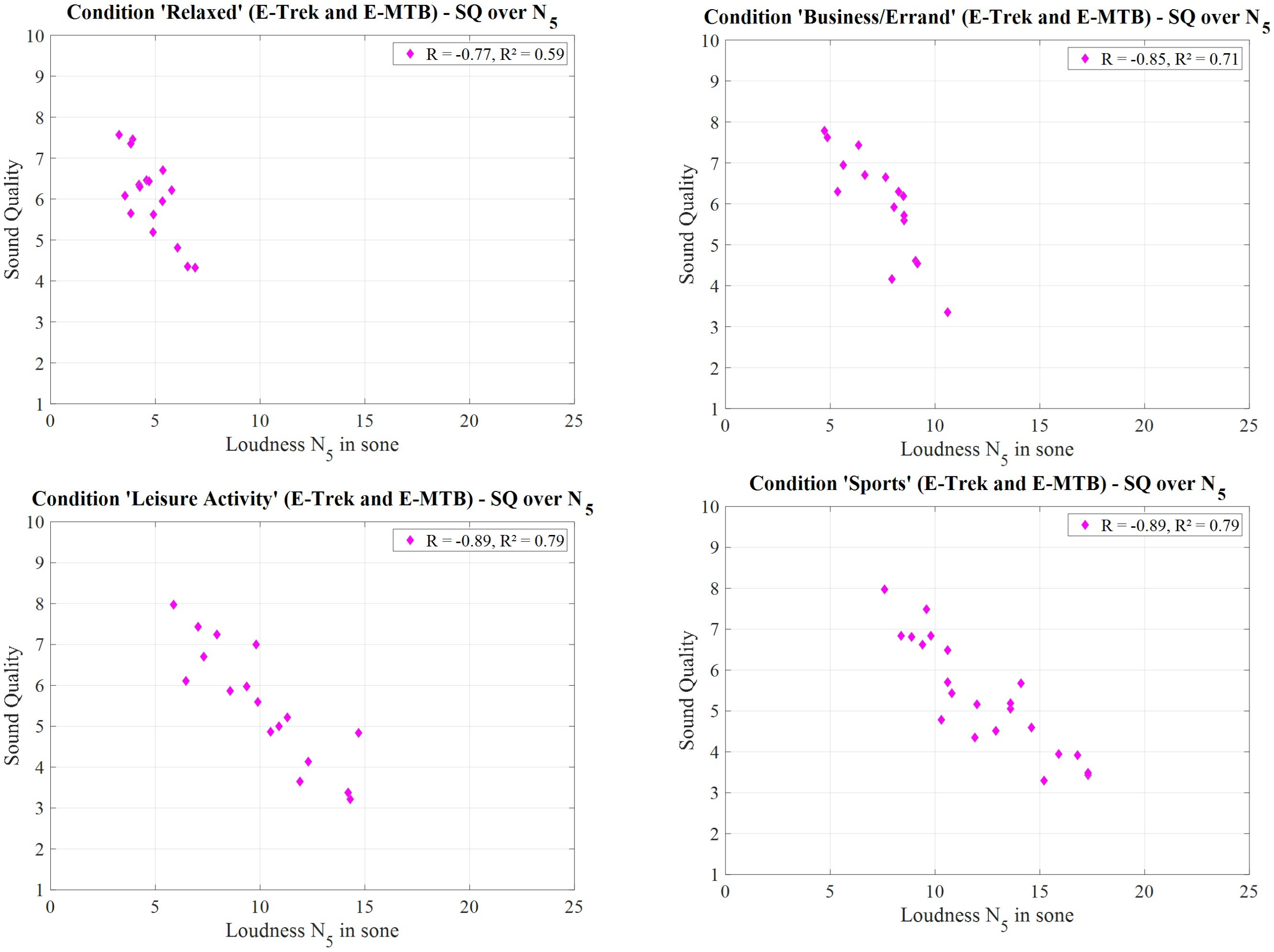


Figure 8: Link between sound quality ratings and loudness N5 according to ISO 532-1 ([10]) over conditions with combined data from E-Trek and E-MTB sounds. From top left to bottom right: Relaxed drive, business/errand drive, leisure activity, sports drive. Average means values of SQ ratings are shown.

