inter.noise
27-30 AUGUST
HONG KONG 2017

Artificial neural network based detectability prediction of synthesized exterior electric vehicle sounds

Lisa Steinbach[1]; Serkan Atamer[2]; M. Ercan Altinsoy[3]
[1,2,3] Institute of Acoustics and Speech Communication, Chair of Acoustic and Haptic Engineering, TU Dresden

ABSTRACT

The safety for traffic participants like pedestrians and cyclists might be effected by the current electrification of vehicles. Detectability of electric vehicle sounds is an important attribute for safety reasons. The aim of this work is to determine the detectability of different electric vehicle sounds for a constant speed, single car pass-by situation. For this purpose, the differences in detection time are investigated with perception studies. The correlation between physical-psychoacoustical parameters and detection time estimations obtained from jury testing is also investigated in this study. Moreover, an artificial neural network (ANN) is also used as a prediction tool of detectability estimations for further evaluations of different possible stimuli. Lastly, advantages and shortcomings of using ANNs for detectability estimations are also discussed.

Keywords: Psychoacoustics, sound quality, detectability, neural networks, safety, electric vehicle

1. INTRODUCTION

In today's urban environment, the inhabitants are permanently exposed by increased noise levels, which are mostly dominated by traffic noise. For years, traffic noise has been one of the most prominent noise sources in urban areas. Recently, thanks to the process of electrification of vehicles, lower noise levels are expected in the city centers in the near future.

However, decreasing noise levels has also a disadvantage. Cars being quieter than before makes them harder to be detected. Especially for the lower driving speeds, where the tire noise component is no longer available, accidents become almost unavoidable. For that reason, car manufacturers are implementing external sound emission systems in electric cars. Regulations are also prepared accordingly, for example, all new electric vehicles in Germany should emit an external sound from 2019. (1)

The external sounds should be designed regarding to the perceptions of the listeners, i.e. pedestrians and residents. The goal of making electric vehicle sounds is both detectability as well as interpretability while keeping the noise levels under allowable limits.

In this study, the detection time of the synthetic vehicle sounds in a background noise is investigated. In order to have a safe traffic situation, the vehicle noise should be detectable before a critical distance where the remaining time is enough to avoid an accident. For this purpose, reaction times were measured in listening tests, with various synthetic external noises within the scope of that study.

2. STIMULI

Within this study 9 different stimuli of electric vehicle sounds were created. For all those 9 stimuli, two different levels were used. The subjects evaluated a total of 18 stimuli. All of them are synthetic sounds. The stimuli we used in this study are general basic vehicle sounds, which can be adapted by the manufacturer to the respective requirements in the sense of a brand design.

Furthermore, a realistic background noise was used for masking. To create a realistic background noise (TU-Dresden Background) acoustic recordings with a dummy head was carried out at different

[1] lisa.steinbach@tu-dresden.de
[2] serkan.atamer@tu-dresden.de
[3] ercan.altinsoy@tu-dresden.de

locations in the city center of Dresden. After listening by experts from the chair, a recording was chosen for the use of the experiment. The homogeneity of the noise was the main criterion, since the dominance of individual events should be avoided. Figure 1 shows the spectrogram of the TU-Dresden Background noise and in Figure 2, some example spectrograms of the stimuli are given.

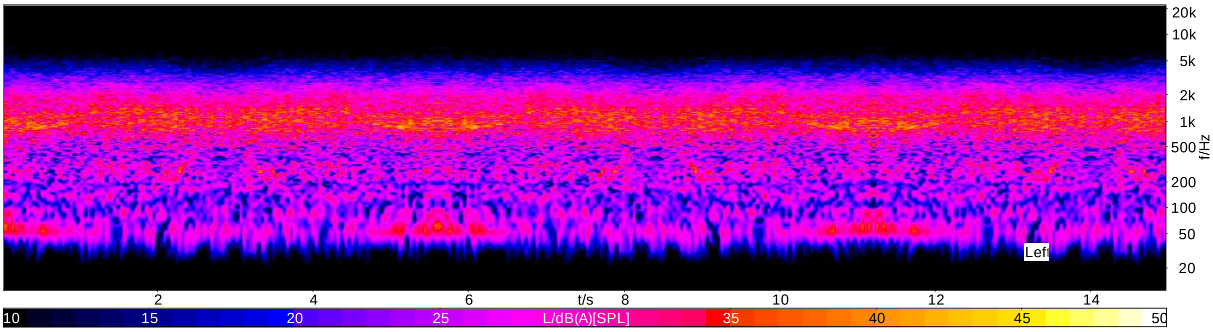

Figure 1. Spectrogram of the background noise (TU-Dresden Background)

