Synthetic Sounds for Automated Vehicles: Loud is Effective

Pavlo Bazilinskyy[a,*], Roberto Merino-Martınez[b], Elif Özcan[c], Dimitra Dodou[d], Joost de Winter[d]

[a]Future Everyday, Department of Industrial Design, Eindhoven University of Technology, Groeneloper 3, 5612 AZ Eindhoven, The Netherlands
[b]Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands
[c]Faculty of Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands
[d]Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract
Synthetic vehicle sounds have been introduced in electric vehicles, and as external human-machine interfaces for automated vehicles. While previous research has studied the effect of synthetic vehicle sounds on detectability and acceptance, the present study takes on a different approach by examining the efficacy of synthetic vehicle sounds in preventing people from crossing the road. An online study was conducted in which 125 participants were presented with different types of synthetic sounds, including sounds of a combustion engine, pure tones, combined tones, and beeps. A vehicle moving in a straight trajectory at a constant velocity of 30 km/h was used, and no visual information was provided. Participants in the role of pedestrians were asked to hold down a key when they felt safe to cross. After each trial, they assessed whether the vehicle sound was easy to notice, whether it gave enough information to realize that a vehicle was approaching, and whether the sound was annoying. The results showed that louder sounds were the most effective in preventing participants from crossing the road, whereas beeps were relatively ineffective in doing so. These results may prove insightful for the improvement of synthetic vehicle sounds.

Keywords: electric vehicles, synthetic sound signals, psychoacoustic annoyance, road safety, listening experiments

* Corresponding author email address: p.bazilinskyy@tue.nl (Pavlo Bazilinskyy)

1. Introduction

More than 270,000 fatal pedestrian traffic accidents occur annually worldwide [1], the majority of which occur during road crossing [2]. Causes of pedestrian-vehicle accidents include underestimation of the crossing gap or the time needed to cross [3], low visibility [4], and visual obstruction of the approaching vehicle [5]. Augmenting the sounds emitted by vehicles could potentially help prevent unsafe crossing. Such solutions have been the topic of investigation for electric vehicles (EVs) as well as in the form of external human-machine interfaces (eHMIs) for automated vehicles (AVs).

1.1. Sound design for electric vehicles

The market penetration of EVs is increasing, and several countries have announced plans to cease production of new internal combustion engine vehicles (ICEV) between 2025 and 2030 [6-9]. However, the lack of combustion noise can make EVs so quiet that they may become unsafe for vulnerable road users (VRUs) with visual impairments [10, 11]. Therefore, additional sound signals are used to make EVs noticeable by VRUs. In 2018, the US National Highway Traffic Safety Administration (NHTSA) declared that all EVs traveling at speeds lower than 30 km/h must emit a sound of an A-weighted sound pressure level of 60 dBA to 64 dBA [12]. In the European Union, EVs are required to have an Acoustic Vehicle Alerting System (AVAS) that makes the vehicle emit a sound of at least 50 dBA at speeds lower than 10 km/h and 56 dBA at speeds lower than 20 km/h [13, 14]. Both regulations share a common feature in that while they describe sound level thresholds and timing, they do not specify which types of sound should be used.

Previous studies have explored various types of artificial sounds for EVs, including engine sounds, hums, whistles, beeps, white noise, as well as music/melodious sounds [15-17]. The effect of spectral characteristics has also been examined, including volume, pitch, frequency, and modulation of frequency and amplitude [11, 18-21], as well as the effect of background noise [11, 18, 22-26]. Outcome measures included detectability (self-reported [10, 20] or in the form of response time [11, 22, 25, 27, 28]), detection rate [18], or detection distance [19]. Other outcome measures studied include acceptance [17], pleasantness [26-28], appropriateness [26], alertness [21], powerfulness [28], and annoyance [21, 29]. The results showed that higher frequencies, particularly when background noise is present [18, 20], frequency modulation [19], and irregular amplitude modulation [11] resulted in better detection. Furthermore, higher frequencies [21], saw-tooth signals [21, 22], and louder signals [29] tend to be perceived as more annoying. Naturally, loudness also affects detectability [30].

