Applied Acoustics 201 (2022) 109136

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Development of electric scooter alerting sounds using psychoacoustical metrics

Tim Walton *, Antonio J. Torija, Andrew S. Elliott

Acoustics Research Centre, University of Salford, The Crescent, Manchester M5 4WT, United Kingdom

ARTICLE INFO

Article history:
Received 11 August 2022
Received in revised form 23 September 2022
Accepted 17 November 2022

Keywords:
Electric scooter
Micromobility
Acoustic vehicle alerting system
Sound quality metrics
Detection
Annoyance

ABSTRACT

In recent years electric micromobility transportation, including electric scooters, has seen a surge in popularity due to technological advances and the move to lower emission transport. Although offering a range of societal benefits, such as reduced pollution and increased personal mobility, concerns have been raised regarding the implications for pedestrian safety, most notably within the blind and partially sighted community. The issue of pedestrian safety is well studied in the context of larger electric vehicles (EVs), and indeed regulations are now in place that specify mandatory Acoustic Vehicle Alerting Systems (AVAS) for such vehicles. However, limited research has been done on the development of acoustic alerting systems for micromobility. In this paper, the development of an electric scooter (e-scooter) AVAS is considered by taking a perception-influenced design approach to designing alert sounds that optimise detectability and annoyance. A listening experiment has been conducted using ambisonic soundscapes and simulated auralisations of e-scooter passes at 20 km/h, in which a detection-based task and annoyance rating task were conducted. Objective metrics for detectability and annoyance were subsequently evaluated in relation to the subjective responses, so as to enable a more focused approach to the development of alert sounds. Results show that without additional alert sounds, the rate of detection for e-scooters in a soundscape of 60 dBA is as low as 23%. Regression analysis showed that the objective metric of Zwicker's psychoacoustic annoyance is a useful predictor of subjective annoyance for AVAS sounds, with a coefficient of determination of $R^2 = 0.96$, and explains more variance than other metrics previously reported in the literature. Partial loudness was also studied as a predictor of detectability, with strong positive association seen ($R^2 \approx 0.9$). Of the alert sounds evaluated, those comprising pure tones with frequency content in the 800 Hz - 1 kHz range, and with amplitude modulation or impulsive characteristics, offered the greatest balance between detectability and annoyance. This study offers much needed research into detectability of electric micromobility transport in a range of environmental noise conditions, and furthermore provides objective metrics for the development of micromobility AVAS sounds going forward.

1. Introduction

Electric scooters (e-scooters) are becoming an increasingly common sight on our streets; as of 2022, there were an estimated 520,000 shared e-scooters across Europe, up from 360,000 in 2021 [1,2]. This increase in popularity can be seen across the electric micromobility[1] sector as a whole, as falling battery prices, improvements in energy density and the move to zero emission transport begin to influence how we move people and goods around our cities [3].

Whilst the take up of electric micromobility has potential advantages, such as the move to 'clean' transport and increased mobility, it is not without its challenges. For example, in a UK Department for Transport survey on perceptions of current and future e-scooter use in the UK, 53% of respondents cited safety issues as one disadvantage of e-scooters [4]. Moreover, UK Government national statistics on road traffic collisions involving e-scooters revealed there were 1,359 casualties in collisions involving e-scooters in 2021, compared to 484 in 2020 [5]. These statistics are based on the definition of an 'e-scooter' as given by the UK Government [6] and this is the definition to which we refer to in this paper. This definition distinguishes e-scooters from other

* Corresponding author.
E-mail addresses: t.walton1@salford.ac.uk (T. Walton), a.j.torijamartinez@salford.ac.uk (A.J. Torija), a.s.elliott@salford.ac.uk (A.S. Elliott).

[1] Defined as "The use of small mobility devices, designed to carry one or two people, or 'last mile' deliveries" [3]

https://doi.org/10.1016/j.apacoust.2022.109136
0003-682X/

two-wheeled electric vehicles by specifying a maximum speed limit of 12.5 mp/h (20 km/h), a mass not exceeding 35 kg and designed to carry one person in a standing position with no provision for seating.

One particular concern of electric vehicles (EVs) and micromobility transportation, is that they pose a challenge to pedestrians, specifically those who are blind and partially sighted. This has been researched in the context of EVs for over a decade [7–11], however has only recently been considered in the context of micromobility [12]. Previous research has highlighted that the risk of road traffic near-misses and accidents involving pedestrians was around 25% more likely when comparing EVs with no alert sounds to internal combustion engine vehicles (ICEVs) due to their inherently lower noise levels [13]. The National Highway Traffic Safety Administration (NHTSA) of the United States conducted early research on the topic of quieter cars and the safety of blind pedestrians [8,9], which resulted in initial specifications for Acoustic Vehicle Alerting Systems (AVAS). There are now a range of regulations in place outlining minimum sound requirements for quiet running vehicles [14–16], such as the United Nations Economic Commission for Europe (UNECE) Regulation 138. In UNECE Regulation 138, minimum AVAS sound levels are specified in one-third octave bands between 160 Hz and 5 kHz, with compliant sounds requiring minimum levels in at least two of the specified bands and with one of them below or within the 1600 Hz one-third octave band [14]. The speed range of required AVAS operation is greater than 0 km/h up to and inclusive of 20 km/h, as above this speed, sound level measurements show little difference between EVs and internal combustion engine vehicles due to an increased contribution from rolling noise and wind noise [10]. Furthermore, the regulation specifies that the AVAS varies proportionally with speed by an average of at least 0.8% per 1 km/h in the speed range from 5 km/h to 20 km/h inclusive when driving in forward direction, so as to alert pedestrians to changes of speed.

Research investigating detectability of smaller electric vehicles is more limited. Sekine et al. [17] investigated the detectability of Electric Motorbikes (EM) and Internal Combustion Engine Motorbikes (ICEM) operating at speeds of 10 km/h and 20 km/h. It was found that the travelling sound of the EM was lower than that of the ICEM by approximately 15 dB at a given speed, and this led to a reduction in detectability distance from 57.9 m for the ICEM to 11.7 m for the EM at 20 km/h. Torija et al. [12] recently presented results from a feasibility study looking at developing a system to generate an awareness sound for e-scooters. A virtual reality experiment was conducted to evaluate pedestrian awareness of an approaching e-scooter with and without additional alert sounds. Initial results indicate that with an additional alert sound, detectability distance was increased by 3.2 m, however it was noted that further research is needed to design awareness sounds with an optimal balance between noticeability and annoyance.

Specific acoustic features of AVAS sounds have previously been researched in the context of detectability and annoyance [18–23]. As part of the eVADER project (electric Vehicle Alert for Detection and Emergency Response), Parizet et al. [18] aimed to develop alert sounds that were readily detectable, but without causing annoyance. Sounds were developed with a range of variables, including number of harmonics, amplitude modulation (AM), and frequency modulation (FM), and the effectiveness of these were evaluated with a detection task in a simulated pass-by scenario. It was found that reaction times were shortest when the alert sounds had a low number of harmonics, when FM is absent, and when AM is prominent and irregular. An extension of this study [19] investigated how the effectiveness of each alert sound was related to its perceived unpleasantness, in which it was found that high levels of detectability were correlated to high levels of unpleasantness. The trade-off between detectability and acceptance for AVAS sounds was also considered by Lee et al. [21], whilst also considering the masking effect of background noise. As in [18], an amplitude modulated signal produced the best performance in terms of both annoyance and detectability.