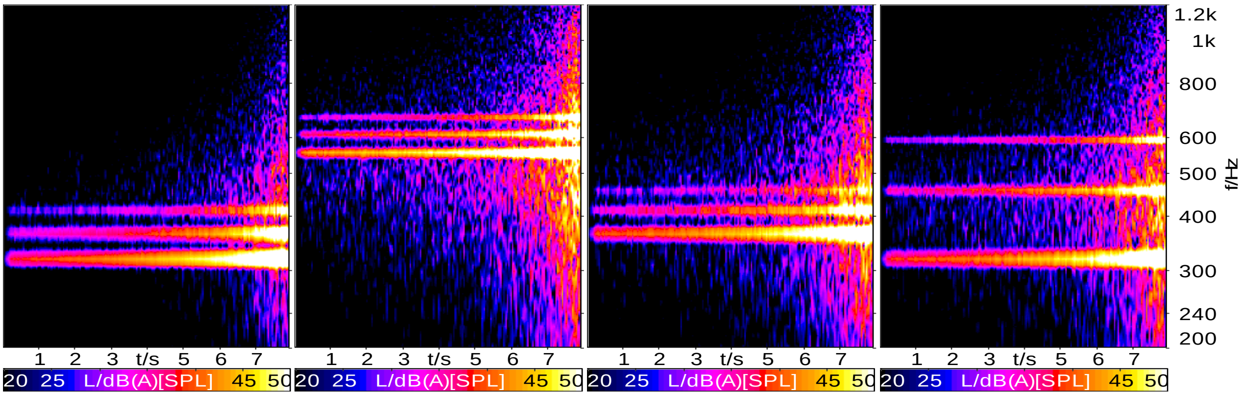

Figure 2. Examples spectrograms from the stimuli used in this study

3. SUBJECTIVE EVALUATIONS

The aim of the experiment is to measure the reaction time required by subjects to react to the auditive recognition of a vehicle passing by in a background noise.

3.1 Experimental procedure

In this study, detection times of approaching vehicles were measured at a constant speed of 10 km/h in a realistic background noises. In urban traffic; secondary roads without traffic lights or pedestrian crossings and a speed limit of 30 km/h are common, so that the situation examined here is a typical traffic situation in which a pedestrian must make a decision to cross the street. The main experimental procedure is similar to the one in (2). In order to prevent the subjects from guessing the beginning of the vehicle sound, randomized delay times between 1 s and 3 s were introduced. The vehicle sounds were randomized in the listening test and presented in a laboratory via a calibrated headphone of Sennheiser type HD 600. In the test setup all delays were determined and compensated. The sum of the uncorrectable hardware-related delay times resulted in an error of the reaction time measurement of below 10 ms which is negligibly small compared to the standard deviation of the results. A graphical user interface was used for the evaluation of the experiments and implemented as a Matlab GUI. The actual listening test was preceded by a training session by presenting all vehicle sounds individually. The subjects were instructed to answer the question "When do you notice the approaching vehicle?" by pressing a button immediately after detecting the approaching vehicle. The critical distance can be calculated with the following equation:

$$s_{veh} = v_{veh} \cdot t_{react} + \frac{v_{veh}^2}{2 \cdot a_{dec}} \qquad (1)$$

For typical reaction times (t_{react}) of concentrated ($t_{react} = 0.7s$) and unconcentrated ($t_{react} = 1.5s$) traffic participant (3), a deceleration of $a_{dec} = 8\ m/s^2$ and a vehicle velocity of $v_{veh} = 10\ km/h$ a critical distance of 5.82 m (concentrated) or 10.26 m (unconcentrated) could be measured.

The subjects were given additional explanations of the traffic scene and received training before the experiment.

3.2 Subjects

A total of 20 test subjects with an average age of 34.3 years (21 to 70) took part in the listening tests. 12 subjects were male and 8 subjects were female. Furthermore, 5 of these 20 subjects were blind or visually impaired. Before the listening tests, the hearing threshold was measured for each subject. Subjects with hearing impairments were excluded as “dropouts” from the study. Only subjects with presbyacusis were allowed.