Only a few studies have investigated the relationship between detectability and annoyance in EVs, and which sound characteristics may result in high detectability, yet low annoyance. In an experiment with 30 participants, Lee et al. [22] reported that sounds with amplitude and frequency modulation led to faster detection and lower perceived annoyance than saw-tooth signals. Petiot et al. [27] used an interactive genetic algorithm to develop sounds: 15 assessors rated sounds in terms of their detectability and pleasantness. The sounds were weighted combinations of four components (a thermic motor sound, a harmonic sound, and two filtered broadband noises). Two different filters were also applied to the final sound. In total, more than 70 parameters were adjustable, including the frequencies and amplitudes of the four components. The genetically evolved sound was found to result in statistically significantly higher ‘fitness’ (i.e., a combination of higher detectability and lower pleasantness) than sounds developed by human designers instructed to develop sounds satisfying these two criteria. It was also found that including the sound of a motor is important for detectability.

The effect of sounds on participants’ willingness to cross has not been extensively investigated. An exception is the on-road study by Emerson et al. [31], in which visually impaired participants were standing next to a road with approaching traffic and asked to indicate their willingness to cross by pressing and releasing a button. The authors reported that vehicles switching to internal combustion later (i.e., after having reached a higher speed) were considerably less detectable than vehicles switching to internal combustion soon after the start accelerating from a stop, indicating the importance of engine sound as a cue of approaching traffic.

1.2. Sound design for automated vehicles

Next to EVs, auditory signals could be useful in the form of auditory external human-machine interfaces (eHMIs) that communicate the state or intention of automated vehicles (AVs) to VRUs. Most eHMIs developed so far are visual, but a number of auditory eHMIs have also been proposed (see [32] for an overview). Mahadevan et al. [33] used the verbal messages “I see you” and “cross” together with visual and tactile (via a mobile phone) feedback and found that a combination of modalities improved pedestrians’ awareness of the approaching AV compared to single modalities. Deb et al. [34] tested a horn, music, and the verbal message “safe to cross” and found that verbal messages were preferred over the abstract sounds tested. Music was tested by Florentine et al. [35], who noted that it helped draw pedestrians’ attention. Inspired by research on EV sounds, Moore et al. [36] proposed the use of a synthetic engine sound to indicate the intention of a driverless AV to stop in front of a pedestrian. The authors tested the concept with a hybrid vehicle in a Wizard-of-Oz naturalistic setting. The results showed that participants in the role of pedestrians rated the clarity of the AV’s intention higher in the presence of engine sound compared to a condition without sound.

1.3. Study aim

The aim of this study is to examine the effectiveness of various types of synthetic sounds that could be used as auditory eHMI to inform pedestrians of an approaching vehicle. We evaluated the effect of pure tones, combined tones, intermittent tones, beeps, and engine/tire sounds, presented with and without background noise, on perception by assessing factors such as whether the sound is easy to notice, whether it gives enough information to realize a vehicle is approaching, and whether it is annoying. Moreover, instead of asking participants to press a key as soon as they detected the sound, we asked them to hold a key when they would be willing to cross the road, in order to examine the efficacy of synthetic vehicle sounds on pedestrian behavior. We expect detectability to increase with loudness and tonal frequency, whereas annoyance is likely to increase with loudness [37-39]. Participants were recruited via crowdsourcing, resulting in a larger sample and higher statistical power compared to previous studies.

2. Method

2.1 Sound stimuli

Thirty sound stimuli were synthetically generated. These sounds can be classified into four categories: (1) continuous pure tones at a single frequency, (2) intermittent pure tones (a 500-ms interval emitting followed by a 500-ms interval not emitting), (3) combined tones, and (4) double beeps. In addition, a stimulus with a diesel engine sound signal [40] was included as a baseline representing an ICEV. A stimulus with only tire noise [41] was also included to assess the performance of a quiet EV/AV.

The tonal sounds in categories (1), (2), and (3) were presented at four frequencies: 350 Hz, 500 Hz, 1000 Hz, and 2000 Hz. The combined tones of category (3) were the same as the continuous pure tones of category (1) but with two additional tones at frequencies ±90 Hz from the main tone. The combined tones were expected to be perceived as more annoying because of the addition of the extra tone at +90 Hz [42]. The double-beep signal of category (4) consisted of eight pairs of 240-ms beeps with a 100-ms interval between beeps in a pair and a 1000-ms interval between pairs. This stimulus was tested by Bazilinskyy et al. [43] (in a series of double-beep stimuli with 2000, 1000, 750, and 430 ms intervals). These authors found that beeps with shorter intervals were perceived as more urgent. For the current study, the stimulus with a medium (1000 ms) interval was selected.

