Do experiments in the virtual world effectively predict how pedestrians evaluate electric vehicle sounds in the real world?

S. Singh[a,*], S. R. Payne[b], J. B. Mackrill[a], P. A. Jennings[a]

[a]*WMG, University of Warwick, Coventry, West Midlands CV4 7AL, UK*

[b]*School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, EH14 4AS, UK*

[*] *Corresponding author at: WMG, University of Warwick, Coventry, West Midlands CV4 7AL, UK.*
Tel.: +44 (0) 24 76151579.
E-mail address: sneha.singh@warwick.ac.uk (S. Singh).

Abstract

New laws stipulate that electric vehicles must emit additional sounds to alert pedestrians of the vehicles' approach to prevent potential collisions. These new sounds will also influence pedestrians' impression of the vehicle brand. A methodology has been developed to evaluate electric vehicle (EV) sounds in a virtual-world environment by assessing; a) detectability and recognisability to ensure pedestrians' safety, and b) emotional evaluation of the sound quality to determine its impact on the perception of the vehicle brand. This experimental study examines external validity of the methodology. Fourteen participants evaluated an EV, emitting three sounds, in a traffic scenario in a real-world and a virtual-world environment. The traffic scenario involved a pedestrian 'standing' at a residential road junction while the EV travelled at 12 mph from behind the pedestrian, arriving at the junction at one of two pre-set times. Results show that the presented virtual-world methodology accurately predicts pedestrians' evaluation of detectability of EV sounds and powerfulness and pleasantness of the vehicle brand in the corresponding real-world scenario. It also predicts the ranked order of sounds in the real-world for detection distance and recognisability. Arguably, for similar methods and setups, virtual-worlds would effectively predict pedestrians' evaluation in the real-world. Interestingly, varying a vehicle's arrival time, just like a real-world scenario, is found to affect pedestrians' detection rate. Unlike experiments in the real-world, the presented methodology for experiments in virtual-world benefits from being reliable, quick, easy to implement, with more experimental control and options to easily manipulate any experiment variables.

Keywords: electric vehicle sounds; virtual world; detection; evaluation; recognition; quiet vehicle.

1. Introduction

Electric Vehicles (EVs) and Hybrid Electric Vehicles (HEVs) are quieter at low speeds compared to Internal Combustion Engine vehicles (ICEVs). The sound pressure level of an EV is usually between 3 to 20 dB(A) lower than an ICEV of comparable physical specifications when operating roughly below 13 to 15 mph (Garay-Vega, Hastings, Pollard, Zuschlag, & Stearns, 2010; JASIC, 2009). At such low speeds pedestrians and cyclists are unlikely to hear Electric Vehicles or Hybrid Electric Vehicles running in electric mode ((H)EVs) sufficiently in advance to be able to avoid a potential collision (Garay-Vega et al., 2010; JASIC, 2009). Indeed, accident statistics worldwide show that pedestrians and cyclists are at a significantly higher risk of colliding with (H)EVs compared to ICEVs (Hanna, 2009; Morgan, Morris, Muirhead, Walter, & Martin, 2011). Particularly, traffic scenarios critical to pedestrians' safety include those where the EVs are moving straight towards a pedestrian's path, turning into a pedestrian's path, accelerating from stop, idling or reversing; in crossroads, junctions, parking lots, and residential roads at or below speeds of 12 mph (Garay-Vega et al., 2010; Hanna, 2009).

To resolve this issue, new vehicle safety standards worldwide stipulate that (H)EVs must emit additional sounds to alert pedestrians, cyclists and other road users of the vehicles' approach to ensure pedestrians' safety (Dalrymple, 2013; MLIT & JASIC, 2010; QRTV, 2011, 2012). The prominent legislation in this regard are Japan's MLIT guidelines for using sound generating devices for (H)EVs namely "Approaching Vehicle Audible Systems" (AVAS) (MLIT & JASIC, 2010), UNECE's Global Technical Regulation (GTR) for the AVAS for European (H)EVs (QRTV, 2011, 2012), and the Federal Motor Vehicle Safety Standard (FMVSS) by the US government (Dalrymple, 2013). (H)EVs' inherent sound level, however, increases with speed as the tyre-road interaction sound becomes dominant thereby eliminating the need for an additional sound to aid detection of these vehicles at higher speeds (Dalrymple, 2013; Garay-Vega et al., 2010; JASIC, 2009; MLIT & JASIC, 2010; QRTV, 2012). The existing variety in (H)EV models and specifications causes variation in their inherent sound level, which in turn varies the speed at which (H)EV become audible "enough" compared to ICEVs. Current legislations are therefore less specific and recommend that additional sounds should be emitted continuously till the vehicle attains a speed somewhere between 12.5 mph to 25.5 mph (≈ 20 to 41 kph) and at idle and reverse (Dalrymple, 2013; MLIT & JASIC, 2010; QRTV, 2011, 2012).

Additionally, the legislations for (H)EV sounds put emphasis on appropriateness of the type of sound to be used for (H)EVs. They recommend that these sounds must be recognisable as a vehicle, so that pedestrians could intuitively associate the sounds to a vehicle in operation (Dalrymple, 2013; MLIT & JASIC, 2010). Particularly, the UNECE and Japan's guidelines prohibit using siren, horn, chime, bell and emergency vehicle sounds; alarm sounds; intermittent sound; melodious sounds, animal and insect sounds; and sounds that confuse the identification of a vehicle and/or its operation (MLIT & JASIC, 2010; QRTV, 2011, 2012). However, an area that remains overlooked by the policy makers is that these sounds should promote positive impressions of the vehicle brand as desired by the manufacturer. A vehicle's sound has always been an important characteristic for reinforcing the brand image of the vehicle. Enhancing a vehicle's sound quality to influence and increase customer satisfaction has been an integral part of the automotive design process (Bisping, 1995; Cerrato, 2009; P. A. Jennings, Dunne, Williams, & Giudice, 2010; Miśkiewicz & Letowski, 1999; Özcan, 2014). People hearing the exterior sounds could evaluate the vehicle as a brand, in terms of simply liking to hear the vehicle pass-by, or as a potential consumer who may want to purchase the vehicle. Therefore, a rigorous methodology is needed to evaluate potential EV sounds on the criteria of both pedestrians' safety and impressions of the vehicle brand.

