Predicting Detectability and Annoyance of EV Warning Sounds Using Partial Loudness

Gustav N. JACOBSEN[1]; Jeong-Guon IH[2]; Wookeun SONG[3]; Ewen N. MACDONALD[1]

[1] Hearing Systems Group, Electrical Engineering, Technical University of Denmark

[2] Acoustics Laboratory, Dept. of Mechanical Engineering, KAIST

[3] Brüel & Kjær Sound & Vibration Measurement A/S

ABSTRACT

At low speeds, electric vehicles (EV) emit less noise compared to internal combustion engine vehicles, making them more difficult for other road users to detect. In this study, the use of a computational partial loudness model for predicting detection and perceived annoyance of EV warning sounds was investigated and compared with the signal-to-noise ratio (SNR). In two experiments, detection thresholds were obtained for 4 different warning sounds in 5 different background noises. In a third experiment, subjective perceived annoyance ratings were obtained for 4 warning sounds in 5 noise conditions. For all experiments, the partial loudness and SNR measures were compared with the listener data. Overall, the detection thresholds in terms of partial loudness were similar for stationary warning sounds. This was not the case with SNR. The perceived annoyance increased with partial loudness as expected and the slope of the increase was similar across different warning sounds in different noise conditions. This was not the case with SNR. Thus, a partial loudness model is a better objective method to predict detection and annoyance of stationary warning sounds. However, more work is needed to improve the prediction of detecting non-stationary warning sounds.

Keywords: Electric vehicle, warning sounds, partial loudness, sound perception, detection, annoyance

I-INCE Classification of Subjects Number(s): 11.9.9, 13.2.1, 63.1, 63.2, 79.9

1. INTRODUCTION

Electric or hybrid vehicles (EVs/HEVs) have significant lower noise emissions at low speeds than conventional internal combustion engine vehicles (ICEVs) (1, 2). This might lead to a higher accident rate with other road users, such as pedestrians and cyclists, due to a reduction of auditory cues (3). Due to this high potential of being a social hazard, legislation have been proposed requiring electric vehicles to emit warning sounds at low speeds (4). The challenge for car makers is to make the warning sounds as detectable as possible while keeping the annoyance at the lowest possible level. EV manufacturers try to develop their own optimal warning sounds in terms of high detectability, high recognizability, low annoyance, and a product sound quality that fits the brand. So far, the evaluation of warning sounds in terms of detectability and annoyance has been based on various experiments and listening tests. These are very time consuming procedures and can be easily influenced by various sources of bias errors. Previous studies have investigated how optimal warning sounds in terms of detectability, annoyance, recognizability, and product sound quality should be designed, but no consensus has been reached yet (5-8).

A model for predicting detectability and annoyance would improve the warning sound design process as it would serve as an early indicator for determining whether a warning sound would be useful or not. Previous studies (9, 10) have indicated that detectability and annoyance are strongly related to loudness, and hence a loudness model can potentially predict detectability and annoyance of warning sounds in quiet. However, as EVs are to operate in noisy urban environments, background noise will influence the detectability and perceived annoyance due to masking effect. Therefore,

[1] gustav.n.jacobsen@gmail.com, emcd@elektro.dtu.dk
[2] j.g.ih@kaist.ac.kr
[3] woo-keun.song@bksv.com

partial loudness, i.e. loudness of a target sound in background noise, needs to be considered.

In this work, the loudness of the 4 warning sounds in different background noise conditions is investigated by using the Moore-Glasberg time-varying partial loudness model (11-13), and it is investigated if the model can appropriately predict detectability or perceived annoyance.

2. PARTIAL LOUDNESS MODEL

We implemented the partial loudness model of the Moore-Glasberg time-varying loudness model (11-13). The implementation is based on a commercial toolbox (14) for the calculation of partial loudness. The Moore-Glasberg loudness model computes loudness following an advanced model of the human ear based on the auditory filter bank concept.

The partial loudness model is an extension to the loudness model as it involves calculating and subtracting the excitation patterns of the background noise from the target signal. To include time-varying sounds is another extension, as it requires the calculation of a spectrogram. A flow chart of the binaural time-varying partial loudness model is shown in Fig. 1.