The set of 15 signals (3 tonal sounds x 4 frequencies + double beeps + ICEV + tires; see Table 1) was offered with background noise (a recording of a quiet street [44]) and without, resulting in a total of 30 stimuli. In all cases with background noise, the tire noise sound was also added (for the tire noise sound stimulus, this means that it was offered once with and once without background noise). In the cases without background noise, no tire noise was added to test the pure effect of the synthetic sound. The stimuli with background noise were presented at a lower volume than their counterparts without background noise (Figure 1) to assess a more realistic signal-to-noise ratio in practical applications. Figure 2 shows the spectrogram of the continuous pure tone at 2000 Hz, with and without background noise. From the spectrogram, the sound of the vehicle can be clearly distinguished. The change in frequency at around 7 s corresponds to the Doppler effect (see Section 2.2). The regular peaks between 3000 and 6000 Hz in the bottom spectrogram correspond to bird tweets in the background noise.

All stimuli were generated at a sampling rate of 44.1 kHz. The duration of each stimulus was 13.6 s.

Table 1
Sound stimuli included in the experiment

	Name	Characteristics
1	Cont350	Pure tone, continuous, 350 Hz
2	Cont500	Pure tone, continuous, 500 Hz
3	Cont1000	Pure tone, continuous, 1000 Hz
4	Cont2000	Pure tone, continuous, 2000 Hz
5	Dscr350	Pure tone, intermittent (13 x [500 ms emitting, 500 ms not emitting]), 350 Hz
6	Dscr500	Pure tone, intermittent (13 x [500 ms emitting, 500 ms not emitting]), 500 Hz
7	Dscr1000	Pure tone, intermittent (13 x [500 ms emitting, 500 ms not emitting]), 1000 Hz
8	Dscr2000	Pure tone, intermittent (13 x [500 ms emitting, 500 ms not emitting]), 2000 Hz
9	Comb350	Combined tone, continuous, 350 Hz (±90 Hz)
10	Comb500	Combined tone, continuous, 500 Hz (±90 Hz)
11	Comb1000	Combined tone, continuous, 1000 Hz (±90 Hz)
12	Comb2000	Combined tone, continuous, 2000 Hz (±90 Hz)
13	Beeps	Double beeps (8 x [240 ms beep, 100 ms pause, 240 ms beep, 1000 ms pause]), 1840 Hz
14	Diesel	Diesel engine
15	Tire	Tires on asphalt

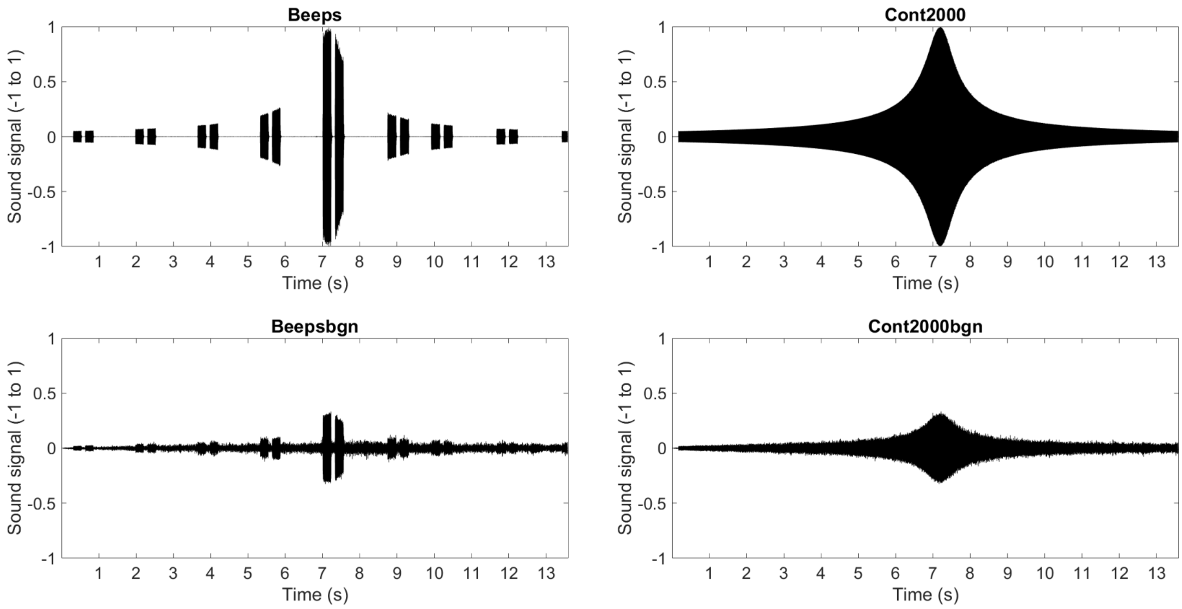

Figure 1. Sound signal as a function of the elapsed time for two example stimuli without background noise (top) and the same stimuli with background noise (bgn) (bottom).