Objective metrics that predict detectability and annoyance of AVAS sounds are desirable as they would allow the development of effective sounds without the need for resource intensive listening experiments. Broadband sound pressure levels, such as L_{Aeq} and L_{AFmax}, offer one type of metric, however they overlook the intricacies of human auditory perception, and as such, additional perception based metrics are needed. Examples of using perception based metrics for the development of AVAS include partial loudness [20], annoyance index [24], whine index [21], spectral flatness and modulation rate [25], and other sound quality metrics (SQMs), e.g. roughness, sharpness and tonality [22].

The study presented here builds on the existing literature by taking a perception-influenced design approach to designing e-scooter alert sounds that optimise detectability and annoyance. Moreover, a range of objective metrics are considered for the development of micromobility AVAS sounds. Specifically, the metric of Zwicker's psychoacoustic annoyance is analysed in correlation with AVAS sound performance, which provides an original contribution to the field. Zwicker's model of psychoacoustic annoyance (PA) combines measures of loudness, sharpness, fluctuation strength and roughness and produces an output of annoyance that can be used to compare different sounds [26]. The metric of PA has also recently been used to analyse human response to drone noise [27]. Furthermore, partial loudness as a predictor of AVAS sound performance for electric micromobility is investigated. To achieve this, a listening experiment has been conducted using a three-dimensional loudspeaker array, including ambisonic soundscapes and simulated e-scooter passes, in which detection and annoyance rating tasks were completed.

The remainder of the paper is organised as follows. Section 2 describes the experimental setup for the listening experiment including baseline measurements of e-scooter passes, Section 3 presents and discusses the experimental results including a description of the objective metrics used, followed by the main conclusions in Section 4.

2. Methodology

This listening experiment was conducted as a laboratory-based study at The University of Salford. Ambisonic soundscapes were reproduced over a three-dimensional loudspeaker array to simulate presence in different acoustic environments and acoustic simulations of passing e-scooters with a variety of added alert sounds were synthesised within the three-dimensional space. Participants were required to complete a detection task as well as an annoyance rating task, with further details provided in the following sections.

2.1. Apparatus

The experiment took place within a listening room at the university, with the setup consisting of 16 Genelec 8030A loudspeakers; 8 loudspeakers are located in the horizontal plane (positioned at azimuths $+0^\circ, \pm45^\circ, \pm90^\circ, \pm135^\circ, +180^\circ$), 4 loudspeakers at $\pm39^\circ$ elevation (positioned at azimuths $\pm45^\circ, \pm135^\circ$), and 4 loudspeakers at -39° elevation (positioned at azimuths $\pm45^\circ, \pm135^\circ$). The loudspeakers in the horizontal plane are at a distance of 1.26 m from the centre of the array, whereas the loudspeakers at $\pm30^\circ$ elevation are at a distance of 1.54 m from the centre of the array. An RME MADIface XT audio interface was used with an RME M-32 DAC to drive the loudspeakers. The loudspeakers were time-aligned and level-aligned to produce the same A-weighted equivalent sound level (L_{Aeq})($\pm$0.5 dB) for a pink noise signal at the central listening position. The reverberation time in the room is approximately 0.1 s.

The graphical user interface was reproduced via a 42.5 inch display mounted on the front wall of the room. Data entry was via a high contrast keyboard and a standard optical computer mouse. The experiment was administered via Cycling '74 MAX/MSP software.

2.2. Stimuli and design

2.2.1. Acoustic Features

A range of e-scooter alert sounds were synthesised using the software MAX/MSP and were designed to incorporate the pertinent psychoacoustic AVAS features as outlined in previous research. In order to investigate the effect of different acoustic features, five main components of the sound were considered at two different levels and these design factors are summarised in Table 1. As the simulation involved e-scooter passes at one speed only, these acoustic features were not speed dependent.

Each sound is based on a fundamental frequency plus two harmonics, as AVAS tones comprising a small number of harmonics have previously been shown to reduce reaction times [18]. Component C1 controls the frequencies of the fundamental and associated harmonics and component C2 controls the wave type of these tones. Component C3 controls the amplitude modulation depth, component C4 controls the amplitude modulation rate, and component C5 controls the synchrony of the amplitude modulation; for level 1 the amplitude modulation of the fundamental and harmonics is in sync, whereas for level 2 the amplitude modulation of the fundamental is 0.5 Hz below the specified AM rate, and the amplitude modulation of the second harmonic is 0.5 Hz above the specified AM rate, so as to produce an asynchronous modulation effect.

A full factorial design of these factors would result in 32 different stimuli (2^5), which was considered impractical for the desired length of experiment. As such, a fractional factorial design was implemented with a design specification of 2_{III}^{5-2} , i.e. a five factor design of resolution III. In resolution III designs, no main effects are aliased with any other main effect, but main effects are aliased with two-factor interactions, and they are often used for screening of important factors [28]. Table 2 outlines the factor combinations used, as taken from [28]. In addition to the eight stimuli produced from the five components in Table 1, two additional stimuli were included. S9 consisted of an impulsive sound based around a 1 kHz tone at a rate of 7 Hz and S10 consisted of a baseline e-scooter sound with no alert sound added, as further discussed in the following section.

Table 1
Levels of acoustic features used for stimuli generation.

Factor	Level 0	Level 1
C1 - Frequency	120 Hz (0 dB), 180 Hz (-3 dB), 240 Hz (-3 dB)	400 Hz (0 dB), 600 Hz (-3 dB), 800 Hz (-3 dB)
C2 - Wave type	Sine	Saw
C3 - AM depth	0.5	1
C4 - AM rate	4 Hz	8 Hz
C5 - AM synchrony	Synchronous	Asynchronous, ±0.5 Hz

2.2.2. Baseline measurements

Audio recordings and sound level measurements of e-scooter operations were undertaken to provide baseline audio and to enable accurate calibration of pass-by levels. On-scooter audio was captured with a class 1 sound level meter (B&K 2250) including windshield, whilst the e-scooter was travelling at 20 km/h on smooth asphalt, with the microphone positioned approximately 1 m from the edge of the scooter and 1.7 m from the ground. This represents the position of a pedestrian and includes any contribution from tyre noise, aerodynamic noise and motor noise. The audio was monitored for wind noise and the windshield used was deemed sufficient to prevent wind noise interference within the audio clip. As the recording was taken on the e-scooter, the audio was relatively constant and could be subsequently processed with the pass-by simulation code. A loop of this recording was used as stimulus S10, and furthermore, this loop was mixed with stimuli S1 to S9 so as to include tyre-road interaction noise and to provide a realistic pass-by sound for the alert sounds. The loop was based on an 8 s section of audio, which was chosen to limit perception of periodicity when repeated, and furthermore the 8 s samples were cross-faded with each other over 1.5 s to reduce perception of connecting points.

Sound level measurements were undertaken with the e-scooter passing 1 m distance from the sound level meter at 10, 15 and 20 km/h; the third-octave frequency spectrum of these is presented in Fig. 1. The 20 km/h pass-by resulted in an overall broadband maximum sound level (L_{AFmax}) of 52 dBA, when propagated to 2 m distance, and this level was used during the calibration stage.

2.2.3. Simulation

The stimuli were processed to obtain a realistic pass-by scenario, with geometry as outlined in Fig. 2. The pass-by simulation was implemented using the SPAT 5 package for MAX/MSP [29], and comprised the following properties:

- The e-scooter was travelling at a constant speed of 20 km/h with the pass-by commencing 50 m from the listener.