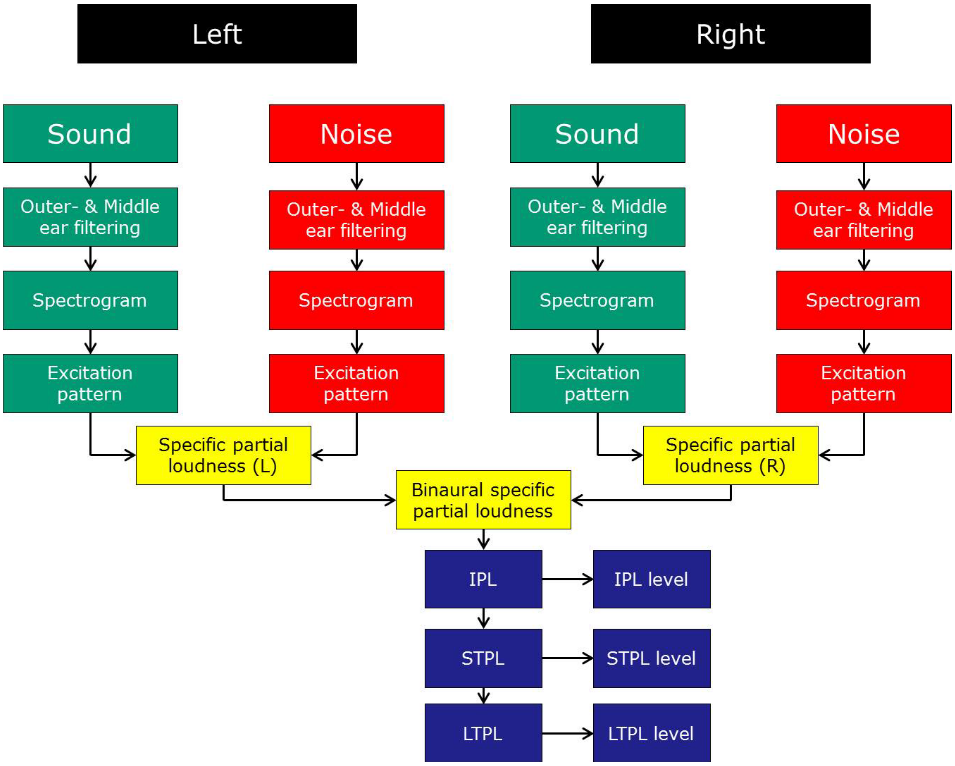

Figure 1 – Schematic of the Moore-Glasberg binaural time-varying partial loudness model used in this study.

In Fig. 1 the last three sections indicated in blue color represent the different time constants that determine the different degree of time dependencies caused by the difference in attack and release times:– instantaneous partial loudness, short-term partial loudness, and long-term partial loudness. In this work, the short-term partial loudness level is employed considering the length of stimuli (12). Further details on the model are explained in (11-13).

3. EXPERIMENTAL SETUP FOR SUBJECTIVE TEST

3.1 Experimental Setup

The experiments were conducted in a listening booth. The stimuli were presented over the headphones (Sennheiser HD-650). A filter compensating for the headphone transfer function was applied as well as a filter representing the free-field HRTF using a head and torso simulator (Brüel & Kjær type 4100) in order to derive the corresponding binaural signals of the monophonic stimuli. The setup was calibrated to ensure correct output levels.

3.2 Subjects

23 European subjects, 17 males and 6 females, ranging from 22 to 36 years of age participated in the experiments. Half the subjects completed experiment 1 first, and the other half completed experiment 2 first. Experiment 3 was always completed as the last. All subjects had normal hearing (audiometric thresholds less than 25 dB HL between 250 and 8000 Hz).

3.3 Warning sounds

Four different synthesized warning sounds were used in the experiments. S1, S3 and S7 are warning sounds previously used in studies by Singh [14], furthermore a 1 kHz pure tone was used as reference. The 1 kHz tone and S1 are stationary sounds while S3 and S7 are time-varying sounds. A brief description of the warning sounds is given below:

S1: Stationary sound with a significant tone at 300 Hz. It has a "sawtooth-like" sound.

S3: Consists of two components. A 500 Hz tone amplitude modulated with 6 Hz modulation frequency and a narrowband noise centered at 1500 Hz amplitude modulated with 2 Hz modulation frequency.

S7: Consists of two components. A narrowband noise centered at 300 Hz amplitude modulated with 2 Hz modulation frequency and a swept sine tone changing frequency from 700 Hz to 500 Hz over 4 seconds which is then repeated – A pitch change.