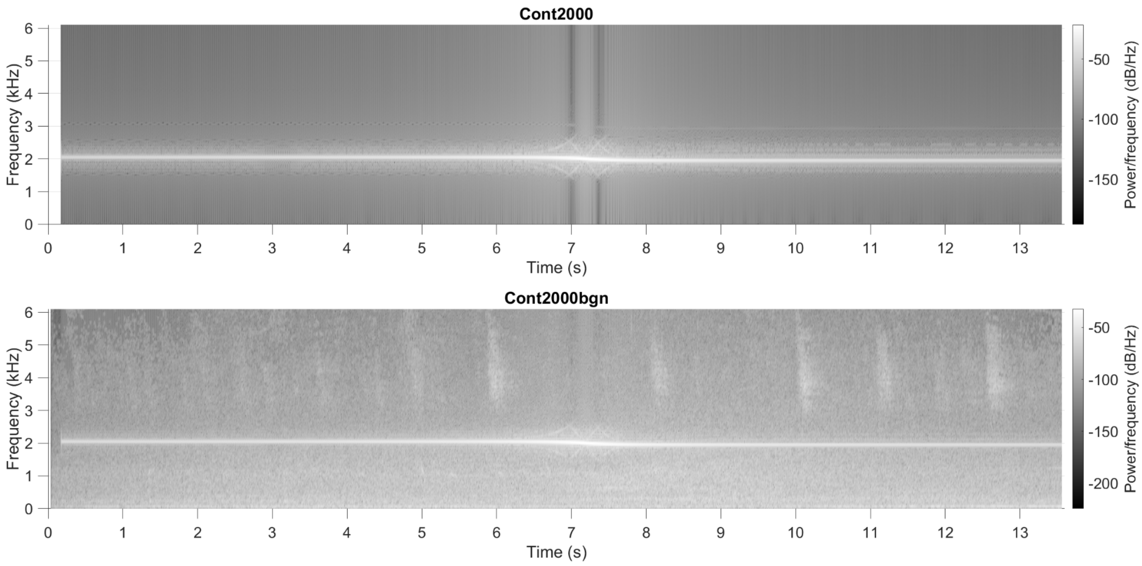

Figure 2. Spectrogram of stimuli Cont2000 (top) and Cont2000bgn (bottom).

2.2. Sound emission from a moving source

This study assumes a relatively simple sound source and observer geometry. A two-dimensional arrangement is employed, where the observer and the sound source are on the same plane, and the reflection of sound from buildings or other structures is neglected. Furthermore, the sound source is assumed to be a monopole moving at a constant velocity $\boldsymbol{V}$ of 30 km/h (≈ 8.33 m/s) in the positive x direction along a straight line at a distance y_s = 3 m from the observer (Figure 3). The initial position of the sound source at t = 0 is defined at ($x_{s,0}$, y_s) and at a general instant t as ($x_{s,0}$ + $V.t$, y_s). For the generation of the sound stimuli, the source moved from $x_{s,0}$ = −60 m to a final position of x_s = 53 m.

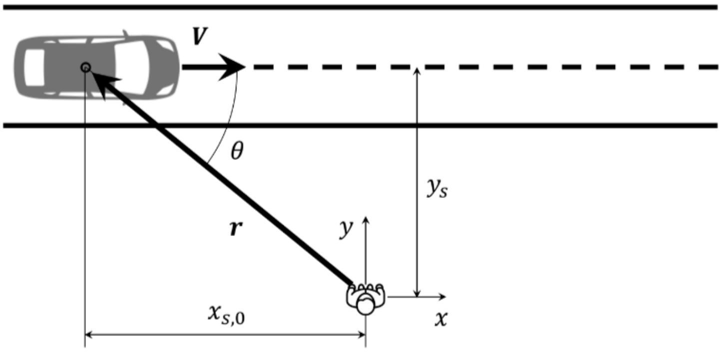

Figure 3. Schematic representation of the vehicle approaching a stationary pedestrian.