Currently, (H)EV exterior sounds are evaluated only to asses pedestrians' safety using detection tests either on-road (Emerson, Kim, Naghshineh, Pliskow, & Myers, 2013; Emerson, Naghshineh, Hapeman, & Wiener, 2011; Goodes, Bai, & Meyer, 2009; Hastings, Pollard, Garay-Vega, Stearns, & Guthy, 2011) or in a laboratory (Ashmead et al., 2012; Barton, Ulrich, & Lew, 2012; Garay-Vega et al., 2010; JASIC, 2009; Morgan et al., 2011; Parizet, Ellermeier, & Robart, 2014). On-road evaluations involve driving the "target vehicle" (the vehicle being evaluated) in real-world scenarios such as parking lots, crossroads and junctions (Emerson et al., 2011; Goodes et al., 2009) or in secured test tracks (Emerson et al., 2013; Hastings et al., 2011) after reserving the test site to allow no traffic interference and very low background sound (Emerson et al., 2013; Goodes et al., 2009; Hastings et al., 2011). The participant, usually blind (folded), detects the vehicle by either raising hands or through some push buttons

(Emerson et al., 2013; Goodes et al., 2009; Hastings et al., 2011; JASIC, 2009). Although conducted in a real-world environment, these evaluations lack appropriate context as here a person evaluates the vehicle in the absence of a visual context and with unrepresentative ambient soundscape (low level background with no transient sounds such as vehicles, construction and nature sounds). A correct context is important for a listening evaluation to achieve more ecologically valid experiment (Singh, Payne, & Jennings, 2014), i.e. to ensure the methods, materials and settings represent the real-world settings (Brewer & Crano, 2014). Some real-world evaluation tests do maintain an appropriate context where the pedestrian detects the target vehicle's sounds in presence of additional traffic and visual context (Emerson et al., 2011). But then, these compromise on the repeatability and consistency across experimental conditions and across participants.

Difficulties in implementation, long experiment durations and the resulting experimenters' and participants' fatigue has made laboratory evaluations a quicker, more convenient alternative to on-road evaluations. Laboratory evaluations involve vehicle detection in controlled laboratories, in the absence of any visuals or traffic sounds, by playing binaurally recorded or simulated vehicle sounds through headphones or speakers (Ashmead et al., 2012; Barton et al., 2012; Garay-Vega et al., 2010; JASIC, 2009; Morgan et al., 2011; Parizet et al., 2014). This environment provides better experimental control, thus more repeatability and consistency. But it lacks appropriate context due to absence of real-life ambient soundscapes and visual scenarios. These tests are usually simple in design and do not use variation in a vehicle's manoeuvre such as a change in the vehicle's time of arrival at the pedestrian's spot (Garay-Vega et al., 2010; JASIC, 2009; Morgan et al., 2011; Parizet et al., 2014). Typical laboratory vehicle detection tests use a "fixed-play" technique and the target vehicle sound is played as soon as a new experimental condition begins. However, varying the arrival time makes the pedestrian-vehicles more realistic. Interestingly, the arrival time may affect pedestrians' detection rate (Singh, Payne, & Jennings, 2014).

A methodology that addresses these issues in evaluating EV exterior sounds has been proposed recently (Singh, Payne, & Jennings, 2014; Singh, Payne, Mackrill, & Jennings, 2014b). This is achieved by evaluating the rate at which pedestrians' detect EV sounds to ensure pedestrians' safety, and evaluating the emotional characteristics of the vehicle sound in the perceptual dimensions of automotive sound quality to understand their impact on pedestrians' impression of the vehicle brand. Testing this methodology(Singh, Payne, & Jennings, 2013, 2014) showed that if EVs emit unrealistic or unrecognisable sounds, in the presence of a real-world ambient soundscape, then participants find the detection task difficult because of the unrecognisability of the EV sounds which they confuse with variations in the ambient soundscape. An option to evaluate the 'detectability' of sounds on a subjective scale along with recording the 'exact' time when the car is detected could enhance the detection task and make the participants feel more confident about their results (Singh et al., 2013; Singh, Payne, & Jennings, 2014). So, based on legislative guidelines and these findings the amended methodology now also includes participants' evaluations of the 'recognisability' and 'detectability' of the sounds. Moreover, the methodology has an option for pedestrians to record the detection time more than once to monitor their self-reported detection errors if, and when they confuse a vehicle sound.

The developed methodology evaluates EV sounds within a virtual-world environment which simulates a pedestrian's viewpoint of realistic traffic scenarios that are critical to pedestrians' safety. The virtual-world environments combine the benefits of on-road and laboratory evaluation methods. They provide an appropriate context for evaluating EV sounds as the inclusion of ambient soundscapes, visuals, and vehicle sounds enables a person to potentially experience the traffic scenario just like a real-world pedestrian. Simultaneously the researcher can fully control experimental conditions with less interference of external factors. However, any experimental methodology is only effective if its results accurately predict or reasonably generalise the real world. This is one of the long debated, yet an unanswered question in the literature on virtual environments (Chan, Pradhan, Pollatsek, Knodler, & Fisher, 2010). Therefore, this experimental study aims at testing external validity of the developed and amended methodology. The primary objective is to determine if, in a given traffic scenario, pedestrians evaluate EVs emitting sounds in a virtual-world environment in the same way as in a real-world environment. Here, the evaluation of EV sounds constitutes the aspects of detection rate, detectability, recognisability and impression of the overall vehicle brand. The other objective of the study is to determine if randomly varying the vehicle's arrival time affects the pedestrians' evaluation. Lastly, the study compares aspects such as the duration, implementation, reliability and control of experiments in a virtual-world with the real-world.

2. Method

A listening evaluation experiment was conducted in two environments, the real-world and virtual-world, using the same methods, stimuli, and participants.

2.1. Participants

All recruited participants self-reported as having no hearing or uncorrected visual impairments. Additionally, those self-reporting as pregnant, unfit, unwell, or with any symptoms of feeling dizzy, nauseated or sick before the evaluation, were excluded from participation for ethical considerations. The final data was collected from 14 participants, 10 males and 4 females. Participants comprised of staff and students at the University of Warwick in the age group of 18 to 55 years, with the modal age of 26-35 years. Based on a 2x2x3 repeated measures Analysis of Variance design, sample size met the minimum number of participants required (n=12) for a minimum statistical power of 0.8 (Cohen, 1988) and type I error probability, $\alpha = 0.05$, with a medium effect size, $f = 0.25$ (Cohen, 1988). This was calculated using Software G*Power 3.1.7 (Faul, Erdfelder, Lang, & Buchner, 2007).

2.2. Environments

2.2.1. Real-world

Participants listened to car sounds while standing at a real world road junction (see Fig. 1) within a secured residential area at the University of Warwick campus. The researcher stood next to the participant (Fig. 1) in order to coordinate the experiment. The target car was an EV that emitted different sounds of which 'sound pressure level' (SPL) and character was controlled using VSound software developed by Brüel and Kjær. For this purpose, the EV was fitted with speakers on its front exterior positioned below the windscreen (Fig. 2). The driver could select various sound profiles (5 to 12 s wave files) from a laptop containing VSound (Fig. 3). VSound took the speed and throttle inputs from the EV and produced the output sound as a continuous emission of the selected sound profile in a speed range of 0 to 20 mph at the desired sound level ($dB(A)_{eq}$) at the external speakers. The frequency modulation and pitch of the output sound varied linearly with vehicle speed in accordance with legislative guidelines (Dalrymple, 2013; MLIT & JASIC, 2010; QRTV, 2012).

Fig. 1: The experiment in the real-world environment. Participant stood at the junction and evaluated the sounds of the EV approaching from behind. A researcher standing beside the participant coordinated the experiment and binaurally recorded ambient and car sounds.

Fig. 2: The electric car used in the experiment in the real-world environment.