Table 3: Prediction of SQ assessment by means of loudness N_5 for conditions including both, E-MTB and E-Trek data

Condition	**model/validation**	**coefficient of determination R^2**	**RMSE**	**MAE**
relaxed drive	model	0.59	0.61	0.53
	validation	0.29	0.68	0.58
business/errand drive	model	0.71	0.66	0.53
	validation	0.55	0.70	0.59
leisure activity drive	model	0.79	0.67	0.52
	validation	0.64	0.78	0.61
sports drive	model	0.79	0.62	0.48
	validation	0.73	0.67	0.54

Similar predictors were always used to develop the metrics, which allow the best possible approximation to the evaluation data. It turned out that the use of two additional parameter yielded the best results with respect to the robustness measured by cross validation methods.
Table 4 displays the results of the iterative metric development by selecting parameters which yielded an improvement of performance regarding the prediction quality of validation data. The parameters chosen for predicting the quality of the pedelec sounds are always loudness, impulsiveness, tonality and sharpness. Figure 9 demonstrates that the predictor loudness is always most important and that the other parameters contributing to the prediction to a lower extent. However, the performance of the sound quality metric with these parameters for the training as well as validation data was better compared to the predictor loudness alone (see table 3). Therefore, the developed sound quality metrics allow predicting the sound quality of pedelecs with an absolute prediction error around 0.3 regarding the 10 pt rating scale for unknown data.

Table 4: Prediction of SQ assessment by means of loudness N_5 for conditions including E-MTB and E-Trek data

Condition	**parameter included in metric**	**model/validation**	**coefficient of determination R^2**	**MAE**
relaxed drive	loudness, impulsiveness, tonality	model	0.87	0.26
		validation	0.65	0.37
business/errand drive	loudness, impulsiveness, sharpness	model	0.92	0.26
		validation	0.78	0.39
leisure activity drive	loudness, impulsiveness, tonality	model	0.92	0.37
		validation	0.75	0.54
sports drive	loudness, impulsiveness, sharpness	model	0.92	0.26
		validation	0.88	0.32

In addition to loudness, sharpness and tonality have a negative effect on the sound quality. The great advantage of the consistent use of these parameters is the reference to standardized calculation rules (ISO 532-1 [10], DIN 45692 [12], ECMA 74 [13]). The parameter impulsiveness can be used to additionally quantify disturbing noise patterns (crackling), which also has a negative effect on the sound quality. As a note, taking into account the cycling cadence, the prediction quality can be further increased, which is particularly relevant when data sets over conditions are merged.

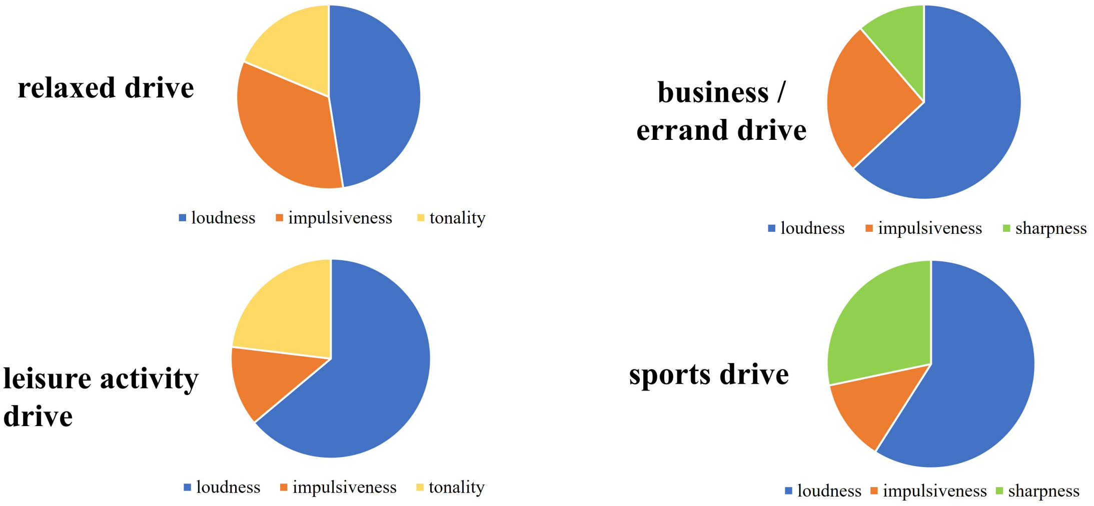
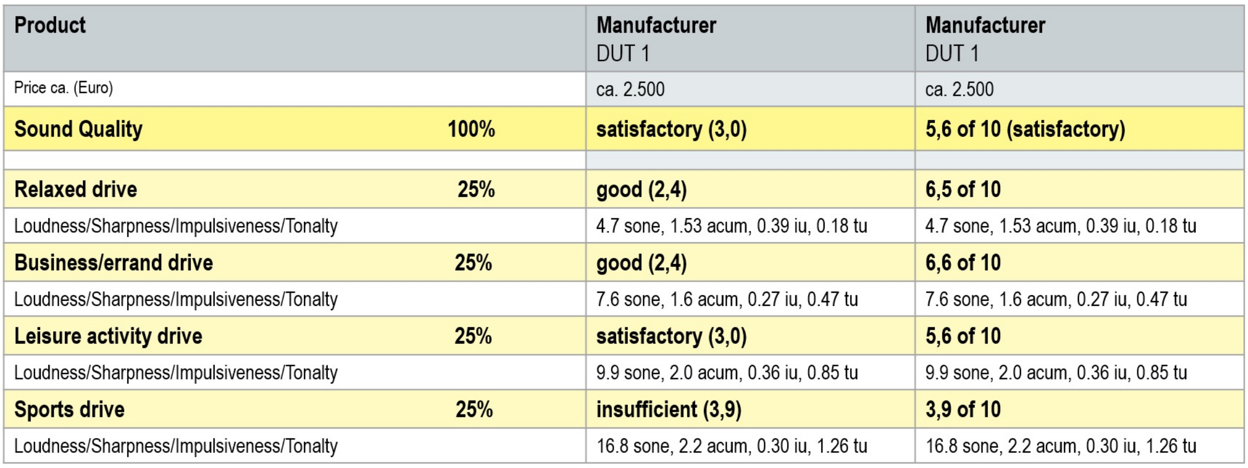