Table 2
Alert sound stimuli and associated component levels.

	Factor level				
Stimulus	C1 - Frequency	C2 - Wave type	C3 - AM depth	C4 - AM rate	C5 - AM synchrony
S1	0	0	0	1	1
S2	1	0	0	0	0
S3	0	1	0	0	1
S4	1	1	0	1	0
S5	0	0	1	1	0
S6	1	0	1	0	1
S7	0	1	1	0	0
S8	1	1	1	1	1
S9	Impulsive sound based around a 1 kHz tone with ADSR envelope settings of 5, 30, 0.01 and 5 ms respectively and a rate of 7 Hz				
S10	Baseline e-scooter audio recording				

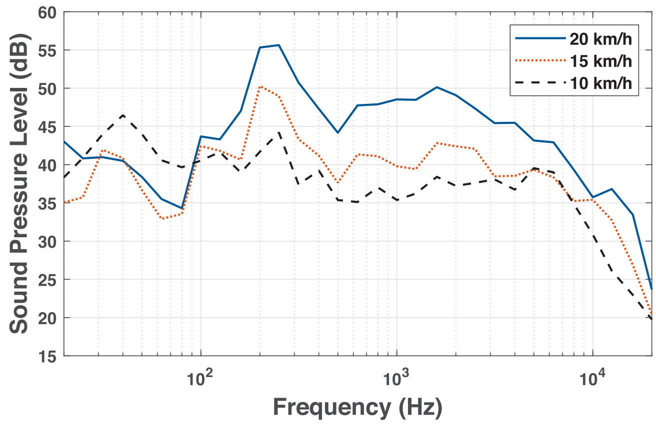


Fig. 1. Third octave frequency spectrum of e-scooter pass-by at 1 m distance L_{ZFmax}.

- The e-scooter passed 2 m in front of the listener, from either left to right, or right to left, with the direction randomly assigned.
- Distance attenuation was applied based on free-field radiation of a monopole, resulting in a sound level that is inversely proportional to the distance to the listener.
- The Doppler effect was modelled, as was high-frequency air absorption.
- The signal was rendered to the loudspeaker array using vector-base amplitude panning (VBAP) [30].

2.2.4. Calibration

The first stage of the calibration process involved calibrating stimulus S10, and the baseline rolling noise component of stimuli S1 to S9, to 52 dBA (L_{AFmax}) ($\pm$0.5 dB) when at 2 m from the listener, as determined by the pass-by sound level measurements. The alert sound components of stimuli S1 to S9 were then added to produce a total level of 56 dBA (L_{AFmax}) ($\pm$0.5 dB) when at 2 m from the listener. This level corresponds to the minimum requirements specified in UN Regulation 138, which specifies a minimum pass-by sound level requirement of 56 dBA when measured at 2 m distance, for quiet running vehicles [14].

2.3. Environmental Noise

The role of environmental noise on vehicle detectability has previously been reported [31,32] and therefore two environmental noise scenarios were used within this study to increase ecological validity of the detection and annoyance rating tasks. The first soundscape consisted of a city park scenario and was calibrated to a sound level of 50 dBA (L_{Aeq}). The second soundscape consisted of a road traffic noise scenario and was calibrated to a sound level of 60 dBA (L_{Aeq}). These two scenarios were chosen as they offer varied but representative use cases for e-scooter passes. Further details can be found in Table 3 and fast Fourier transforms of the environmental noise clips can be found in Fig. 3.

The soundscapes were recorded using a Soundfield ST450 ambisonic microphone and a Zoom F8n Field Recorder. Short segments of 10–15 s were looped so as to create excerpts of approximately constant level with minimal attention attracting events. Such soundscapes have previously been labelled as “amorphous sequences” [33] and are referred to as background noises in which no specific event can be isolated. As a measure of variation in level, the difference between the 90th percentile and 10th percentile of the A-weighted sound pressure level was calculated, with values of 1.6 dBA and 1.7 dBA for clips N1 and N2 respectively. The loop segments were chosen to limit perception of periodicity when repeated and were cross-faded to reduce perception of connecting points. The B-format ambisonic recordings were rendered to the loudspeaker array using the SPAT 5 renderer for MAX/MSP [29].

2.4. Environmental Process

The experiment consisted of four sections: familiarisation, training, detection and rating. The familiarisation page consisted of all 10 stimuli randomly assigned to buttons ’A’ through ’J’ on the interface, with no environmental noise present. Pass-by simulation effects were not applied to the stimuli at this stage, and therefore the perception can be considered to be from the perspective of the e-scooter rider. Participants were instructed to famil-

Table 3
Environmental noise characteristics.

ID	Description	Level (L_{Aeq})
N1	City park soundscape, characterised by by birdsong and distant road traffic noise.	50 dBA
N2	Soundscape of passing traffic, as recorded next to a medium to busy city road.	60 dBA

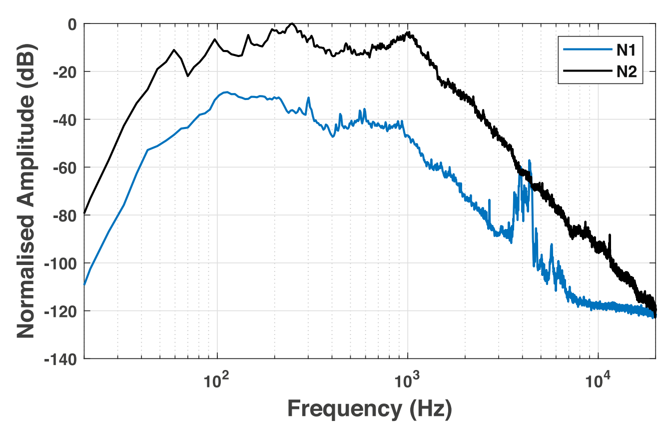


Fig. 3. Frequency characteristics of environmental noise excerpts. The increase in energy at 4 kHz for N1 has been identified as birdsong.

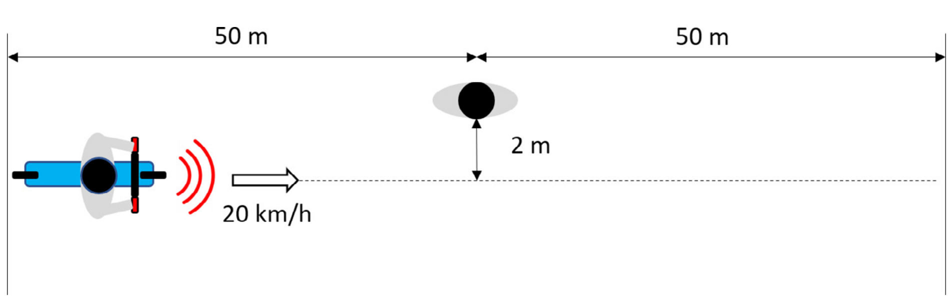


Fig. 2. E-scooter pass-by scenario as acoustically simulated for detection tasks.

iarise themselves with the various alert sounds in preparation for the detection tasks.

For the training and main detection sections, each page corresponded to a single detection task. After pressing start, there was a random time delay of between 0 and 3 s before the pass-by event started from either the left or the right of the listener, as determined randomly. As soon as the participant detected the alert sound, they were instructed to press the left arrow key on the keyboard if the e-scooter was approaching from the left, or the right arrow key on the keyboard if the e-scooter was approaching from the right. Participants were free to move on to the next page within their own time.