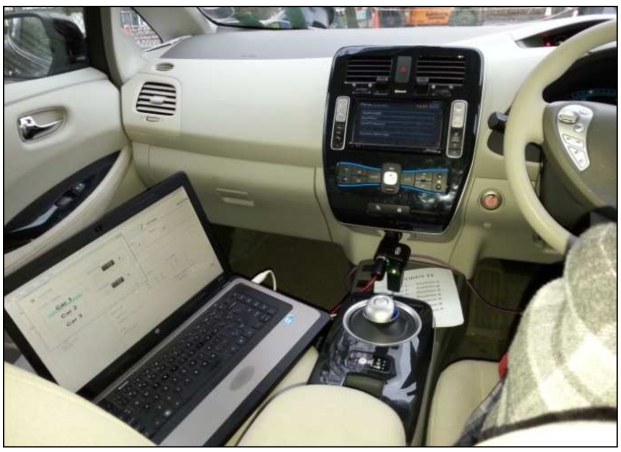

Fig. 3: The sound delivery set-up inside the car in the real-world experiment.

2.2.2. Virtual-world

A virtual-world was created using an Exterior Sound Simulator (ESS) software installed in a soundroom laboratory at the University of Warwick (see Fig. 4) (Singh et al., 2013; Singh, Payne, & Jennings, 2014). ESS, a software tool developed by Brüel and Kjær, simulates a virtual environment from a pedestrian's perspective. It simulated the visuals of a residential road junction that was similar in layout to the junction in the real-world environment. Participant sat on a chair at the centre of eight floor speakers arranged in a regular octahedron. In front of him/her were three adjoining screens where the visuals were projected at resolution of 1280X1024 pixels per screen. Sounds synthesized from ESS were played through the floor speakers by the standard technique of virtual sound source positioning using vector base amplitude panning (Pulkki, 1997). The visuals and sounds were calibrated at the participant's sitting position to correspond to a pedestrian of height 1.6 m standing at the residential junction. Participants therefore experienced vehicles as if they were standing at a real-world junction.

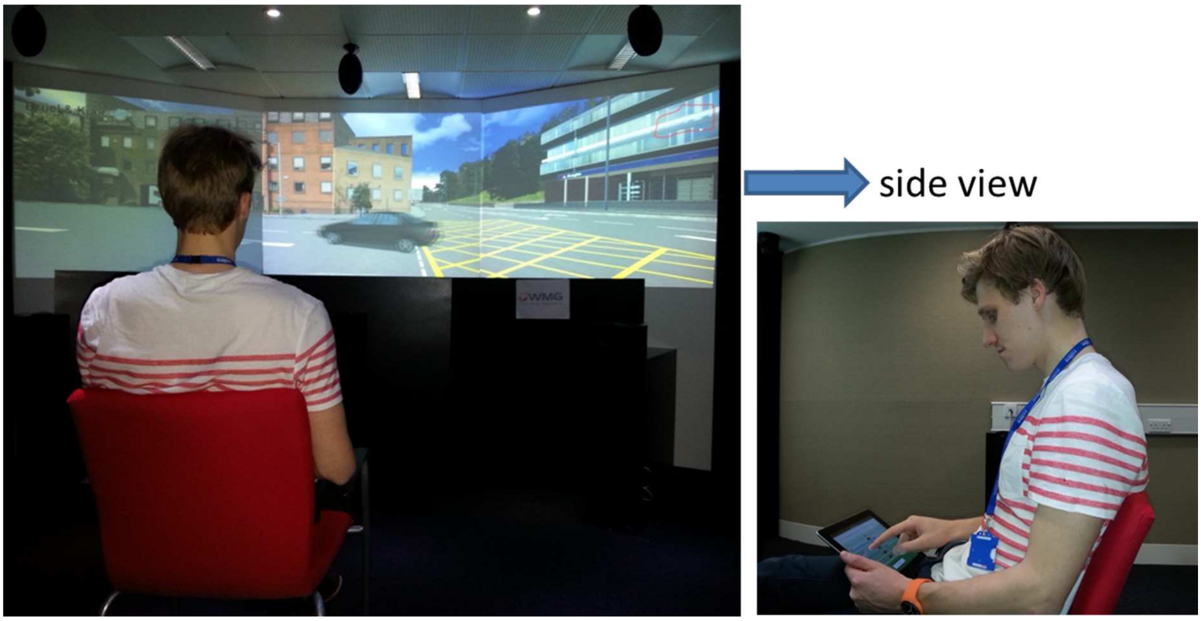

Fig. 4: The experiment set-up for virtual-world environment. The participant, while being exposed to stimuli, evaluated the EV sounds on ESS evaluation interface.

2.3. Stimuli

Visual stimuli: The visual scenario represented a pedestrian standing on the pavement of a residential road junction ("A" in Fig. 5). An electric car, started from one of two different starting positions ("S1" and "S2" in Fig. 5) situated behind the pedestrian on the adjacent parallel road and travelled at 12 mph. This meant the car arrived at the junction 21s or 29.5s from the beginning of a particular experimental condition thus setting the two arrival time conditions.

In the virtual-world, ESS software accurately synchronised the vehicle speed and arrival times. In the real-world the drivers were instructed to try and maintain a speed of 12 mph (note a digital speed dial was used). A GPS data logger recorded the instantaneous speed of the vehicle during every experimental run. Only those experimental conditions where the vehicle speed was 12 ± 1 mph, for at least 90% of that experimental condition, were used for analysis.

Auditory stimuli: Three sounds, denoted as sound 1, sound 2 and sound 3, from an electric car manufacturer were used as the three auditory conditions. The SPL of these sounds was fixed within 57 – 59 dB(A) to comply with the recommended SPL by the AVAS guidelines and FMVSS guidelines (Dalrymple, 2013; MLIT & JASIC, 2010). The loudness of the sounds was between 4.5 – 6.5 sones, sharpness was between 0.4 – 0.8 acum, and the roughness was between 0.0 – 0.1 asper. To comply with the legislation, these sounds were broadband with at least 1 signal in the range 160 – 5000 Hz and did not contain siren, horn, chime, bell, alarm, animal, insect or melodious sounds. The total exterior sound of the target EV comprised mainly of the EV's tire-road interaction sound and the additional sound emitted from its speakers whose level, modulation frequency and pitch varied with speed.

In the real-world environment participants were exposed to the ambient soundscape of the experiment location. The ambient soundscape comprised of wind, occasional birdsong, geese calling, and distant traffic and construction sounds. The ambient soundscape and sounds of the approaching EV for every participant were binaurally recorded during the real-world experimental conditions. The recording (Fig. 1) was made using Brüel & Kjær Sonoscout NVH Recorder - Type 3663. Ambient soundscape recordings were reproduced at the same SPL ($dB(A)_{eq}$) in the virtual-world environment, to match with the corresponding participant and experimental condition during the real-world experiment.