3.3 Results

Figure 3 shows the experimentally determined perception intervals of the detection experiment as boxplots over all test subjects. Box plots are used for graphical display of the distribution of the detection times. It summarizes different robust scattering dimensions in a representation. The blue box corresponds to the area in which the average 50% of the data is located. The length of the box corresponds to the interquartile range (IQR). The IQR is a measure of the dispersion of the data. In addition, the median is indicated as a continuous red stroke in the box. Due to its location within the box, one gets an impression of the skewness of the data underlying the data. The lines (“whiskers”) show the values outside the box. The length of the whiskers is determined by the data values and not by the interquartile distance alone. This is also the reason why the whiskers do not have to be the same length on both sides. If there are no values outside the limit of 1.5 × IQR, the length of the whisker is determined by the maximum and minimum values. Otherwise, the values outside the whiskers are entered separately into the diagram. These outliers are shown as red crosses.

The x-axis shows the vehicle sounds and the y-axis shows the detection times in seconds. The two horizontal lines mark the critical distances for unconcentrated (black) and concentrated (gray-dashed) traffic participant for a constant speed of 10 km/h. The sounds above of the lines are only audible at a too low distance, so it would be difficult to prevent an accident. The detection times of all sounds were determined with a background noise with 55 dB(A).

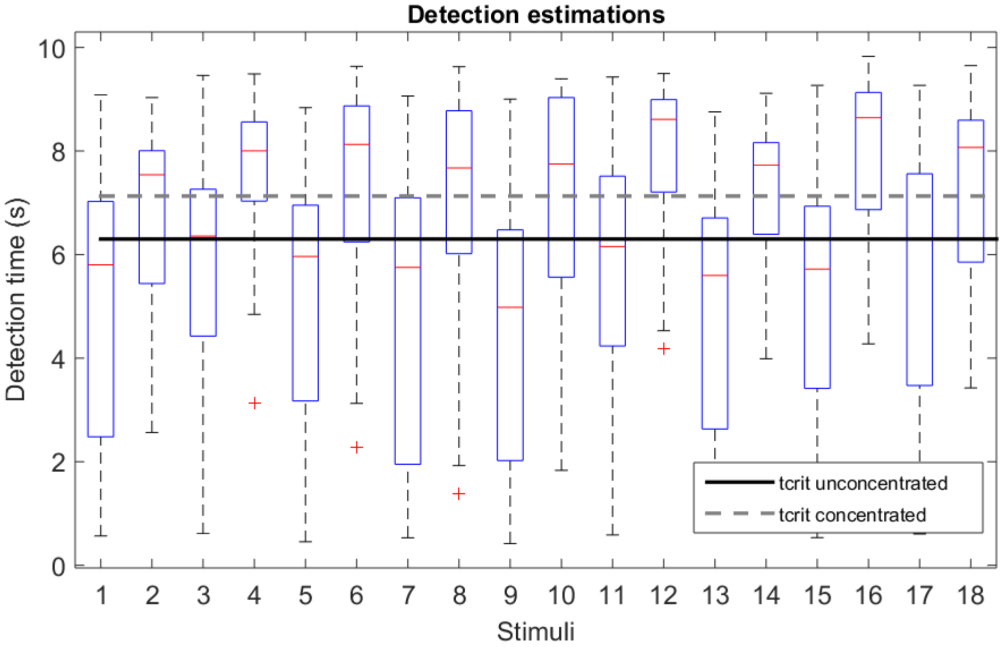

Figure 3: Results of the experimentally determined perceptive detection times for approaching vehicles in boxplots.

4. PSYCHOACOUSTICAL PARAMETERS

Psychoacoustical parameters are calculated in Head ArtemiS software. Considered acoustical variables are A-weighted sound pressure level, loudness over time, sharpness over time, tonality over time and roughness over time. It is assumed through this study, that the car, while driving with constant speed of 10 km/h, generates the constant level of external noise. Only difference in levels for the stimuli used for listening tests, are obtained because of the distance between car and listener. In order to calculate the single values from the stimuli, only the maximum values of the all aforementioned parameters are used. Those values are assumed to be the ones, when the car is at the same point with the listener, basically representing the generated noise itself, disregarding the effect of the distance.