The training section consisted of four detection pages, each with a predetermined stimulus (S2, S6, S7 and S9) and environmental noise (N1). The purpose of this section was for participants to learn how to use the detection task interface. The results from this section were not considered during the subsequent data analysis.

The main detection section consisted of 40 detection pages. The initial environmental noise condition was randomly chosen and then for this condition, detection tasks of all 10 stimuli were completed. The order of the stimuli was randomly chosen. This was then repeated for the second noise condition. The whole process with alternating environmental noise conditions and random stimuli was repeated to produce a total of 40 tasks (2 environmental noise x 10 stimuli x 2 repeats).

Following the detection page, participants made annoyance ratings of the stimuli. The rating page was presented as a multiple stimuli method, where all 10 stimuli were rated on the same page, with an annoyance scale ranging from 0 ('not at all annoying') to 100 ('extremely annoying'). Intermediate verbal anchors were provided at 25 ('slightly annoying'), 50 ('moderately annoying') and 75 ('very annoying'). An open text box was also provided to elicit comments about the given ratings. The annoyance ratings were made in the presence of environmental noise N1 so as to increase ecological validity of the ratings.

For all pages, spoken instructions were available via a text-to-speech generator and the experiment could be navigated using keyboard shortcuts or mouse input.

2.5. Participants

A total of 30 subjects participated in the study. Four of these were excluded due to their responses containing a high number of outliers (see Section 3.1), leaving a sample of 26 participants for the following analysis. Of these, 20 were male and 6 were female. Age data was recorded in ranges, with the youngest participants falling within the age range of 18–25 and the oldest within the range 56–65. All participants were fluent in English and self reported normal hearing. Participants were recruited from a listening experiment participant database and included subjects who were both internal and external to the university. Subjects received a small monetary compensation for their participation.

It should be noted that although none of the participants were blind or partially sighted, this is not expected to have negatively influenced the applicability of the results, as previous research has shown no significant difference between sighted and blind participant groups for AVAS reaction times [34,18].

3. Results and Discussion

3.1. Participant Screening

Prior to further analysis, participant reliability was checked by comparing directional error rates (i.e. recording opposite direction to actual pass) across participants, as well as comparing missed detections, defined as responses after the e-scooter had passed the listener's position (reaction distance RD < 0). The error rates from four participants were identified as outliers when compared to the sample as a whole (greater than 1.5 times the interquartile range above the upper quartile), based on a median of 8 and an upper adjacent of 13. These participants were subsequently excluded from further analysis as particularly high error rates could indicate that a participant had difficulty completing the task, or did not fully understand the instructions.

3.2. Analysis of Missed Detections

Missed detection rates including incorrect directional responses were compared across all stimuli and environmental noise conditions for the remaining participants, Fig. 4. It is seen that stimulus S10 (baseline condition) resulted in a missed detection rate of 29% for noise condition N1 and 77% for noise condition N2. The addition of alert sounds (S1 - S9) increased detectability in all cases with fewer missed detections, however, it is apparent that the missed detection rate varied markedly by stimulus. Stimuli S1 and S5 show an increased missed detection rate for both noise conditions in comparison to the other stimuli, with rates above 50% for noise condition N2. These stimuli share the factors of C1 = low frequency and C2 = sine wave, suggesting that at the presented levels, this combination of features results in low detection rates, especially for noise conditions comprising road traffic noise.

For the subsequent analysis, trials with no response recorded and trials with directional errors (11.6% of trials) have been excluded from the dataset.

3.3. Descriptive statistics of detection distance

Fig. 5 presents the mean detection distance by stimulus and noise. A 'risk' area is highlighted in red (RD $<$ 7.5 m); if a pedestrian steps in front of the oncoming e-scooter within this distance, they

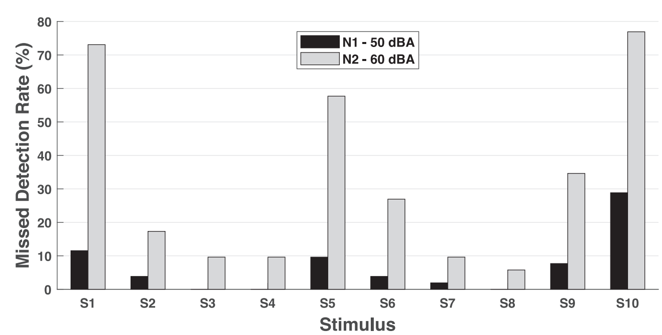


Fig. 4. Missed detection rate by noise and stimulus.

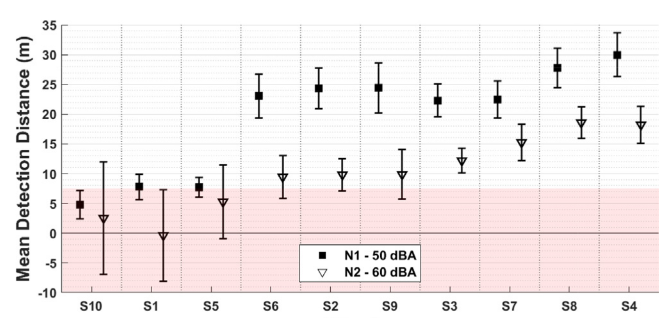


Fig. 5. Mean detection distance by noise and stimulus. Red shading indicates 'risk' area. Error bars show 95% confidence interval.

are at risk of being hit by the e-scooter given the reaction time needed for the rider to start breaking, as based on stopping distances established in [35] for cars. It should be noted that this risk area is based on reaction times and breaking distances for larger vehicles, and therefore the risk area may be conservative for e-scooters, which have a considerably lower mass.

The baseline condition, S10, has mean detection distances within the risk area for both environmental noise conditions, with a mean detection distance in the presence of a 50 dBA environmental noise of 4.8 m, falling to 2.5 m in 60 dBA noise, albeit with a large variance for the louder noise condition. This suggests that e-scooters travelling at 20 km/h without additional alert sounds do not provide sufficient auditory warning for pedestrians to react in a timely manner, when in a typical city soundscape.

For the stimuli that contain an additional alert sound (S1-S9), there is a large distribution for mean detection distances, highlighting the fact that the metric of L_{AFmax} alone is insufficient for predicting detectability. Stimuli S1 and S5 are seen to provide mean detection distances within the risk area for N2, whilst being close to the risk area for N1. Along with the data presented in Fig. 4, this suggests that the combination of C1 = low frequency and C2 = sine wave, results in low detectability for the sound levels presented. On the other hand, stimuli S8 and S4 provide the best detectability performance, which correspond to C1 = high frequency and C2 = saw wave. The detectability performance of the stimuli is therefore considered to be dependent upon the spectral characteristics of the stimuli and masking environmental noise, with lower frequency components being more effectively masked by the ambient soundscape, thus being less detectable. The degree to which the stimuli are masked can be characterised by the metric ‘partial loudness’, which is discussed in greater detail in Section 3.7.2.

3.4. Descriptive statistics of annoyance ratings

Fig. 6 presents the mean annoyance ratings by stimulus, where 0 corresponds to ’not at all annoying’ and 100 corresponds to ’extremely annoying’. Compared to the baseline (S10), all stimuli with additional alert sounds have an increased mean annoyance rating. Stimuli S5, S9, S1, S6 and S2 have annoyance ratings closest to the verbal anchor ’slightly annoying’, whereas stimuli S3, S7, S4 and S8 have mean ratings between ’moderately annoying’ and ’very annoying’. All of this later group correspond to the saw wave type, suggesting that this is a prominent factor in the subjective annoyance ratings.