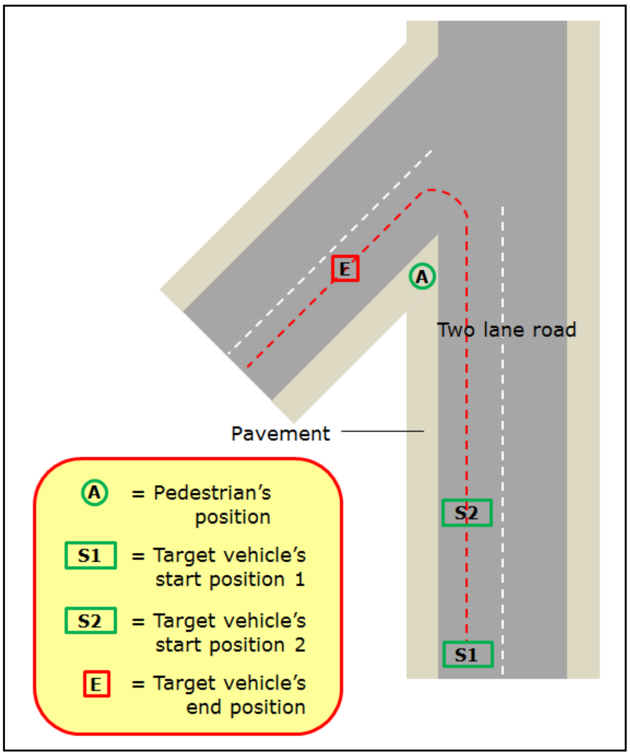

Fig. 5: Schematic of the visual scenario. Red dotted lines indicate the EV's manoeuvre (Singh, Payne, Mackrill, & Jennings, 2014a).

2.4. Measures

2.4.1. Detection

Detection distance: In line with research involving 'quiet' vehicles (Garay-Vega et al., 2010; Hastings et al., 2011), this study used the 'detection distance' to assess the detectability of the target EV sounds. The 'detection distance' is defined here as the distance of the target vehicle from the pedestrian's position at the instance the pedestrian indicates detection. The time difference between the instances of vehicle detection and vehicle's arrival at the junction (time-to-vehicle-arrival) was multiplied with the vehicle's speed, 12 mph, to calculate the detection distance.

In the real-world environment, the participant was asked to press a buzzer on an electronic tablet interface (Fig. 6) as soon as (s)he heard or saw the target EV. Following this the researcher pressed a buzzer as soon as the EV arrived at the junction. These buzzer sounds were heard on the binaural recordings for every experiment and were used to calculate detection distance.

The virtual-world environment used a touch screen evaluation interface that was linked to the ESS software, and synchronized with the experiment condition. The interface had a 7-point semantic scale: "not heard – heard" (Fig. 7) that the participant was asked to slide as soon as (s)he heard or saw the target EV. The interface recorded the time of every instance the scale was moved. If the participant later thought (s)he had incorrectly perceived hearing the car or pressed the buzzer/ moved the scale mistakenly, (s)he was instructed to do this again when the participant thought (s)he started hearing the EV. The detection time was calculated from the last instance the participant pressed the buzzer/ moved the "not heard – heard" scale.

Recognisability: The pedestrian's evaluation of the recognisability of the target EV sound as a vehicle was collected using a 7-point scale of "not recognisable as vehicle – recognisable as vehicle". In the real-world environment a paper questionnaire contained the scale (Fig. 8). In the virtual-world environment the participant recorded the ratings by moving the scale on the ESS evaluation interface to the appropriate value (Fig. 7).

Detectability: The pedestrian's evaluation of the detectability of the target EV sound was collected using a 7-point scale of "not detectable – detectable". This was on the paper questionnaire in the real-world, and on the ESS interface in the virtual-world environment.

2.4.2. Impressions of the vehicle brand

"Powerfulness" and "pleasantness" are well-established perceptual dimensions of vehicle sound quality (Bisping, 1995, 1997; Västfjäll, Gulbol, Kleiner, & Gärling, 2002) that could help understand a listeners' impression of the vehicle brand. Powerfulness and pleasantness of the EV based on listening to its sound was evaluated using 7-point scales of "weak – powerful" and "unpleasant – pleasant" on the paper questionnaire in the real-world, and on the ESS interface in the virtual-world.

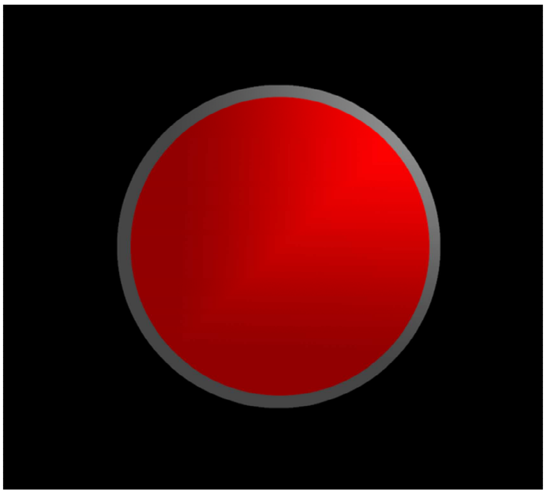

Fig. 6: Buzzer interface used in the real-world environment.

Fig. 7: ESS-linked evaluation interface for virtual-world environment.

Based on your listening to the car's sound, evaluate the car on the following attributes on a scale of 1 to 7.

Car 1:

Not recognisable as vehicle	1	2	3	4	5	6	7	**Recognisable as vehicle**
Not detectable	1	2	3	4	5	6	7	Detectable
Weak	1	2	3	4	5	6	7	Powerful
Unpleasant	1	2	3	4	5	6	7	Pleasant

Fig. 8: Paper questionnaire used in the real-world environment.

2.5. Experimental Design

A repeated measures design was used with the environment (real-world and virtual-world), the target car's arrival time (21s and 29.5s) and the target car's sound (sound 1, 2, or 3) as independent variables. Thus, a 2X2X3 repeated measures design gave 12 different experimental conditions. Within each environment, (real-world and virtual-world) one experimental condition, namely target car emitting sound 1 and arriving at 29.5s, was repeated to check external reliability of the experiment. Therefore, each participant was exposed to 14 experimental conditions (7 experimental conditions per environment). The same presentation order was maintained for each participant during the real-world and the virtual-world environment but order effects were controlled for using 7X7 balanced Latin square (Goodwin, 2010).

2.6. Procedure

The experiment was performed with one participant at a time. Informed consent was obtained from the participant. The participant was briefed about the experiment and was instructed to first detect the EV aurally or visually (whichever was first) and then rate the target EV sounds on the semantic scales. Upon completion of the experimental task, participant was thanked for their participation.

The experiment in the real-world environment was completed first, followed by a two month gap before completing the experiment in the virtual-world environment. The two month gap allowed sufficient time for participants to forget the stimuli. In the real-world environment the researcher and the driver communicated via Bluetooth and confirmed that no other passing vehicles were visible nearby. The driver then selected a sound (1, 2 or 3) from VSound laptop corresponding to the experimental condition, began driving and reached the desired 12 mph speed. As soon as the front of the car approached the corresponding starting position (S1 or S2, Fig. 5) the driver communicated this to the researcher. The researcher then immediately announced to the participant that the experimental condition had begun. In the virtual-world environment these experimental conditions were synchronised using ESS.