3.4 Background noises

White noise was low pass filtered with an IIR filter such that it had a similar frequency shape as stationary urban noise (15). This simulated urban noise is labelled *N*. Furthermore four noise recordings at a busy urban road with a speed limit of 40 km/h were used – *N1*, *N2*, *N3* and *N4*.

4. EXPERIMENTS AND DISCUSSION ON THE RESULTS

4.1 Experiment 1: Detection threshold

The purpose of Experiment 1 was to determine if a general threshold based on the maximum short-term partial loudness level (STPL level) can be used to predict the detection of warning sounds in background noise.

First the subjects listened to the warning sounds in quiet in order to familiarize them with the stimuli. The subjective reaction time was measured by asking the subject to press a button as soon as he/she heard a short noise impulse. 6 repetitions were made and the average of the last 4 was used.

The subject was presented with one of the four warning sounds in presence of one of the five background noises totaling 20 combinations. The stimuli were binaural and presented diotically, i.e. the same signal is presented at both ears. The warning sound was initially presented at a level below the masking threshold but increased linearly by 40 dB over 15 seconds. The noise was held at a constant level. The subject pressed a button when he/she was 'absolutely certain' that he/she detected the warning sound. The stimuli were presented in a balanced Latin square design to avoid carry over effects. Moreover, one of four gain factors (-4.5 dB, -1.5 dB, +1.5 dB and +4.5 dB) was used to adjust the starting level of the target signal in order to change the point in time at which detection occurs and thus minimize learning effects. For each listener, after the short training session, two identical sessions of the 20 combinations, which were unknown to the listener, were conducted. The procedure is presented in figure 2.

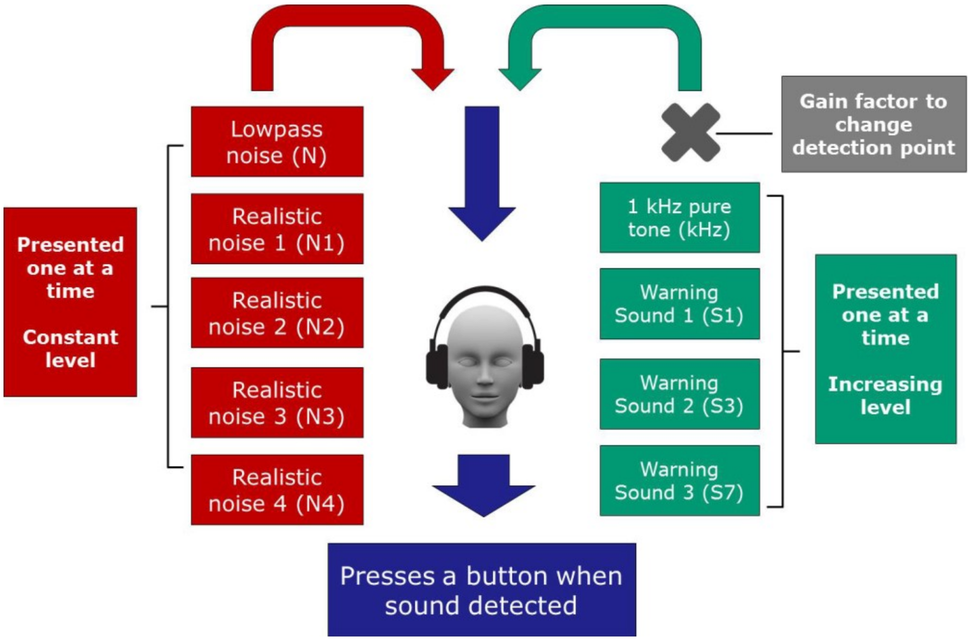

Figure 2 – Schematic of procedure in experiment 1.

Based on the detection points in time it is possible to obtain the maximum signal-to-noise ratio (SNR) before detection. Similarly it is possible to obtain the maximum STPL level before detection using the partial loudness model. The individual reaction time is subtracted from the detection points before this step in order to reduce bias due to motor latency.

The STPL-level and SNR for detection are shown in figure 3 and 4. From figure 3 it is seen that a detection threshold of approximately 27 phon exists for the stationary warning sounds (1 kHz and S1) independent of background noise as the horizontal bars as well as the black points are fairly well aligned. The time-varying warning sounds (S3 and S7) have a detection threshold higher than the stationary warning sounds at approximately 35 phon. The increase is possibly because the subjects want to be 'absolutely certain' and hence waits for the next peak level of the amplitude modulation.