Figure 9: Weights (expressed in size of angle) of the SQ metric parameters

5 Proposal for a sound quality label

A transparent and easily understandable representation of the acoustic evaluation of an e-bike is desirable and would increase the acceptance of a label for acoustic comfort/sound quality. An orientation towards the product evaluation procedure at Stiftung Warentest can lead to increased comprehensibility and acceptance of an newly introduced product sound evaluation. Moreover, the reason for low or high product ratings needs to be understandable and plausible in order to be accepted by manufacturer, traders as well as customers.



Product		**Manufacturer** DUT 1	**Manufacturer** DUT 1
Price ca. (Euro)		ca. 2.500	ca. 2.500
Sound Quality	**100%**	**satisfactory (3,0)**	**5,6 of 10 (satisfactory)**
Relaxed drive	**25%**	**good (2,4)**	**6,5 of 10**
Loudness/Sharpness/Impulsiveness/Tonalty		4.7 sone, 1.53 acum, 0.39 iu, 0.18 tu	4.7 sone, 1.53 acum, 0.39 iu, 0.18 tu
Business/errand drive	**25%**	**good (2,4)**	**6,6 of 10**
Loudness/Sharpness/Impulsiveness/Tonalty		7.6 sone, 1.6 acum, 0.27 iu, 0.47 tu	7.6 sone, 1.6 acum, 0.27 iu, 0.47 tu
Leisure activity drive	**25%**	**satisfactory (3,0)**	**5,6 of 10**
Loudness/Sharpness/Impulsiveness/Tonalty		9.9 sone, 2.0 acum, 0.36 iu, 0.85 tu	9.9 sone, 2.0 acum, 0.36 iu, 0.85 tu
Sports drive	**25%**	**insufficient (3,9)**	**3,9 of 10**
Loudness/Sharpness/Impulsiveness/Tonalty		16.8 sone, 2.2 acum, 0.30 iu, 1.26 tu	16.8 sone, 2.2 acum, 0.30 iu, 1.26 tu

Figure 10: Exemplary SQ assessment tables resembling the typical display from Stiftung Warentest. Left: Category including weighting, middle: Use of school grades, right: Use of 10-pt rating scale

A reference to the German school grading system makes is easy to understand for German-speaking countries, but appears unfavorable in an international context due to different school grading systems. It may be more appropriate to refer to a certain range, such as 1 to 10, as the range of values from poor to excellent. By referring to relevant psychoacoustic variables (including the values), the reason for complaints and poor SQ ratings becomes transparent and understandable. Figure 10 shows a comparison of school grades with a point systems using a range from 1 to 10. The design of the table is exemplary adapted to the way Stiftung Warentest presents product testing results.
All in all, the inclusion of the psychoacoustic measured values when evaluating product sound quality and the integration into a kind of label for informing customers bear some advantages. Displaying the determined psychoacoustic values, on the one hand, would lead to an increasing spread and acceptance of the proposed parameters. And, on the other hand, the measured acoustic values are directly comparable and can be used later for comparison purposes, e.i. for customers and manufacturers. Moreover, manufacturers have then specific reference values for their product noise optimization. A reference to the well-known EU energy label, as exemplary shown in figure 11, would be easy to understand and to interpret for the public.