3. Results

3.1. External reliability

Repeated-measures-ANOVA found no significant differences in the detection distance, or ratings of recognisability, detectability and powerfulness upon repeating an experimental condition within the real-world and the virtual-world environment (Table 1). Although there was no significant difference in pleasantness ratings upon repeating the experimental condition within the real-world, it significantly differed for the virtual-world. This was considered as an experimental anomaly. As only one of ten results significantly differed, overall, both experiments were considered reliable. The mean data of the repeated experimental conditions were used for further analysis.

Standard deviations for the detection distances were large for the repeated experimental condition. Thus, there was a high variability across the participants for detection distances. However, individual participants themselves were consistent, thus the study was considered to have external reliability.

Table 1: Repeated measures ANOVA results for the repeated experimental condition in each environment.

Environment	Measure	Repetition 1		Repetition 2		F	p	Partial η^2
		Mean	SD	Mean	SD			
Real-world	Detection distance (m)	49.10	24.27	48.51	23.70	.007	.933	.001
	Recognisability	4.14	1.35	4.79	1.42	3.545	.082	.214
	Detectability	4.00	1.66	4.43	1.65	1.721	.212	.117
	Powerfulness	3.00	1.41	3.43	1.40	3.545	.082	.214
	Pleasantness	4.71	1.27	4.93	1.07	.511	.487	.038
Virtual-world	Detection distance (m)	28.01	25.78	24.80	30.30	.191	.669	.014
	Recognisability	3.71	1.64	4.36	1.98	1.918	.189	.129
	Detectability	3.71	1.77	3.86	1.75	.134	.720	.010
	Powerfulness	3.29	1.38	3.71	1.33	1.219	.290	.086
	Pleasantness	4.07	1.21	4.86	0.95	5.026	$.043^*$	.279

alpha = 0.05, df = 1, and df_{error} = 13, $^*p<.05$

3.2. Effect of environment

Table 2 shows the repeated-measures-ANOVA results for effect of evaluation environment and target car's arrival time. Participants detected the target car at significantly larger distances (faster detection) and also found the target sounds significantly more 'recognisable as a vehicle' in the real-world than in the virtual-world. No significant differences were found between the real-world and the virtual-world environment in participants' detectability ratings of the target car's sound nor the powerfulness and pleasantness of the target car.

The significant differences in the detection distances and recognisability were further explored. The difference in the detection distance between the real-world and virtual-world ranged from -89.05 m to 72.74 m and was inconsistent throughout the experimental conditions, mean Pearson's correlation coefficient, $r = .14$, $p > .05$ (Field, 2009). Similarly, the difference in the recognisability between the real-world and virtual-world ranged from -4 to 5 and was inconsistent, mean Pearson's correlation coefficient, $r = .16$, $p > .05$ (Field, 2009). A comparison of the ranking of sounds based on detection distances showed that in both environments sound 2 was detected the fastest compared to sound 1 and sound 3 (Fig. 9). The ranking of sounds based on being recognisable as vehicle were same for both environments, sound 1 being most recognisable followed by sound 3, and sound 2 being the least recognisable (Fig. 10).

3.3. Effect of target car's arrival time

The participants detected the target car at significantly larger distances (faster detection) when the target car arrived at 21 seconds compared to 29.5 seconds. However, the target car's arrival time had no significant effect on participant's rating of the target car sounds' recognisability, detectability or powerfulness and pleasantness (Table 2).

Table 2: Results of repeated measures ANOVA of the independent variables.

Independent Variables	Measure	Level 1		Level 2		F	p	Partial η^2
		Mean	SD	Mean	SD			
Evaluation environment	Detection distance (m)	47.38	2.46	31.96	3.07	11.427	$.005^{**}$	.468
	Recognisability	4.49	0.15	3.78	0.17	7.069	$.020^{*}$	.352
	Detectability	4.42	0.17	4.20	0.14	.481	.500	.036
	Powerfulness	3.51	0.14	3.58	0.14	.185	.674	.014
	Pleasantness	4.70	0.14	3.92	0.15	3.693	.077	.221
Arrival time	Detection distance (m)	43.66	3.13	35.69	2.59	7.126	$.019^{*}$	.354
	Recognisability	4.17	0.17	4.11	0.16	.187	.673	.014
	Detectability	4.27	0.16	4.35	0.16	.201	.661	.015
	Powerfulness	3.50	0.15	3.60	0.14	.556	.469	.041
	Pleasantness	4.26	0.14	4.36	0.15	.229	.640	.017

alpha = 0.05, df = 1, and $df_{error} = 13$, Level 1 = real-world/ 21 s, Level 2 = virtual-world/ 29.5 s, $^{*}p<.05$, $^{**}p<.01$

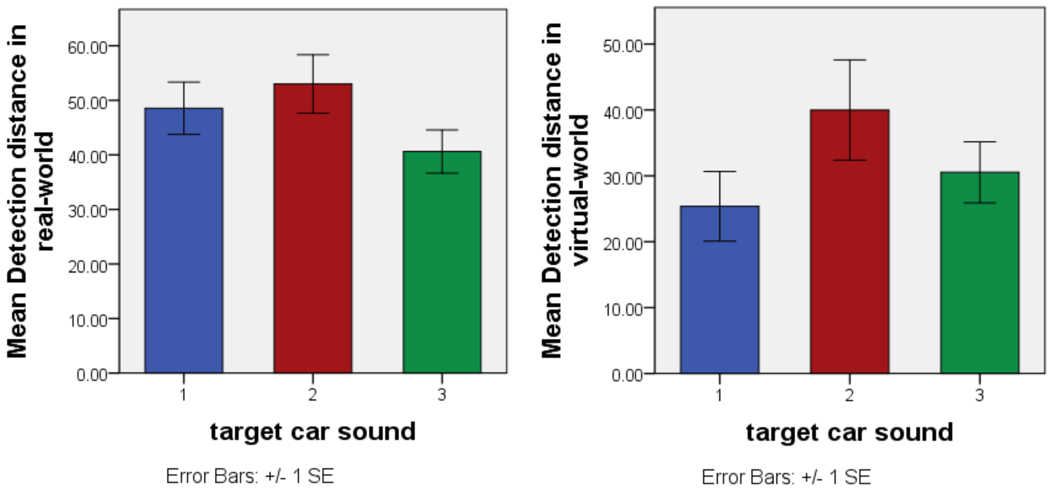

Fig. 9: The target sounds' ranking based on the target car's detection distance.