Comparing figure 3 with figure 4 it is seen that the SNR for detection are more dependent on the background noise as the black points are not aligned to the same degree as in figure 3 and hence the partial loudness model is a better model for predicting detection in presence of different maskers.

The mean STPL levels for detection across warning sounds were in the range 26 to 36 phon. The mean SNR levels for detection across warning sounds were in the range -17 to -12 dB.

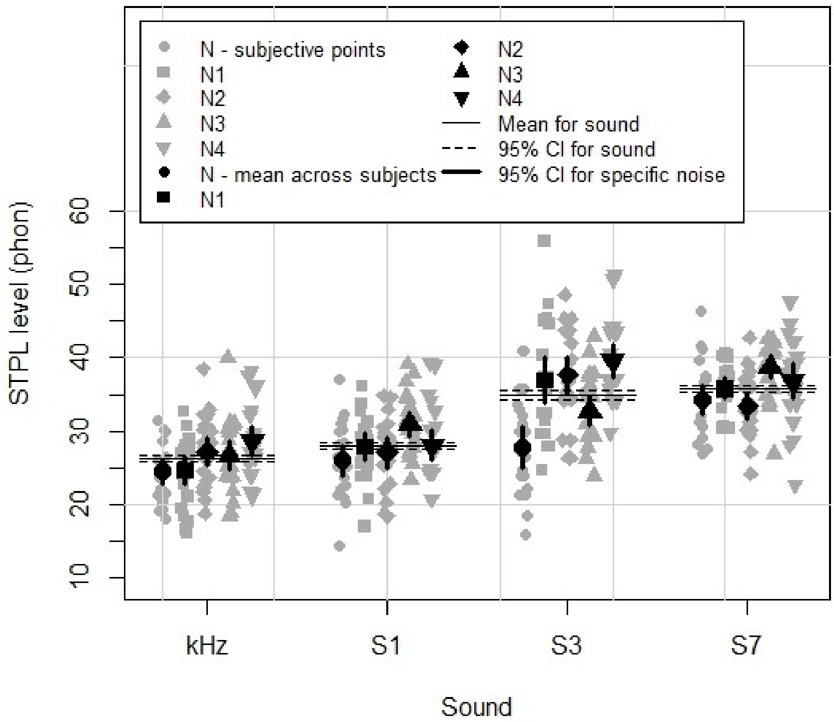

Figure 3 – STPL level at the subjective detection threshold across warning sounds (x-axis) and background noises in experiment 1. Black points represent the mean for a specific background noise for a specific warning sound. The vertical bars represent the 95% confidence interval. The horizontal bars represent the mean across all background noises for a given warning sound. The dotted horizontal bars represent the 95% confidence intervals.