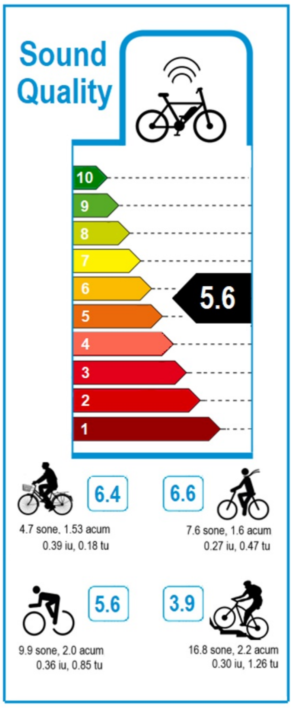


Figure 11: Potential display adapted to the design of the EU Energy label of the sound quality of a pedelec

6 Summary

The sound quality of pedelecs is more and more important as the market becomes competitive, indicated by the increased sales numbers. Pedelecs are increasingly popular for older people but also for younger persons as well and in this context product sound counts to one of the most important product features for purchase decisions. Thus, the information about the product sound quality is of interest for potential buyers. Therefore, the sound quality of different pedelecs were investigated. The considered E-MTB and E-Treks bikes were measured on a test bench and subject to extensive listening experiments. For it, a standardized procedure in collecting judgments about the sound quaylity of the pedelecs sounds under controlled conditions were developed and applied. Based on the jury experiments with 37 participants it was possible to determine perception relevant noise characteristics and to develop a kind of sound quality label for providing product information to lay persons.

It turned out that the psychoacoustic parameter loudness is most important for the assessment of pedelecs noise with respect to all driving conditions. Further psychoacoustic parameters, such as sharpness and tonality have a negative effect on the sound quality as well as those properties attract unwanted attention. Moreover, disturbing noise patterns lead to a decrease of the product sound quality, which can be captured by the parameter impulsiveness.

Based on these parameters the collected judgmental data can be reasonably well predicted. The information about the sound quality of pedelecs needs to be conveyed to lay-persons in a way easy to understand. The use of established labels like the wide-spread EU energy label will facilitate communication about and the use of the sound quality of pedelecs.

7 Acknowledgment

We would like to take this opportunity to thank the HEAD-Genuit-Foundation for supporting the research.

References

[1] E. Brust: Zahlen - Daten - Fakten zum Fahrradmarkt in Deutschland 2020. Zweirad-Industrie-Verband, Berlin, 2021.

[2] C. Nobis: Mobilität in Deutschland. MiD Analysen zum Radverkehr und Fußverkehr. Studie von infas, DLR, IVT und infas 360 im Auftrag des Bundesministeriums für Verkehr und digitale Infrastruktur, FE-Nr. 70.904/15, Bonn, Berlin, 2019.

[3] ZIV: Zahlen-Daten-Fakten zum Fahrradmarkt in Deutschland 2021, Zweirad-Industrie-Verband, Berlin, 2022.

[4] AWA 2022: Allensbacher Markt- und Werbeträger-Analyse - AWA 2022. IfD Allensbach, Allensbach am Bodensee, 2022.

[5] F. Kamp, A. Fiebig: Entwicklung von Störgeräuschmetriken unter Anwendung eines Verfahrens zur Robustheitsanalyse, Magdeburger Symposium, Tagungsband, Magdeburg, 2016

[6] D. Twisk, A. Stelling, P. Van Gent, et al.: Speed characteristics of speed pedelecs, pedelecs and conventional bicycles in naturalistic urban and rural traffic conditions, Accident; analysis and prevention, DOI:10.1016/j.aap.2020.105940, 2020.

[7] J.C. Bourne. The impact of e-cycling on travel behaviour: A scoping review. J. Transp Health. 1, 2020

[8] H.T. Stenner, J. Boyen, M. Hein. Everyday pedelec use and its effect on meeting physical activity guidelines, Int. J. Environ. Res. Public Health 2020,17, 4807; doi:10.3390/ijerph17134807, 2020.

[9] W. Friedrich. Optimales Sportwissen. Grundlagen der Sporttheorie und Sportpraxis. Spitta, 2016.

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

[11] S. Becker. B. Berchtenbreiter, F. Krömer, A. Lodermeyer, J. Riedel, A. Renz. Akustik von E-Bikes. Akustik Journal 02, 41-49,2020.

[12] DIN 45692. Measurement technique for the simulation of the auditory sensation of sharpness, Beuth Verlag, Berlin, 2009.

[13] ECMA 74. Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment. ECMA International, 2021.