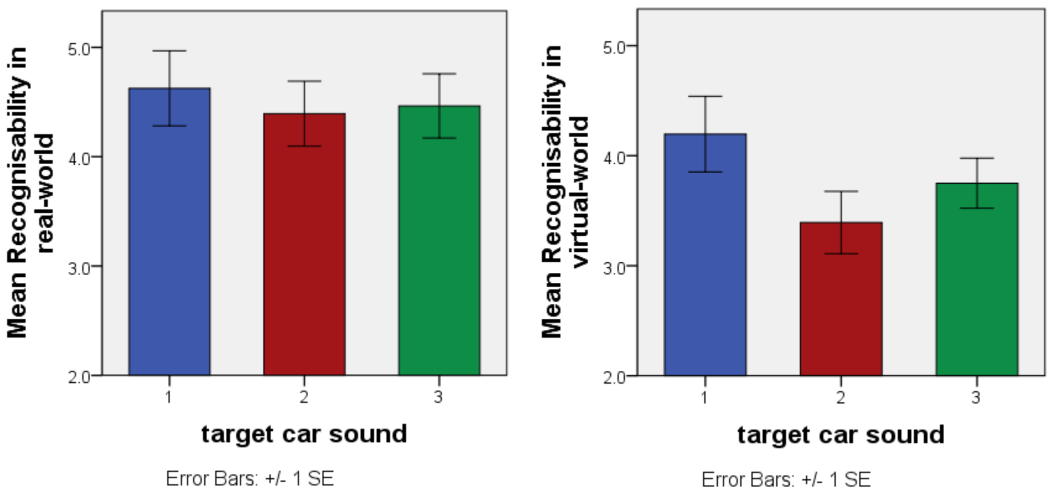

Fig. 10: The target sounds' ranking based on their recognisability as a vehicle

4. Discussions

This study begins to answer a fundamental and most debated question concerning the external validity of simulators and virtual environments: "If the results from experiments in virtual world environments accurately predict or generalise pedestrians' evaluation of EVs emitting sounds in a traffic scenario in the real world?"

It is found that participants evaluated EV sounds whilst being pedestrians in a virtual-word environment in a similar way as when they evaluated the sounds in a real-world environment. However, in the virtual-world environment they found it harder to recognise the EV sounds and took longer to detect them than in the real world. But, they did detect and recognise these sounds in a similar order as in the real-world environment. The results therefore partly support the use of virtual world environments as an equivalent to real-world testing conditions for evaluating EV exterior sounds. The study also highlights EV's arrival time as a key methodological aspect that affects pedestrians' detection rate and makes visual context more realistic; thus it should be manipulated in future experiments.

4.1. Testing external validity of virtual-world environments

Pedestrians' evaluation of the detectability of the target car's sound and the powerfulness and pleasantness of the target car in a virtual-world environment did not significantly differ to a comparable scenario in a real-world environment. This suggests participants were responding similarly in the virtual world as in the real world.

Recognisability of the EV sounds, however, was significantly higher in the real-world than in the virtual-world environment. This means that the same simulated sounds seem more recognisable in the real-world than in a virtual-world. Another explanation is that by the time participants made their recognisability ratings they had seen a real EV emitting these sounds in the real-world thus increasing their association of these sounds to a vehicle. A person without an automotive background is unlikely to know all of the numerous automotive sounds in existence and thus unable to rate a set of sounds on an absolute scale. Therefore, process of evaluating automotive sound quality on perceptual attributes is essentially a process of providing relative ratings to a set of candidate vehicle sounds (Otto, Amman, Eaton, & Lake, 1999). Similarly, in this study, a pedestrian provides relative scores on the perceptual attributes, based on the vehicle sounds they have been exposed to previously, or during the experiment. Therefore, though less important than actual scores of emotional evaluations, the ranked order of sounds being evaluated is valuable information which the virtual-world experiment accurately predicted for recognisability of sounds in the real-world. It is worth noting that this experiment used three very acoustically similar sounds (see section 2.3) developed for a single EV brand. On the contrary, many previous studies have used a very diverse set of sounds such as engine, melodious, bell, pure tones and nature sounds (Goodes et al., 2009; Hastings et al., 2011; Singh, Payne, & Jennings, 2014), or sounds from different brands (Barton et al., 2012). With an extremely diverse set of sounds it is easier for a method to predict their ranked order or ratings in the real world. However, this study differentiates and predicts ranked order and ratings of very similar sounds, which highlights the efficiency of the methodology.

Overall, it can be said that experiments conducted in the virtual-world using presented methodology will produce results that effectively predict pedestrians' emotional evaluations of vehicle exterior sounds in real-world conditions. However, current evaluation tests of EV sounds (Barton et al., 2012; Emerson et al., 2013, 2011; Garay-Vega et al., 2010; Goodes et al., 2009; Hastings et al., 2011; JASIC, 2009; Parizet et al., 2014) do not combine detection tests with assessing a listener's emotional responses to EV exterior sounds. Including emotional evaluations in EV exterior sound studies will make them similar to the focus of sound quality evaluation tests, usually conducted for interior sounds, in a vehicle's design process (Bisping, 1995, 1997; Cerrato, 2009; P. A. Jennings et al., 2010; Miśkiewicz & Letowski, 1999; Otto et al., 1999). In that way, this methodology can be directly integrated within automotive industries' vehicle design and development processes.

In this study, the EV was detected at a significantly greater distance by a pedestrian in a real-world scenario than in a comparable virtual-world environment. This suggests that virtual–world environments may not be accurate at representing how fast pedestrians react to and detect a vehicle in the real world. However, there were two potential human-related errors in the real-world condition that were absent in the virtual-world condition and these could have affected the measured values of detection distances. Firstly, in the real-world condition there was the potential for "operator's manual control error" (Wickens & Hollands, 2000) if the driver of the EV deviated from driving the vehicle at 12 mph, and four different drivers were used during the study. If we take the mean detection distance observed as 47.38 m (see Table 2), then an error of 1 mph speed deviation would have caused a detection distance error of 3.95 m. In contrast, in the virtual-world condition the EV was 'driven' by the software at a constant 12 mph, to ensure the car arrived exactly at one of two ‘arrival time’ conditions (21s or 29.5s). Secondly, in the real-world condition there was the potential for “human observer measurement error” (Goto & Mascie-Taylor, 2007) as the driver needed to verbally state when the car crossed a given line (experiment’s starting position), and the researcher needed to state when the car had arrived at a given point by pressing a buzzer. Here, even an error of 1 s in researcher’s or driver’s observation would have caused a detection distance error of 5.4 m each. Whilst in the virtual-world, ESS software accurately monitored these two instances thereby eliminating measurement discrepancies.

Different methods had to be used for this detection distance measurement as the ESS data collection facility is not available for use in the real-world. Both of the discussed errors are by definition random errors. When the differences in the detection distances were analysed individually it was found that they too are random errors as they are inconsistent throughout the experimental conditions (low correlation) and bi-directional (negative and positive errors). Thus, the differences in the detection distances are likely to be caused by random errors such as human-related measurement errors, and/or effect of uncontrolled external factors such as weather, lighting, traffic, etc. Also they are not likely to be caused due to a problem in the presented methodology because then the differences would be systematic across all conditions.

Therefore, the study suggests that people react faster and are more aware of a vehicle approaching in the real-world than in the virtual-world. However, these differences might be due to error in measuring their detection rate rather than people responding differently due to the different environments. Despite the significant differences in participants’ reaction time to an EV approaching in the real and virtual-world environments, the ranked order of EV sounds based on their detection distance was same in both environments. This suggests that the results from the virtual-world environment are still generalizable to the real-world, and thus supports the developed methodology.