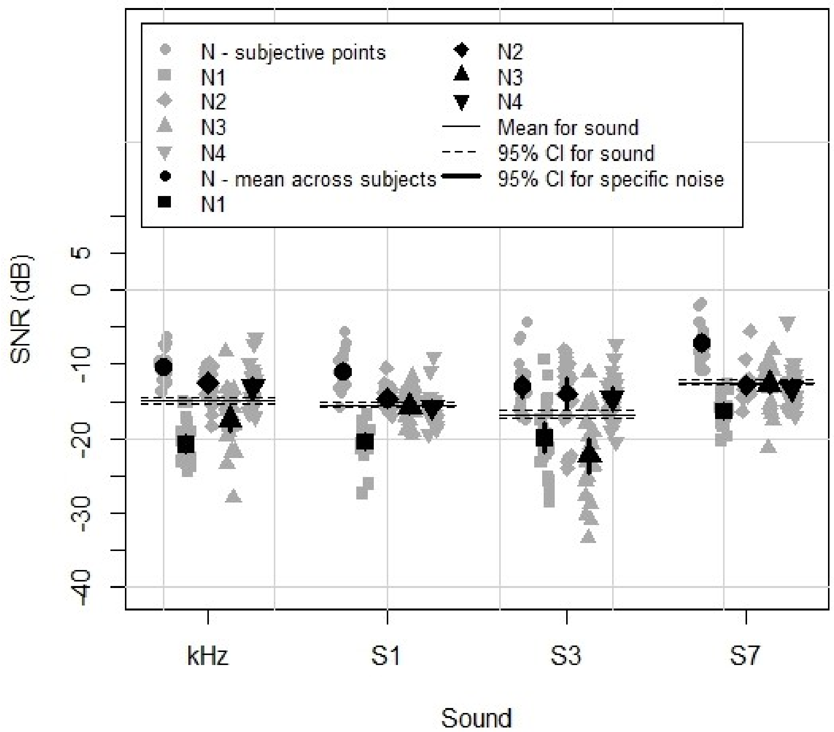

Figure 4 – SNR level at the subjective detection threshold across warning sounds (x-axis) and background noises in experiment 1. Black points represent the mean for a specific background noise for a specific warning sound. The vertical bars represent the 95% confidence interval. The horizontal bars represent the mean across all background noises for a given warning sound. The dotted horizontal bars represent the 95% confidence intervals.

4.2 Experiment 2: Detection threshold

The purpose of experiment 2 was similar to that of experiment 1. However, in Experiment 2 an adaptive force choice paradigm was used to reduce the possible effects of individual bias when detecting warning sounds in background noise.

The subject was presented with three consecutive sound samples. All three samples contained the same 1 second segment of the simulated urban noise (*N*) but one sample also contained a 1 second segment of one of the warning sounds. The subject had to indicate which sound sample contained the warning sound (3AFC). An adaptive one-up-two-down method was used, yielding the 70.7% detection threshold. A step size of 8 dB was used until two reversals had occurred, then 4 dB for one reversal and finally a step size of 2 dB until two reversals occurred (8, 8, 4, 2, 2 dB). The detection threshold was determined as the mean of the last 6 trials' SNR. Three runs were conducted for each warning sound yielding a total of 12 runs for each subject. For each trial a new 1 second segment of the stationary noise was used in order to avoid frozen noise adaptability.

Based on the obtained SNR for detection it is possible to calculate the STPL level for detection using the partial loudness model. The results are shown in figures 5 and 6. From figure 5 it is seen that a similar detection threshold of approximately 17 phon was obtained for the stationary sounds (1 kHz and S1) and S3. The threshold for the S7 sound is significantly higher at 29 phon - no obvious explanation can be given for this. The thresholds obtained in experiment 2 are substantially lower than the ones obtained in experiment 1, which is expected due to the lower point on the psychometric function.

From figure 6 it is seen that the detection thresholds using SNR are similar for the stationary warning (1 kHz and S1) sounds at approximately -15 dB whereas the thresholds differ for the time-varying sounds (-12 and -19 dB for S3 and S7 respectively). Hence, neither the partial loudness model nor the SNR model are adequate for predicting detectability across all conditions. However, the partial loudness model proved suitable for predicting detectability when limiting the warning sounds to be stationary.

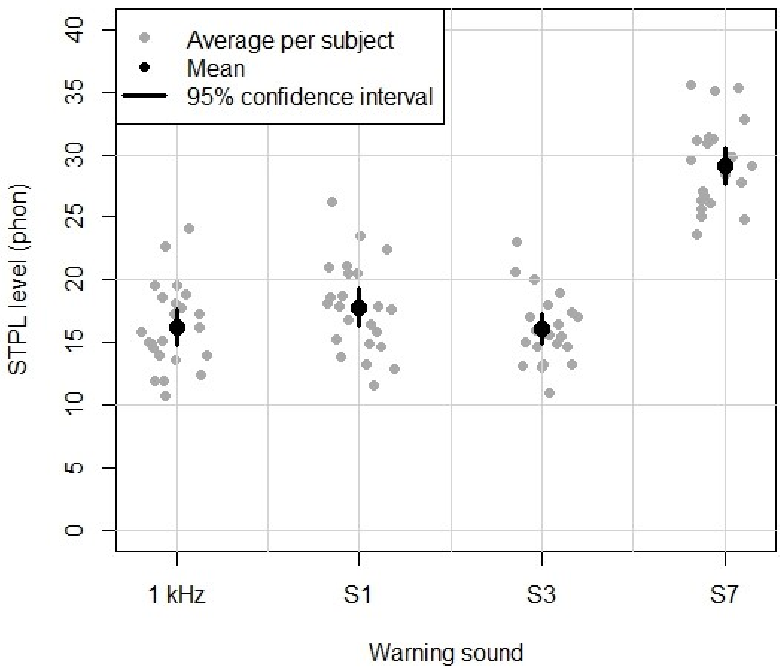

Figure 5 – STPL level for detection across warning sounds (x-axis) in experiment 2. Black points represent the mean for a specific warning sound. The vertical bars indicate 95% confidence intervals.