3.5. Analysis of sound feature significance

In order to evaluate the significance of the sound features on detectability and annoyance in more depth, a linear mixed model

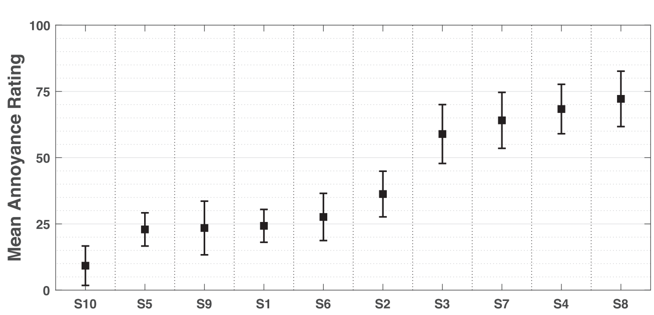


Fig. 6. Mean annoyance ratings by stimulus; 0 corresponds to ’not at all annoying’ and 100 corresponds to ’extremely annoying’. Error bars show 95% confidence interval.

analysis was conducted on responses including stimuli S1 to S8. As fixed effects in the model, the components C1 to C5 were used, with main effects calculated only, due to aliasing of interactions (see Section 2.2.1). To account for differences between individuals, variable Participant was used as a random effect in the model, including intercepts. Visual inspection of residual plots did not reveal any deviations from homoscedasticity or normality.

Type III tests of fixed effects revealed that the main effects of frequency, wave type and AM rate were statistically significant ($p < .05$) for detectability, and the main effects of frequency and wave type were statistically significant ($p < .05$) for annoyance, see Table 4.

Descriptive statistics of the significant effects are presented in Table 5. These results highlight that, of the components studied, the higher frequency level (fundamental of 400 Hz) and saw wave type level improve detectability whilst at the same time increasing annoyance. Moreover, the AM rate has a small significant influence on detectability, with the lower rate (4 Hz) performing slightly better than the higher rate (8 Hz).

3.6. Relationship between detectability and annoyance

The relationship between detectability and annoyance is presented in Fig. 7, using ratings made in the presence of environmental noise N1. As found in previous studies [19], there is a noticeable correlation between detectability and annoyance. By analysing the location of the stimuli within this perceptual space, it is possible to assess how well suited they are for use as alert sounds that minimise annoyance. A cluster of stimuli (S10, S5 and S1) can be seen at low detection distances and low annoyance ratings, although as these fall within the risk area (RD < 7.5 m), they are not considered to provide sufficient detectability performance. On the other hand, a cluster of stimuli (S3, S7, S4 and S8) can be seen at high detection distances and high annoyance ratings; these provide maximal detectability at the cost of increased annoyance. Finally, a cluster of stimuli (S9, S6 and S2) is seen with high detection distances and relatively low annoyance ratings; out of the stimuli tested, these offer the greatest balance between detectability and annoyance. Stimuli S6 and S2 are characterised by the higher frequency level and the sine wave type. S9, which offers the second lowest

Table 4
Type III fixed effects for dependent variables detection distance and annoyance ratings.

Variable	Source	df	F	p
Detectability	C1 - Frequency	1/ 727	106.14	<**.001**
	C2 - Wave type	1/ 727	140.78	<**.001**
	C3 - AM depth	1/ 727	0.09	0.769
	C4 - AM rate	1/ 727	9.83	**0.002**
	C5 - AM type	1/ 727	2.89	0.089
Annoyance	C1 - Frequency	1/177	12.87	<**.001**
	C2 - Wave type	1/177	255.21	<**.001**
	C3 - AM depth	1/177	0.01	0.917
	C4 - AM rate	1/177	0.01	0.929
	C5 - AM type	1/177	0.82	0.638

Table 5
Mean detection distance (m) and annoyance rating values (scale from 0–100) for significant sound features.

Variable	Source	Level 0	Level 1
Detectability (m)	C1 - Frequency	12.6	20.5
	C2 - Wave type	11.8	21.0
	C4 - AM rate	17.6	15.7
Annoyance	C1 - Frequency	42.6	51.1
	C2 - Wave type	27.8	65.9

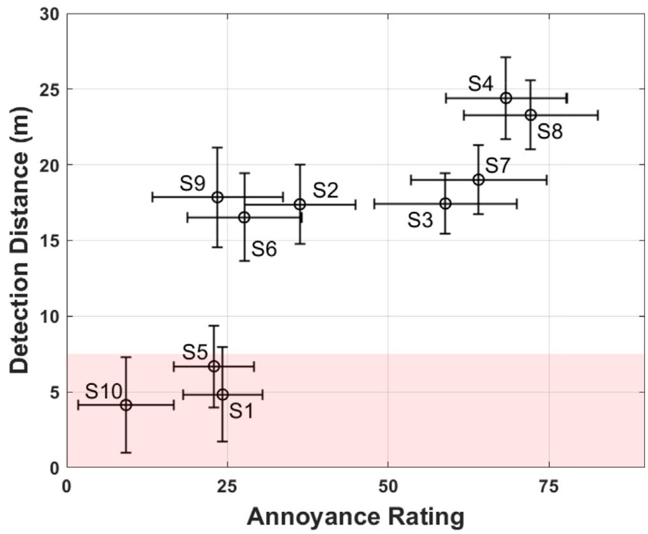


Fig. 7. Mean detection distances versus mean annoyance ratings. Error bars show 95% confidence interval. Red shading indicates 'risk' area.

annoyance ratings yet fifth best detectability of the additional alert sounds tested, is characterised by an impulsive sound based around a 1 kHz tone.

3.7. Relationship between objective metrics and subjective performance

3.7.1. Sound Quality Metrics

The psychoacoustic metrics of the e-scooter stimuli were calculated using HEAD Acoustics ArtemiS Suite 12.5 software. The calculation of loudness (units of sone) was made according to DIN 45631/A1 [36], which is based on Zwicker's loudness model and includes a modification for time varying signals. There are no standard methods for calculating roughness (asper) and fluctuation strength (vacil) and therefore these two metrics were calculated according to the hearing model given by Sottek [37], as was tonality (tuHMS). Sharpness (acum) was calculated according to the Aures method [38], which takes into account the influence of absolute loudness on the sharpness perception. The Aures model of sharpness was considered appropriate in this instance as the stimuli were presented in the presence of a masking noise, potentially increasing the importance of loudness on the perception of attributes at high frequency.

Zwicker's model of psychoacoustic annoyance [26] is calculated using the terms loudness (N), sharpness (S), fluctuation strength (F) and roughness (R), and is given by:

$$PA = N_5\left(1 + \sqrt{w_S^2 + w_{FR}^2}\right), \tag{1}$$

where N_5 is the 5th percentile of the loudness (sone) and

$$w_S = \begin{cases} (S - 1.75) \cdot 0.25 \log(N_5 + 10), & \text{if } S > 1.75 \\ 0, & \text{if } S \leqslant 1.75 \end{cases}, \tag{2}$$

$$w_{FR} = \frac{2.18}{N_5^{0.4}}(0.4F + 0.6R). \tag{3}$$

Fig. 8 plots the mean subjective annoyance ratings versus the metric psychoacoustic annoyance, as calculated by Zwicker's model. Simple linear regression was used to test if PA significantly predicted subjective annoyance responses. The results of the regression indicated that the model explained 96.1% of the variance and that the model was significant, $F(1, 194) = 4513.5, p < .001$. It was found that PA significantly predicted subjective annoyance ratings for the presented e-scooter sounds ($\beta 1 = 7.94, p < .001$).