Loudness calculations are based on DIN 45631 and is given in soneGF; sharpness calculations are based on DIN 45692; roughness calculations are based on Aures, tonality calculations are including 50% overlapping. For roughness calculations, in order to obtain reasonable results, first 0.5 seconds are not considered. (4)

4.1 Correlations with subjects estimations

The software IBM SPSS statistics 23 was used for the statistical analysis. The calculation of correlation can be used to show a relationship between two metric variables. This consists of a correlation coefficient and a p-value. The correlation coefficient indicates the strength and direction of the relationship. It is between -1 and 1. A value close to -1 indicates a strong negative relationship. A value close to 1 indicates a strong positive correlation. If there is no connection, the value is close to 0. The p-value indicates whether the correlation coefficient differs significantly from 0, i.e. whether there is a significant correlation. Usually, p-values less than 0.05 (*) are referred to as statistically significant and p-values less than 0.01 (**) are referred as highly significant. The Pearson correlation coefficient is used when the data is normally distributed and there is a linear relationship between the two variables.

Table 1 shows the correlation between the median values of the subjective estimations and the calculated psychoacoustic parameters. The highest negative correlation show the median values with the A-weighted levels and loudness. The detection estimations also show significant negative correlations with the calculated roughness. No significant correlation was found with the sharpness and the tonality.

Table 1. Correlation (Pearson) between calculated psychoacoustical parameters and median values of detection times obtained from listening tests

	A-weighted level (dB(A))	**Loudness (sone)**	**Roughness (asper)**	**Sharpness (acum)**	**Tonality (tu)**
Detection Time (median values)	-0.700**	-0.601**	-0.495*	-0.245	0.140
Significance	p=0.001	p=0.008	p=0.037	p=0.327	p=0.579

5. ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial neural networks are computational units, developed by imitating on the biological structure and operation methodology of a biological neuron. They consist of layers of interconnected neurons including activation functions, and all the connections have some mathematical expressions called as weights. They act like a usual biological neuron: getting information from the other neurons, processing the data in the cell body and transferring it to another neuron.

Neural networks are being used in different studies to perform complex curve fitting, data clustering, speech recognition, image recognition purposes etc. With the help of enough training data, artificial neural networks can be developed to perform aforementioned tasks, basically by adjusting the connections between the neurons, called weights. (5) In this study, an artificial neural network is tailored to have nonlinear curve fitting between the calculated psychoacoustical parameters of the synthesized stimuli and detection time estimations obtained from listening tests. There are different parameters need to be considered for tailoring a neural network architecture, such as training method, size of the network -particularly number of neurons in the hidden layer, division of the input data into training and validation parts and function type used within the tests. Most of the former studies using ANNs usually in sound quality estimations use directly the most efficient neural network architecture and the estimation results obtained by using this single ANN design. However in this study, selection procedure of the best performing network is also investigated in detail, by keeping some of the design parameters of a neural network parametric. At the end, 30 different neural network are obtained for each estimation and performance of the different neural networks are compared with each other to find the most efficient ANN architecture. Table 2 shows the values that kept parametric during the study, while division of data into training and validation is similar for all cases (70% training, 30% validation and test) and function type within the cells are being kept as constant (sigmoid functions).

Table 2. Parameters considered for comparing different ANN architectures

Number of training sets	5 different states
Training function	•Levenberg-Marquardt •Bayesian regularization •Scaled conjugate gradient
Network size – hidden layer size	•2 neurons in hidden layer •3 neurons in hidden layer
Result: 5*3*2 = 30 different neural networks	

In order to compare the performances of different neural network architectures, mean squared error values (MSE) between the calculated outputs from ANN and targets obtained from listening tests are used. Performance values are calculated by using the all stimuli for all 30 different neural network structures. Results are compared at the end, by using the different neural network architecture to understand the importance of tailoring process of a neural network.

For this study, A-weighted sound pressure levels, loudness, sharpness, roughness and tonality values are used as inputs for neural networks. As the target values, the median values of reaction time experiments are used. Figure 4 shows the two layer, feed-forward ANN architecture used in this study. It should be noted that, the number of neurons used in hidden layers are varied as 2 and 3.