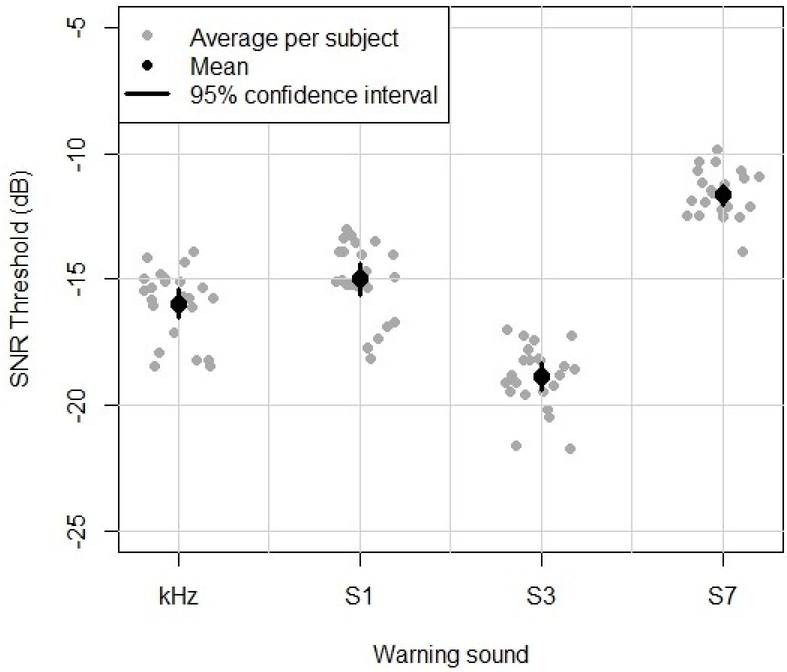

Figure 6 – SNR for detection across warning sounds (x-axis) in experiment 2. Black points represent the mean for a specific warning sound. The vertical bars indicate 95% confidence intervals.

4.3 Experiment 3 – Annoyance rating

The purpose of Experiment 3 was to investigate if perceived annoyance was correlated with partial loudness and/or SNR.

The subjective annoyance of the previously tested warning sounds (1 kHz, S1, S3 and S7) were evaluated in five different noise conditions – quiet (no noise), high and low SNR with simulated urban noise (*N* from Experiments 1 and 2) and high and low SNR with an urban noise recording (*N3* from Experiments 1 and 2).

The subjects were told to imagine themselves in an urban environment (e.g., sitting outside of a café or waiting for green light at a crosswalk). The annoyance of the warning sound was evaluated on a continuous scale that had a range of 0-50; 0 corresponding with not being annoying (very little annoyance) and 50 corresponding with being very strongly annoying. The subject was told to rate only the annoyance of the warning sound and not the background noise. The subject rated the annoyance of the four warning sounds in presence of the same noise condition in a trial. A total of five trials were to be completed each representing one of the five noise conditions. For each trial the warning sounds were assigned randomly to four buttons. The order of the trials followed a balanced latin square design.

The mean STPL level and mean SNR for each sound/noise combination and the corresponding annoyance rating is plotted in figure 7 and 8. Linear regression lines for each warning sound are added. From figure 7 and 8 it is obvious that the slopes of the linear regression are more alike for the mean STPL level than the mean SNR, furthermore Pearson's R^2 coefficients are much higher for the STPL level, ranging from 0.64 to 0.91 compared to a range of 0.28-0.60 for the SNR. However, while the slopes are similar, it is also seen that partial loudness alone is not sufficient for predicting perceived annoyance – The 1 kHz is consistently perceived as the most annoying sound while S1 is consistently perceived as the least annoying. Thus, other factors than partial loudness have an impact on annoyance.