Previous studies have shown that by using broadband sound pressure levels as a predictor of AVAS annoyance, coefficients of determination from regression models have ranged from $R^2 = 0.34$ when using the average sound pressure level (L_{Aeq}) as a predictor, to $R^2 = 0.65$ when using the maximum sound pressure level (L_{Amax}) [24]. Furthermore, in the same study, Altinsoy developed a metric labelled as the "annoyance index", which used loudness, tonality, roughness and fluctuation strength, with linear contributions from each. This resulted in a coefficient of determination of $R^2 = 0.91$. Although Zwicker's model of PA used in the study presented here did not include a tonality component, the non-linear contribution from the SQMs (with the exception of loudness) are believed to have resulted in a stronger model as this aligns more closely with auditory perception [26]. In the case where the AVAS stimuli set contains a mixture of strongly tonal and non-tonal sounds, it may be necessary to include a tonality term to Zwicker's PA model, as outlined in [39] and further assessed in the context of drone noise annoyance in [27].

3.7.2. Partial Loudness

The term 'partial loudness' (PL) refers to the perceived loudness of a target sound against a background of other masking sounds. As the level of the masking sound increases, the partial loudness of the target sound is therefore reduced [40]. Partial loudness has previously been used as a metric for detectability and annoyance of AVAS sounds, and was shown to provide a better prediction metric than signal-to-noise-ratio (SNR) [20], as well as giving good performance for the audibility prediction of a wider range of technical signals in real-world background noises [41].

To investigate the use of partial loudness as an objective metric for e-scooter alert sound performance, a partial loudness model implemented in Python / C++ was used [42], as based on Glasberg and Moore's model of partial loudness for time-varying sounds [43]. The model was calculated for each stimuli and noise combination, with the environmental noise summed to mono and the original stimuli loops without pass-by processing. The output of the model was 'Short Term Partial Loudness' (STPL).

Figs. 9a and 9c plot partial loudness versus mean detection distance for environmental noise conditions N1 and N2, with simple linear regression models included. When including all stimuli, the regression model has coefficients of determination of $R^2 = 0.597$ and $R^2 = 0.852$ for noise conditions N1 and N2 respectively. It is notable that stimuli S9 is an outlier to the regression model, and its detection performance is not well predicted by PL, especially for the lower noise condition. This is consistent with the literature, which shows that current loudness models, includ-

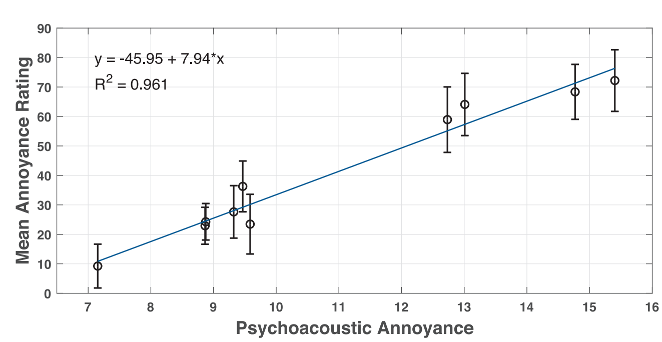
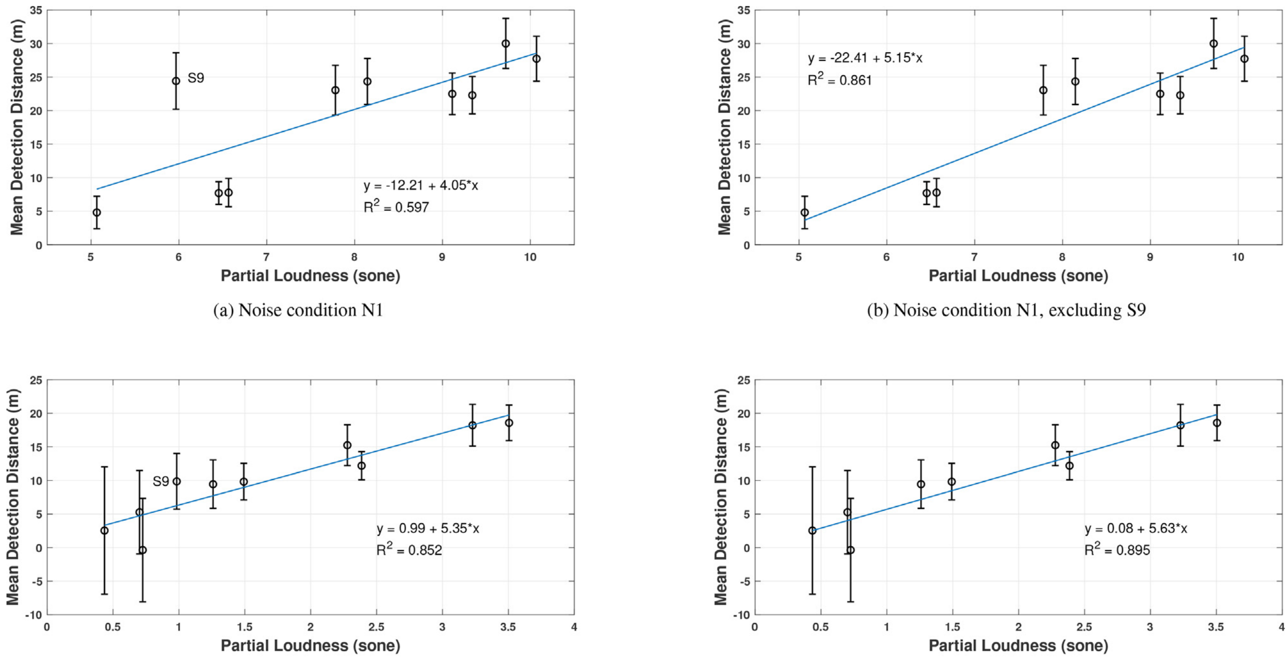


Fig. 8. Mean subjective annoyance ratings versus the metric psychoacoustic annoyance, with simple linear regression fitted.



(a) Noise condition N1

(b) Noise condition N1, excluding S9

(c) Noise condition N2

(d) Noise condition N2, excluding S9

Fig. 9. Partial loudness versus mean detection distance for environmental noise conditions N1 and N2, with and without impulsive stimulus S9.

ing Glasberg and Moore (2002), often show discrepancies for strongly time-varying signals [44,45]. As such, linear regression was recalculated excluding stimuli S9 to assess PL as a predictor for detectability for the non-impulsive sounds, Figs. 9b and 9d. When excluding S9, the regression model has coefficients of determination of $R^2 = 0.861$ and $R^2 = 0.895$ for noise conditions N1 and N2 respectively, and is significant ($p < .001$) in both cases. Partial loudness is therefore a good predictor for detectability of AVAS sounds in the presence of environmental noise, and outperforms the metrics of loudness and psychoacoustical annoyance, as summarised in Table 6.