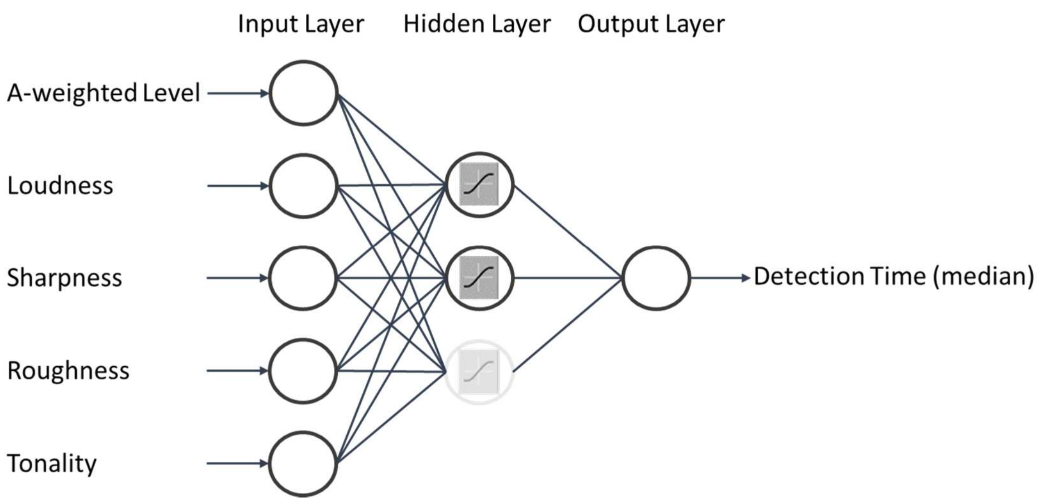

Figure 4. ANN architecture used in the study. Note that number of neurons in hidden layer is varied

5.1 DETECTABILITY PREDICTION BASED ON ANN

The estimations obtained from 30 different ANN architectures are given in Figure 5. It can be seen that, for the neural networks which have lower performance values, discrepancies between the target values obtained from listening tests and outputs obtained from ANNs are quite high. Even for some neural network architectures, training state is so irrelevant, such that the network gives the same output regardless of the input of all stimuli. However, the general tendencies of the results obtained from neural networks can be regarded to lie in the acceptable range.

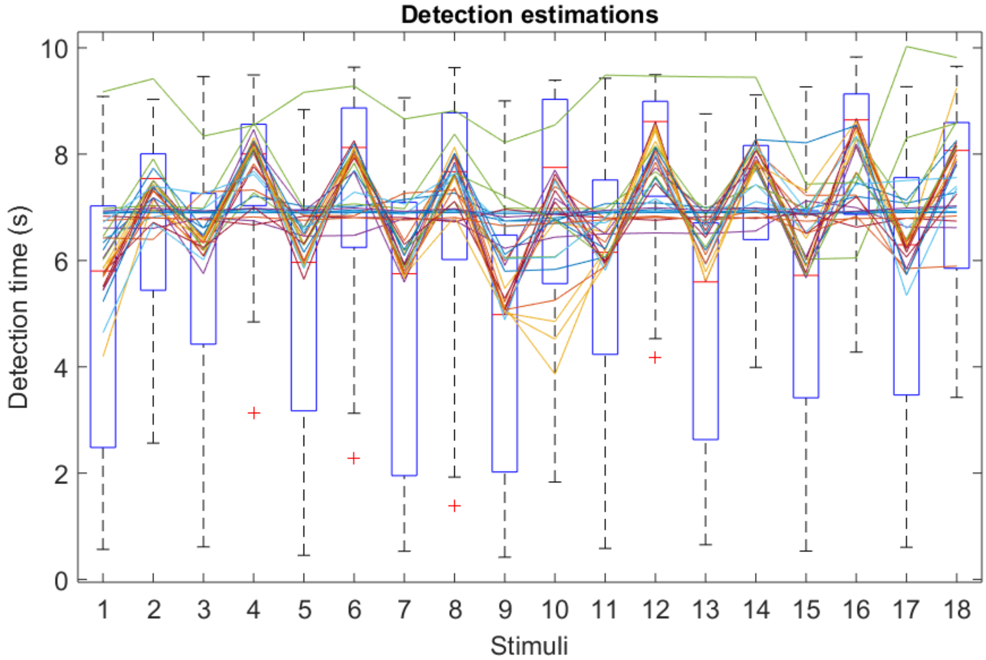

Figure 5: Results obtained from 30 different ANNs compared to subjective detection times.