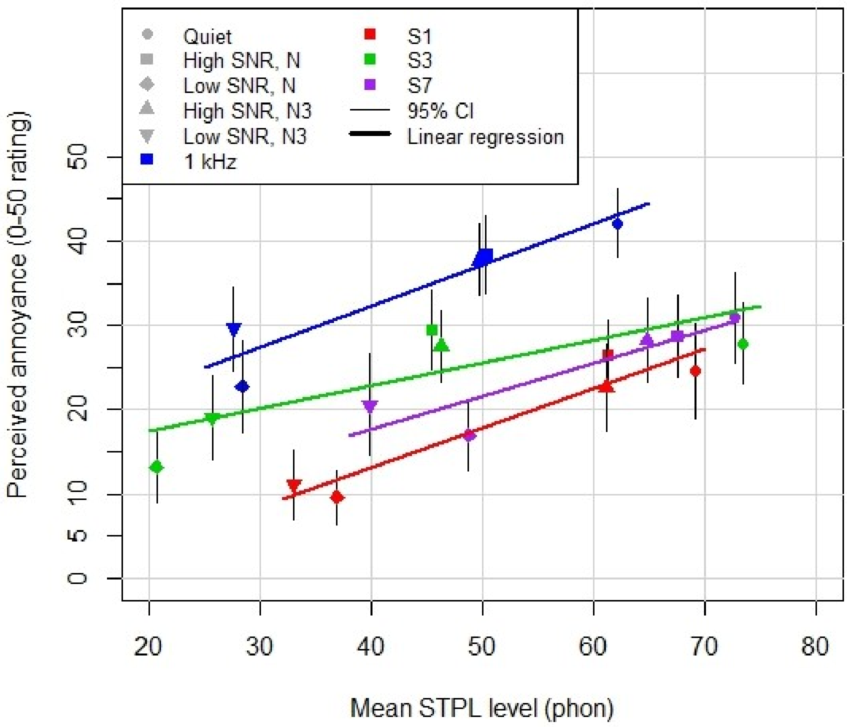

Figure 7 – Annoyance rating plotted against the mean STPL level for each warning sound in the five noise conditions. Linear regression for each warning sound is added. Each color represents a specific warning sound. Each symbol represents a specific noise condition.

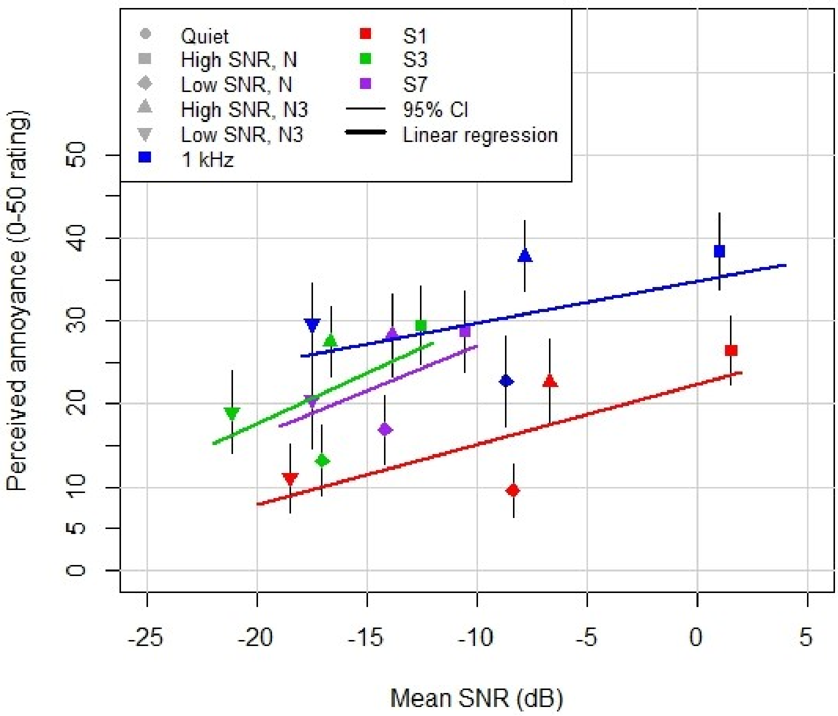

Figure 8 – Annoyance rating plotted against the mean SNR for each warning sound in the five noise conditions. Linear regression for each warning sound is added. Each color represents a specific warning sound. Each symbol represents a specific noise condition.

Table 1 – Slope of linear regression and R^2 values for partial loudness and SNR

Warning Sound	Slope STPL level	Slope SNR	R^2 STPL level	R^2 SNR
1 kHz	+0.49 point/phon	+0.51 point/dB	0.88	0.28
S1	+0.47 point/phon	+0.72 point/dB	0.91	0.60
S3	+0.27 point/phon	+1.22 point/dB	0.64	0.34
S7	+0.39 point/phon	+1.08 point/dB	0.81	0.33

The average of the linear regression slopes based on STPL is 0.40 point/phon, which, when this slope is used, yield R^2-values of 0.86, 0.89, 0.48 and 0.81 for the 1 kHz, S1, S3 and S7, respectively. The average of the linear regression slopes based on SNR is 0.88 point/dB which yield R^2-values all below 0.5.