The relationship between partial loudness and annoyance was also investigated, with the linear regression model resulting in a coefficient of determination of $R^2 = 0.925$ ($p < .001$). This is consistent with results presented by Jacobsen et al., who reported values between $R^2 = 0.64$ and $R^2 = 0.91$ [20]. Although not as strong as PA as a predictor of subjective annoyance, this result confirms the use of loudness-based metrics for the prediction of subjective annoyance.

4. Summary and Conclusions

In this paper, the development of an e-scooter AVAS has been considered by taking a perception-influenced design approach to designing alert sounds that optimise detectability and annoyance. A listening experiment has been conducted using ambisonic soundscapes and simulated auralisations of e-scooter passes at 20 km/h, in which a detection-based task and annoyance rating task were conducted by 30 participants.

Table 6
Objective metric comparison for the prediction of detectability.

Noise Condition	Predictors	R^2
N1	N	0.528
	PA	0.559
	PL	0.597
	PL (excl. S9)	0.861
N2	N	0.662
	PA	0.821
	PL	0.852
	PL (excl. S9)	0.895

Detectability results showed that, when no additional alert sound was included, missed detection rates were 29% for a 50 dBA environmental noise condition, and 77% for a 60 dBA environmental noise condition. When combined with mean detection distances of 4.8 m (50 dBA) and 2.5 m (60 dBA) for this baseline condition, it can be considered that e-scooters travelling at 20 km/h without additional alert sounds do not provide sufficient auditory warning for pedestrians to react in a timely manner, when in a typical city soundscape. These detection rates and mean detection distances are likely to be even further reduced in many city environments when the ambient noise level is greater than the 60 dBA level used in this study.

With the addition of a 56 dBA L_{AFmax} alert sound, e-scooter passes at 20 km/h showed improved detectability rates and mean detection distances for all alert sounds investigated. In the case of the best performing AVAS sound, mean detectability distance increased to 30.0 m in the 50 dBA noise condition and 18.2 m in the 60 dBA noise condition, an improvement of 4.9 s in terms of reaction time for the worst case scenario. Specific acoustic features of the alert sound had a significant influence on detectability performance, with high frequency, saw wave type sounds providing the best detectability performance out of those studied. Statistically significant AVAS components for detection were frequency, wave type and AM rate. This result highlights that broadband sound level metrics such as L_{AFmax} are inadequate for predicting detectability performance, as all stimuli were calibrated to the equivalent L_{AFmax} level.

By plotting detection distance versus annoyance, stimuli could be evaluated within a perceptual space so as to optimise detectability and annoyance. Whilst detectability and annoyance were seen to be highly correlated, a cluster of stimuli offered good detectability with relatively low annoyance ratings. Two of these sounds were characterised by higher frequency sine tones (400

+ 600 + 800 Hz) with amplitude modulation, with the most optimal being characterised by an impulsive 1 kHz tone at a rate of 7 Hz. This suggests that modulated or impulsive tones with frequency content in the 800 Hz - 1 kHz range, and with sine type characteristics, may provide optimal micromobility AVAS sounds, however further research is needed. Comparing these sounds to the EV AVAS regulations as set out in UNECE 138 [14], all of the continuous stimuli tested (S1 - S8) would fulfil requirements, subject to calibration, as they span multiple third-octave frequency bands, however, the impulsive 1 kHz tone (S9) would not. Adding further frequency components to this sound should be investigated, as this may mean it is more detectable for individuals with frequency-specific hearing loss, and satisfy current AVAS regulations.

Regression analysis showed that the objective metric of Zwicker's psychoacoustic annoyance is a useful predictor of subjective annoyance for AVAS sounds, with a coefficient of determination of $R^2 = 0.96$. When compared to other metrics for the prediction of subjective annoyance, such as annoyance index in [24], PA is seen to explain more variance and should therefore be considered when developing micromobility AVAS sounds. Likewise, partial loudness was studied as a predictor of detectability, with strong positive association seen ($R^2 \approx 0.9$). By evaluating signals with these objective metrics, it could be possible to assess the suitability of AVAS sounds, in terms of optimisation between detectability and annoyance, without the need for resource intensive listening experiments. However, current loudness models often show discrepancies for strongly time-varying signals [44,45], and their applicability to impulsive type AVAS sounds needs to be further investigated.

Due to practicalities in experiment duration, a single AVAS sound level was considered in this study (56 dB L_{AFmax} at 2 m); this corresponded to current AVAS regulations for EVs [14]. Further work is needed to optimise AVAS sound pressure levels for micromobility transportation to further balance detectability and annoyance, and the objective metrics of PA and PL will likely be useful tools for this. For instance, by interpreting the slope of regression between PL and detection distance, it is possible to calculate the required PL for a given detection distance in a specific noise environment. Such analysis should be considered alongside the distribution of responses, as basing alert sound levels on mean results may produce inadequate detectability for a set percentage of the population.

Further work should also consider directivity of AVAS loudspeakers for micromobility. This topic has been considered previously in the context of EVs [46,47], however unlike traditional EVs where the driver is acoustically insulated from the reproduced AVAS sound, riders of e-scooters will be more exposed to generated alert sounds, in part due to proximity to the loudspeaker. Beamforming and array technology should be considered with the purpose of increasing detectability for pedestrians whilst simultaneously reducing annoyance for the rider.

Funding

The authors would like to acknowledge the funding provided by the University of Salford's Higher Education Innovation Fund and emTransit B.V (Dott).

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Tim Walton: Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Antonio J. Torija:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Andrew S. Elliott:** Conceptualization, Methodology, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Maria Luiza de Ulhôa Carvalho for kindly providing the ambisonic soundscape recordings.