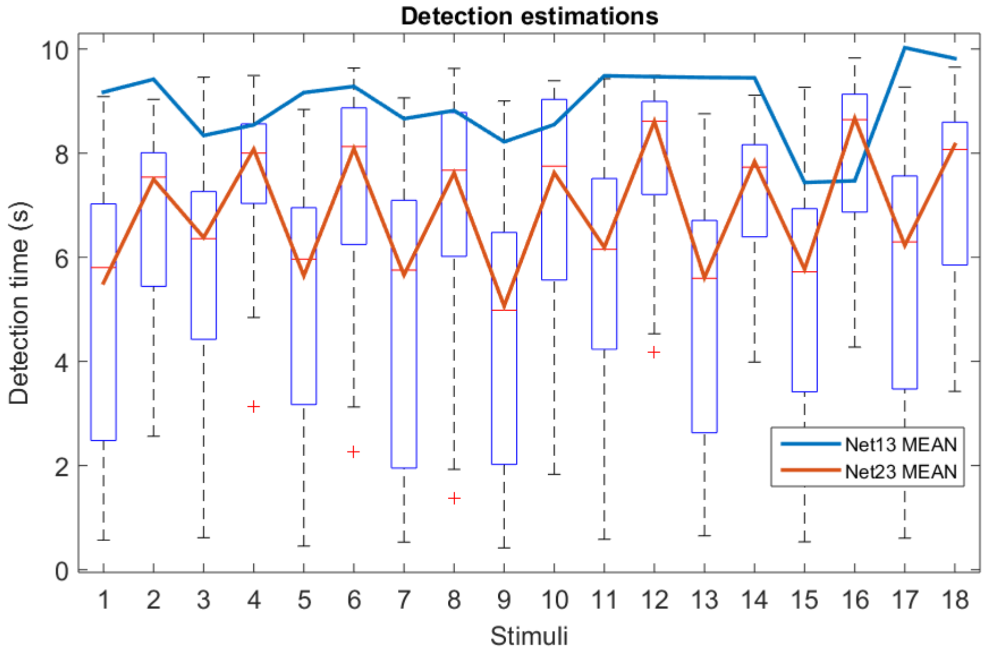

Figure 6. Results obtained from the worst and best performing network compared to subjective detection times

In order to see the differences between the best and worst performing network, results obtained from both network architectures are given in Figure 6 in detail. The blue line corresponds to the results obtained from worst performing network while the best is represented with red. Here it can be clearly deduced that, performance of the network is crucial for detection time evaluations. Moreover, it is clear to see in Figure 6 that, despite high values of standard deviations, the best performing neural network can fit the data of median values quite well.

6. CONCLUSION

The main aim of this study is to understand the detection time of different synthesized external car noises and understand the effectiveness of using artificial neural networks (ANNs) for predicting detection time estimations. For that reason, detectability estimations of electric vehicle sounds are obtained by subjective jury testing. 18 synthesized stimuli is used for listening tests with 20 test subjects including 5 blinds and visually impaired people. All the stimuli are presented with a realistic background noise and obtained results are given in boxplots.

Maximum A-weighted sound pressure levels, loudness, roughness, sharpness and tonality values are calculated for each stimuli and the correlation values are obtained between calculated psychoacoustical parameters and median values of detection times obtained from listening tests. All the regulations related to this topic includes minimum A-weighted sound pressure levels to be implemented in the cars to increase detectability, however, the correlation studies suggests that, loudness and roughness have also a strong effect on detectability issues.

For the second part, 30 different artificial neural network architecture are obtained and results are compared. It is stated that, tailoring a neural network is an important part of a study using ANNs as a prediction tools, since it includes different parameters to be considered. Comparing the MSE values, best performing neural network is selected within this study. Best performing neural network fits the data obtained from listening tests quite well.

ACKNOWLEDGEMENTS

The authors would like to acknowledge their debt to Dipl. –Ing. Margitta Lachmann for her technical assistance and thank all the subjects participated in listening tests.

REFERENCES

1. ECE, UN. Proposal for a new Regulation concerning the approval of quiet road transport vehicles (QRTV). ECE/TRANS/WP. 29/2016/26. Geneva, 2016.
2. Steinbach L., Altinsoy M.E, Rosenkranz R. “Elektromobilität: Angepasste Geschwindigkeits-Pegelskalierung erhöht die Sicherheit”, DAGA 2017 - 43rd German Annual Conference on Acoustics
3. Green M.: "How long does it take to stop?" Methodological analysis of driver perception-brake times. Transportation human factors, 2:195-216, 2000.
4. Zwicker, Eberhard, and Hugo Fastl. Psychoacoustics: Facts and models. Vol. 22. Springer Science & Business Media, 2013.
5. Atamer S., Altinsoy M.E, „ Application of Artificial Neural Networks for Understanding the Quality and Masculinity Perception of Electric Shavers“, DAGA 2017 - 43rd German Annual Conference on Acoustics