5. CONCLUSIONS

In experiment 1 and 2 similar detection thresholds in terms of STPL level were obtained for the stationary warning sounds (1 kHz and S1). The detection threshold for the stationary warning sounds was 27 phon in experiment 1 (for listeners to be absolutely certain they detected the warning sound) and 17 phon in experiment 2 (70.7% point on the psychometric function). No consistent detection threshold was obtained for the time-varying warning sounds suggesting further work investigating the time-varying effects such as amplitude and frequency modulation is needed. In experiment 3 it was shown that partial loudness alone cannot completely predict annoyance of a sound. However, there was a higher correlation between partial loudness and perceived annoyance within each warning sound than for the SNR and perceived annoyance. Increasing the partial loudness of a warning sound with 1 phon increases the perceived annoyance with approximately 0.4 points on a 50 point rating scale.

ACKNOWLEDGEMENTS

This project was conducted as part of a master's thesis project carried out at the Technical University of Denmark (DTU), Korea Advanced Institute of Science and Technology (KAIST) and Brüel & Kjær Sound and Vibration Measurement A/S. The authors would like to thank Sneha Singh and the University of Warwick for providing the warning sounds used in this project.

REFERENCES

1. Dubois F, Baudet G, Chamard J C. EVADER: Electric Vehice Alert for Detection and Emergency Response, Société Française d'Acoustique. Acoustics 2012, April 2012, Nantes, France. <hal-00810760>.
2. Altinsoy M E. The detectability of conventional, hybrid and electric vehicle sounds by sighted, visually impaired and blind pedestrians. Inter-Noise 2013, September 2013.
3. U.S. Department of Transportation National Highway Traffic Safety Association. Incidence of pedestrian and bicycle crashes by hybrid electric passenger vehicles. Technical report. September 2009
4. U.S. Department of Transportation National Highway Traffic Safety Association. Minimum sound requirements for hybrid and electric vehicles. January 2013.
5. Altinsoy M E, Landgraf J, Lachmann M. Investigations on the Detectability of Synthesized Electric Vehicle Sounds - Vehicle Operation: Approaching at 10 km/h. Forum Acusticum, September 2014.
6. Parizet E, Robart R, Ellermeier W, Janssens K, Haider M, Leduc Y, Quinn D, Chamard J C. Additional efficient warning sounds for electric and hybrid vehicles. Forum Acusticum, September 2014.
7. Sakamoto I, Sekine M, Morita K, Houze H. Research for standardization of measures against quiet HV/EV in Japan. Forum Acusticum, September 2014.
8. Yasui N, Miura M. Relationship between characteristics of amplitude fluctuation and detectability of warning sounds for electric vehicle. Forum Acusticum, September 2014.
9. Yamauchi K, Sano T, Hasegawa S, Tamura F, Takeda Y. Detectability and hearing impression of additional warning sounds for electric or hybrid vehicles. Inter-Noise 2014, November 2014.
10. Zwicker E, Fastl H. Psychoacoustics, Facts and Models. Springer, 2nd edition, 1999.

11. Moore B C J, Glasberg B R, Baer T. A model for the prediction of thresholds, loudness and partial loudness. Journal of the Audio Engineering Society, 45(4), April 1997.
12. Moore B C J, Glasberg B R. A model of loudness applicable to time-varying sounds. Journal of the Audio Engineering Society, 50(5), May 2002.
13. Moore B C J, Glasberg B R. Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds. Journal of the Audio Engineering Society, 53(10), October 2005.
14. GENESIS, "Loudness Toolbox" - A MATLAB toolbox provided by sound company GENESIS.
15. Singh S, Payne S R, Jennings P A. Toward a methodology for assessing electric vehicle exterior sounds. IEEE Transactions on Intelligent Transportation Systems, 15(4), August 2014.
16. Can A, Leclercq L, Lelong J, Botteldooren D. Traffic Noise spectrum: Dynamic modelling vs. experimental observations. Applied Acoustics, 71(8), August 2010.