References

[1] Zag Daily. European update: 360,000 e-scooters available across the continent [online] (2021). Accessed: 2022–08-02.
[2] Zag Daily. Zag data: Over half a million shared e-scooters on Europe's streets [online] (2022). Accessed: 2022–08-02.
[3] Department for Transport. Future of mobility: Urban strategy [online] (2019). Accessed: 2022–08-02.
[4] Department for Transport. Perceptions of current and future e-scooter use in the UK - Summary report [online] (2021). Accessed: 2022–08-02.
[5] Department for Transport. Reported road casualties Great Britain: e-scooter factsheet 2021 [online] (2022). Accessed: 2022–08-02.
[6] Department for Transport. Legalising rental e-scooter trials [online] (2020). Accessed: 2022–09-21.
[7] Goodes P, Cerrato G, Bai YB, Investigation into the detection of a quiet vehicle by the blind community, in: Proceedings of InterNoise 2008, 2008, pp. 133–140.
[8] Garay-Vega L, Hastings A, Pollard JK, Zuschlag M, Stearns MMD, Quieter cars and the safety of blind pedestrians: Phase 1. DOT HS 811 304, Tech. rep., United States. National Highway Traffic Safety Administration (2010). URL: https://rosap.ntl.bts.gov/view/dot/9474
[9] Hastings A, Pollard, JK Garay-Vega L, Stearns MD, Guthy C, Quieter cars and the safety of blind pedestrians, phase 2: Development of potential specifications for vehicle countermeasure sounds. DOT HS 811 496, Tech. rep., United States. National Highway Traffic Safety Administration (2011). URL:https://rosap.ntl.bts.gov/view/dot/9475
[10] Sakamoto I, Tanaka T, Sekine M, Morita K, Reviews on countermeasure to safety risk associated with quietness of hybrid or electric vehicles in Japan, in: Proceedings of InterNoise 2010, 2010.
[11] Kim DS, Emerson RW, Naghshineh K, Pliskow J, Myers K. Impact of adding artificially generated alert sound to hybrid electric vehicles on their detectability by pedestrians who are blind. J Rehabil Res Dev 2012;49:381. https://doi.org/10.1682/JRRD.2011.03.0041.
[12] A.J. Torija, A.S. Elliott, L. Harris, Z. Podwinska, C. Welham, Generation and analysis of artificial warning sounds for electric scooters, in: Proceedings of InterNoise 2022, 2022.
[13] Karaaslan E, Noori M, Lee JY, Wang L, Tatari O, Abdel-Aty M. Modeling the effect of electric vehicle adoption on pedestrian traffic safety: An agent-based approach. Transp Res C: Emerg Technol 2018;93:198–210. https://doi.org/10.1016/J.TRC.2018.05.026.
[14] UNECE, Regulation No 138 of the Economic Commission for Europe of the United Nations (UNECE) – Uniform provisions concerning the approval of Quiet Road Transport Vehicles with regard to their reduced audibility [2017/71] (2017).
[15] MLIT, Guideline for the approaching vehicle audible system (Japanese), Ministry of Land, Infrastructure, Transport and Tourism (2010).
[16] NHTSA, Minimum sound requirements for hybrid and electric vehicles, National Highway Traffic Safety Administration, Department of Transportation, US (2016).
[17] Sekine M, Sakamoto I, Houzu H, Nishi T, Morita K, Necessity of acoustic vehicle alerting system for electric motorcycles to ensure pedestrian safety, in: Proceedings of InterNoise 2017, 2017.
[18] Parizet E, Ellermeier W, Robart R. Auditory warnings for electric vehicles: Detectability in normal-vision and visually-impaired listeners. Appl Acoust 2014;86:50–8. https://doi.org/10.1016/J.APACOUST.2014.05.006.
[19] Parizet E, Robart R, Pondrom P, Chamard JC, Baudet G, Quinn D, Janssens K, Haider M. Additional efficient warning sounds for electric and hybrid vehicles. Energy Environ 2016:501–10. https://doi.org/10.1002/9781119307761.CH32.

[20] Jacobsen GN, Ih, Song W, Macdonald EN, Predicting detectability and annoyance of EV warning sounds using partial loudness, in: Proceedings of InterNoise 2016, 2016.
[21] Lee SK, Lee M, Shin T, Han M. Objective evaluation of the sound quality of the warning sound of electric vehicles with a consideration of the masking effect: Annoyance and detectability. Int J Automot Technol 2017;18:699–705. https://doi.org/10.1007/s12239-017-0069-6.
[22] Steinbach L, Atamer S, Altinsoy E, Evaluation of electric vehicle sounds, new concepts regarding speed-dependency and prediction of the warning effect, in: Proceedings of Meetings on Acoustics, Vol. 30, Acoustical Society of America ASA, 2017, p. 040006. doi:10.1121/2.0000567.
[23] Petiot JF, Legeay K, Lagrange M, Optimization of the sound of electric vehicles according to unpleasantness and detectability, in: International Conference on Engineering Design, ICED19, 2019, pp. 5–8. doi:10.1017/dsi.2019.402.
[24] Altinsoy ME. The evaluation of conventional, electric and hybrid electric passenger car pass-by noise annoyance using psychoacoustical properties. Appl Sci 2022;12:5146. https://doi.org/10.3390/APP12105146.
[25] Misdariis N, Cera A, Sound signature of quiet vehicles: state of the art and experience feedbacks, in: Proceedings of InterNoise 2013, 2013.
[26] Fastl H, Zwicker E, Psychoacoustics: Facts and models, 2nd Edition, Springer-Verlag, 1999. doi:10.1007/978-3-540-68888-4/COVER.
[27] Torija AJ, Chaitanya P, Li Z. Psychoacoustic analysis of contra-rotating propeller noise for unmanned aerial vehicles. J Acoust Soc Am 2021;149:835. https://doi.org/10.1121/10.0003432.
[28] Montgomery DC. Design and analysis of experiments. (8th edition),. John-Wiley & Sons; 2012.
[29] Thibaut C, A new implementation of Spat in Max, in: 15th Sound and Music Computing Conference (SMC2018), 2018, pp. 184–191.
[30] Pulkki V. Virtual sound source positioning using vector base amplitude panning. J Acoust Soc Am 1997;45:456–66.
[31] Yamauchi K, Menzel D, Takada M, Nagahata K, Iwamiya SI, Fastl H. Psychoacoustic examination of feasible level of additional warning sound for quiet vehicles. Acoust Sci Technol 2015;36:120–5. https://doi.org/10.1250/AST.36.120.
[32] Poveda-Martínez P, Peral-Orts R, Campillo-Davo N, Nescolarde-Selva J, Lloret-Climent M, Ramis-Soriano J. Study of the effectiveness of electric vehicle warning sounds depending on the urban environment. Appl Acoust 2017;116:317–28. https://doi.org/10.1016/J.APACOUST.2016.10.003.
[33] Dubois D, Guastavino C, Raimbault M. A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories. Acta Acust united Ac 2006;92:865–74.
[34] Altinsoy E, The detectability of conventional, hybrid and electric vehicle sounds by sighted, visually impaired and blind pedestrians, in: Proceedings of InterNoise 2013, 2013.
[35] Glaeser KP, Marx T, Schmidt E, Testing the sound detection of electric vehicles by blind or visually impaired persons, in: Proceedings of Forum Acusticum, 2014.
[36] Deutsches Institut Für Normung, DIN 45631/A1-2010, Calculation of Loudness Level and Loudness From the Sound Spectrum-Zwicker Method–Amendment 1: Calculation of the Loudness of Time-Variant Sound, 2010.
[37] Sottek R, Models for signal processing in the human auditory system, Ph.D. thesis, RWTH Aachen University, Aachen, Germany (1993).
[38] Aures W. Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale. Acustica 1985;59.
[39] Di GQ, Chen XW, Song K, Zhou B, Pei CM. Improvement of Zwicker's psychoacoustic annoyance model aiming at tonal noises. Appl Acoust 2016;105. https://doi.org/10.1016/j.apacoust.2015.12.006.
[40] Moore BC, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness, and partial loudness, AES. J Audio Eng Soc 1997;45.
[41] Schell-Majoor L, Rennies J, Ewert SD, Kollmeier B. Application of psychophysical models for audibility prediction of technical signals in real-world background noise. Appl Acoust 2015;88:44–51. https://doi.org/10.1016/J.APACOUST.2014.08.001.
[42] Ward D, Applications of loudness models in audio engineering, Ph.D. thesis, Birmingham City University (2017).
[43] Glasberg BR, Moore BCJ. Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds, AES. J Audio Eng Soc 2005;53:906–18.
[44] Rennies J, Verhey JL, Appell JE, Kollmeier B, Loudness of complex time-varying sounds? a challenge for current loudness models, in: Proceedings of Meetings on Acoustics, Vol. 19, 2013. doi:10.1121/1.4799514.
[45] Sottek R, Moll T, Loudness perception and modeling of impulsive sounds, in: Proceedings of Euronoise, 2015.
[46] Quinn D, Mitchell J, Clark P, Garcia JJ, eVADER - development and initial evaluation of a next generation pedestrian alert solution for quiet electric vehicles, in: Proc. FISITA, 2014.
[47] Kournoutos N. Design of electric vehicle warning sound systems to minimise drive-by noise. University of Southampton; 2020. Ph.D. thesis.