4.2. Effect of arrival time on pedestrians’ detection rate

Pedestrians took longer to detect the target car, responding at a slower rate when the car arrived later in time. This supports the relationship between the target car’s arrival time and pedestrians’ detection rate in a previous study (Singh, Payne, & Jennings, 2014). This implies that the attention level of pedestrians reduced with the passage of time within an experimental condition. In most conventional listening tests, the car sound to be detected is present from the very beginning of the stimulus, and has a fixed arrival time at the pedestrian’s spot (Garay-Vega et al., 2010; JASIC, 2009; Morgan et al., 2011; Parizet et al., 2014), so after a few trials, participants begin to expect to hear the car straight away. Therefore, participants may pay more attention towards hearing the car. These conventional tests do not represent a real-world pedestrian-vehicle interaction where a car may approach a crossroad/junction at any time. Thus, conventional listening test methods may produce results different from a real-world traffic scenario.

On the other hand, this methodology involves randomly altering the target car’s arrival time throughout the experiment, just as it would be in a real-world scenario especially at un-signalised crossroads and junctions. Thus, it enables participants to think, react and pay similar attention as a pedestrian in the real-world who is unsure of the time of a car’s approach and his/her expectation will not be as evident. Thus with time the participant’s attention may shift from focussing on detecting the target car’s sound, to the perception of other stimuli such as, visual and ambient sounds. Reduced expectations and decreased attention may have caused the participants to react slower towards the end of a particular experimental condition. The same is also expected in a real-world scenario. Therefore, varying the target car’s arrival time is recommended for future studies.

4.3. Virtual-world versus real-world: a methodological comparison

It is usually critiqued that real-world experiments or field studies have much lower reliability compared to a laboratory study (P. Jennings et al., 2007; Singh, Payne, & Jennings, 2014). Yet, the methodology used in this study has successfully achieved protocols for conducting experiments in a virtual-world environment and in the real-world that have significant external reliability. However, a large variability was observed in the detection distances across individual participants for the repeated experimental condition (Table 1). The sound used for this condition was too low that many participants only heard the tire rolling before the actual sound from the speaker. The variability in the detection distances could be due to the ambiguity of the sound used for this particular condition. No such variability was observed across individual participant scores for the overall experiment (low standard deviation in Table 2). Experiments in the virtual-world environment were quicker compared to the real-world environment, taking only 8 minutes per participant for completion of the seven experimental conditions. In contrast, in the real-world, the completion of these seven experimental conditions took between 30 to 45 minutes per participant, excluding the time to arrive or leave the experiment site. This was because of the periodic interruptions by the University approved vehicles passing the experiment location and time taken to achieve the car's desired manoeuvre before beginning a new experimental condition. Furthermore, the virtual-world experiments for all 14 participants were completed in a one week. In comparison, the real-world experiments took one month to complete because of implementation difficulties. The real-world experiments were cancelled on two occasions due to problems with vehicle charging and tyre puncture, thus causing a ten day delay. Moreover, one participant's data was not used due to heavy winds that interfered with car driving and binaural recording. Initially, it was decided to maintain a low speed condition of 10 mph for experiments just as in the previous study (Singh et al., 2013; Singh, Payne, & Jennings, 2014). But the pilot study showed that in the real-world drivers found 10 mph difficult to maintain therefore 12 mph were chosen for actual experiments with a speed tolerance of ± 1 mph (note a digital speed dial was used).

Overall we can say that methodologically virtual environments seem a preferred alternative to real-world studies as they are quick, easy to implement and provide better experimental control. Additionally, virtual environments allow easy manipulation of factors such as vehicle's arrival time, direction (Singh, Payne, & Jennings, 2014), weather, and ambient conditions, which is difficult to achieve in the real-world and also in conventional laboratory listening tests.

5. Limitations and Future Work

The current study does not accurately predict the distance at which pedestrians detect an EV in the real-world. As discussed, these differences in results may be due to inaccurate detection measurement and uncontrolled external factors present in the real-world.

Currently, all real-world detection studies have limitations in implementing an accurate measurement method for vehicle detection, as it involves a trade-off between accuracy, cost, feasibility, and installation issues. The measurement method for virtual worlds in this study balances feasibility, accuracy, and cost. Other real-world detection studies have primarily used the inaccurate, but economical, methods such as video recordings of the experiment with road markings to estimate the detection distance (Emerson et al., 2013; Goodes et al., 2009; Hastings et al., 2011). More accurate methods such as monitoring vehicle's position with photoelectric sensors, marking instances of detection using push buttons, and storing the data in a synchronized data acquisition software, are relatively costly and difficult to implement (Hastings et al., 2011). These methods could have their own errors and problems. The best method for collecting detection distance in the real-world therefore needs further investigation. Hopefully, future studies in the real-world could be improved, but currently, virtual world environments seem more usable for doing such studies.

6. Conclusions

This study applies a methodology where people can evaluate EV exterior sounds as a pedestrian in a virtual-world traffic scenario. The study confirms that the developed virtual world methodology produces results that accurately predict pedestrians' real-world evaluations of detectability of EV sounds and their real-world impression of the powerfulness and pleasantness of the vehicle brand. The results also generalise how recognisable pedestrians find these sounds in the real-world. Arguably, for comparable methods and setups, experiments in virtual worlds would effectively predict pedestrian's evaluation of EV sounds in the real world. As the methodology randomly alters vehicle's arrival time over the experimental conditions, just like in a real world scenario this helps participants think, respond and pay similar attention as a real-world pedestrian. Therefore, this virtual world methodology can help transportation and traffic safety researchers, manufacturers, and the legislators of new EV sounds to understand how pedestrians in the real-world detect and evaluate a vehicle, emitting new sounds.

Acknowledgements

This work is part of a doctoral project funded through WMG – IIT Kharagpur collaboration. Thank you also to Brüel and Kjær and Nissan, UK for supporting the project.

References

Ashmead, D. H., Grantham, D. W., Maloff, E. S., Hornsby, B., Nakamara, T., Davis, T. J., … Rushing, E. G. (2012). Auditory Perception of Motor Vehicle Travel Paths. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(3), 437–453.

Barton, B. K., Ulrich, T. A., & Lew, R. (2012). Auditory detection and localization of approaching vehicles. *Accident Analysis & Prevention*, *49*, 347–353. doi:10.1016/j.aap.2011.11.024

Bisping, R. (1995). Emotional Effect of Car Interior Sounds: Pleasantness and Power and Their Relation to Acoustic Key Features. *SAE International*, Paper No. 951284. doi:10.4271/951284

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

Brewer, M. B., & Crano, W. D. (2014). Research Design and Issues of Validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed., pp. 11–26). NY: Cambridge University Press.

Cerrato, G. (2009). Automotive Sound Quality –Powertrain, Road and Wind Noise. *Sound & Vibration*, *43*(4), 16–24.

Chan, E., Pradhan, A. K., Pollatsek, A., Knodler, M. a, & Fisher, D. L. (2010). Are driving simulators effective tools for evaluating novice drivers' hazard anticipation, speed management, and attention maintenance skills? *Transportation Research. Part F, Traffic Psychology and Behaviour*, *13*(5), 343–353. doi:10.1016/j.trf.2010.04.001

Cohen, J. (1988). The concepts of power analysis. In *Statistical power analysis for the behavioral sciences* (2nd ed., pp. 1–17). Hillsdale, New Jersey; London, England: Lawrence Erlbaum Associates.

Dalrymple, G. (2013). *MINIMUM SOUND REQUIREMENTS FOR HYBRID AND ELECTRIC VEHICLES: Draft Environmental Assessment*. Washington D.C.

Emerson, R. W., Kim, D. S., Naghshineh, K., Pliskow, J., & Myers, K. (2013). Detection of Quiet Vehicles by Blind Pedestrians. *Journal of Transportation Engineering*, (January), 50–56. doi:10.1061/(ASCE)TE.1943-5436.0000478.

Emerson, R. W., Naghshineh, K., Hapeman, J., & Wiener, W. (2011). A Pilot Study of Pedestrians with Visual Impairments Detecting Traffic Gaps and Surges Containing Hybrid Vehicles. *Transportation Research. Part F, Traffic Psychology and Behaviour, 14*(2), 117–127. doi:10.1016/j.trf.2010.11.007

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Field, A. (2009). Correlation. In *Discovering Statistics Using SPSS* (3rd ed., pp. 166–196). London: SAGE Publications Limited.

Garay-Vega, L., Hastings, A., Pollard, J. K., Zuschlag, M., & Stearns, M. D. (2010). *Quieter Cars and the Safety Of Blind Pedestrians : Phase I*. doi:DOT HS 811 304

Goodes, P., Bai, Y. B., & Meyer, E. (2009). Investigation into the Detection of a Quiet Vehicle by the Blind Community and the Application of an External Noise Emitting System. *SAE Technical Paper 2009-01-2189*. doi:10.4271/2009-01-2189

Goodwin, C. J. (2010). Control Problems in Experimental Research. In *Research in Psychology: Methods and Design* (6th ed., pp. 205–232). Hoboken, NewJersey: John Wiley & Sons Ltd.

Goto, R., & Mascie-Taylor, C. G. N. (2007). Precision of Measurement as a Component of Human Variation. *Journal of PHYSIOLOGICAL ANTHROPOLOGY*, *26*(2), 253–256. doi:10.2114/jpa2.26.253

Hanna, R. (2009). *Incidence of Pedestrian and Bicyclist Crashes by Hybrid Electric Passenger Vehicles*. doi:DOT HS 811 204

Hastings, A., Pollard, J. K., Garay-Vega, L., Stearns, M. D., & Guthy, C. (2011). *Quieter Cars and the Safety of Blind Pedestrians , Phase 2 : Development of Potential Specifications for Vehicle Countermeasure Sounds-Final Report*. Washington D.C. doi:DOT HS 811 496

JASIC. (2009). A Study on Approach Warning Systems for hybrid vehicle in motor mode. Retrieved from http://www.unece.org/fileadmin/DAM/trans/doc/2009/wp29grb/ECE-TRANS-WP29-GRB-49-inf10e.pdf

Jennings, P. A., Dunne, G., Williams, R., & Giudice, S. (2010). Tools and techniques for understanding the fundamentals of automotive sound quality. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 224*(10), 1263–1278. doi:10.1243/09544070JAUTO1407

Jennings, P., Giudice, S., Fry, J., Williams, R., Allman-ward, M., & Dunne, G. (2007). Developing Best Practice for Sound Evaluation using an Interactive NVH Simulator. *Transactions of Society of Automotive Engineers of Japan*, *38*(1), 31–36.

Miśkiewicz, A., & Letowski, T. (1999). Psychoacoustics in the automotive industry. *Acta Acustica United with Acustica*, *85*(5), 646–649.

MLIT, & JASIC. (2010). Guideline for Measure against Quietness Problem of HV , etc . Retrieved from http://www.unece.org/fileadmin/DAM/trans/doc/2010/wp29grb/QRTV-03-01e.pdf

Morgan, P. A., Morris, L., Muirhead, M., Walter, L. K., & Martin, J. (2011). *PUBLISHED PROJECT REPORT PPR525: Assessing the perceived safety risk from quiet electric and hybrid vehicles to vision-impaired pedestrians*. doi:PPR 525

Otto, N., Amman, S., Eaton, C., & Lake, S. (1999). Guidelines for Jury Evaluations of Automotive Sounds. *SAE Technical Paper 1999-01-1822*. doi:10.4271/1999-01-1822

Özcan, E. (2014). THE HARLEY EFFECT : INTERNAL AND EXTERNAL FACTORS THAT To love or not to love machinery sounds. *Journal of Sonic Studies*, *6*(1), a07. Retrieved from http://journal.sonicstudies.org/vol06/nr01/a07

Parizet, E., Ellermeier, W., & Robart, R. (2014). Auditory warnings for electric vehicles: Detectability in normal-vision and visually-impaired listeners. *Applied Acoustics*, *86*, 50–58. doi:10.1016/j.apacoust.2014.05.006

Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, *45*(6), 456–466.

QRTV. (2011). *PROPOSAL FOR GUIDELINES ON MEASURES ENSURING THE AUDIBILITY OF HYBRID AND ELECTRIC VEHICLES* (Vol. 1). Informal document GRB-53-09-Rev.1, Presented at 53rd GRB (UNECE's Working Party on Noise), 15-17 February 2011, agenda item 10. Retrieved from http://www.unece.org/trans/main/wp29/wp29wgs/wp29grb/grbinf53.html

QRTV. (2012). *Draft Recommendations for a Global Technical Regulation Regarding Audible Vehicle Alerting Systems for Quiet Road Transport Vehicles*. Informal document GRB-55-14, Presented at 55th GRB (UNECE's Working Party on Noise), 7-9 February 2012, agenda item 10. Retrieved from http://www.unece.org/trans/main/wp29/wp29wgs/wp29grb/grbinf55.html

Singh, S., Payne, S. R., & Jennings, P. A. (2013). Detection and emotional evaluation of an electric vehicle ' s exterior sound in a simulated environment. In *Internoise 2013; Paper 0807* (pp. 1–9). Innsbruck, Austria.

Singh, S., Payne, S. R., & Jennings, P. A. (2014). Toward a Methodology for Assessing Electric Vehicle Exterior Sounds. *IEEE Transactions on Intelligent Transportation Systems*, *15*(4), 1790–1800. doi:10.1109/TITS.2014.2327062

Singh, S., Payne, S. R., Mackrill, J. B., & Jennings, P. A. (2014a). Evaluation of electric vehicle exterior sounds in virtual and real-world environments – a comparative study. In *Aachen Acoustics Colloquium 2014* (pp. 82–89). Aachen, Germany.

Singh, S., Payne, S. R., Mackrill, J. B., & Jennings, P. A. (2014b). Using virtual environments as a tool to evaluate electric vehicle sounds. In *5th IET Hybrid and Electric Vehicles Conference (HEVC 2014)*. London: IET. doi:10.1049/cp.2014.0948

Västfjäll, D., Gulbol, M. -a., Kleiner, M., & Gärling, T. (2002). Affective Evaluations of and Reactions to Exterior and Interior Vehicle Auditory Quality. *Journal of Sound and Vibration*, *255*(3), 501–518. doi:10.1006/jsvi.2001.4166

Wickens, C. D., & Hollands, J. G. (2000). Manual Control. In *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.