For a stationary observer, the amplitude of a sound emitted by a moving source changes over time due to the relative motion of the source. The amplitude of the sound pressures observed by the pedestrian, $A(t_{obs})$, relative to the amplitude of the sound pressures observed by the pedestrian if the vehicle would be standing still (M = 0) in front of the pedestrian at coordinate (0, y_s) is given by Equation (1) [45, 46].

$$\frac{A(t_{obs})}{A_0} \approx \frac{y_s}{r(1-M.cos(\theta))^2} \quad (1)$$

, in which M is the Mach number, defined as $V/c \approx \frac{8.33\ m/s}{343\ m/s} \approx 0.024$.

The sound signal is observed with a slight delay with respect to the emission time:

$$t_{obs} = t + r/c \quad (2)$$

Equation 1 describes that the observed sound amplitude increases as the vehicle gets closer to the pedestrian and decreases once the vehicle has passed the pedestrian. Furthermore, the effect of the source motion causes a frequency shift due to the Doppler effect, as described by Equation 2 [45].

$$f_{observed} = \frac{f_{emitted}}{1-M.cos(\theta)} \quad (3)$$

2.3. Participants

Participants subscribed to the study through the crowdsourcing service Appen (https://appen.com). They could become aware of this research by logging into one of the channel websites (e.g., https://www.ysense.com) where our study was presented on a list of other projects available for completion. We allowed contributors from all countries to participate. Participants were not allowed to complete the study more than once using the same worker ID. A payment of 0.50 USD was offered after the completion of the experiment.

In addition to the crowdsourced participants, a small number of participants recruited among students conducted the same study. These participants answered the same pre-experimental questions as the crowdsourced participants, but in a Google form questionnaire instead of in Appen. The experiment itself was presented in the same online environment as that of the crowdsourced participants.

The research was approved by the Human Research Ethics Committee of the Delft University of Technology (reference number 1233).

2.4. Procedure

The study was presented in English. At the top of the page, contact information was provided. Participants were informed that they could contact the investigators to ask questions about the study and that they had to be at least 18 years of age. Information about anonymity and the voluntary nature of participation was also provided.

The experiment was created using a modified version of the framework based on jsPsych [47] that was used in a previous study on the measurement of reaction times to auditory, visual, and multi-modal stimuli [48], as well as in studies investigating the willingness of pedestrians to cross in front of an automated vehicle, using the same key-press method as here [49, 50]. The sounds were pre-loaded before the start of the experiment to prevent delays during the experiment.

Participants first provided demographic information about their age, gender, and driving experience. Next, they were asked to leave the questionnaire by clicking on a link that opened a webpage with the experiment and were presented with the following instructions:

"*Imagine that you are a pedestrian standing on the side of the road. You will listen to 60 sounds of vehicles driving by you. When the sound is playing, press and HOLD 'F' when you feel safe to cross the road in front of the car. You can release the button and then press it again multiple times during the sound. After each sound, you will be asked to answer a few questions. After each 10 sounds you will be able to take a short break. Sometimes you will be asked to listen to a phrase and type what was said.*

Please make sure that your audio is on. On the next page, you will listen to a song. When you will be listening to the song, adjust your volume level to be able to hear the song clearly. Do NOT change your volume level till the end of the experiment. Press 'C' to proceed." The song used to adjust the volume was taken from [51].

The participants had to respond to 60 sounds presented in blocks. Participants were randomly assigned to either listen to 30 sound stimuli with background noise first, followed by 30 without background noise, or vice versa. Each sound was presented twice per block, presented in a random order that differed for each participant. Before each stimulus sound, the participants were instructed as follows: "*Start by HOLDING the 'F' key. Release the key when it becomes unsafe to cross; press again when safe to cross*". The instruction remained visible throughout the duration of the stimulus. After each stimulus, they were asked to rank the sound based on three criteria: (1) "easy to notice", (2) "gave me enough information to realize that a vehicle was approaching", and (3) "annoying". Each criterion was ranked by moving a slider between 0 and 10. The participant could not proceed to the next stimulus before having moved all the sliders (Figure 4). Participants did not receive feedback on their responses.

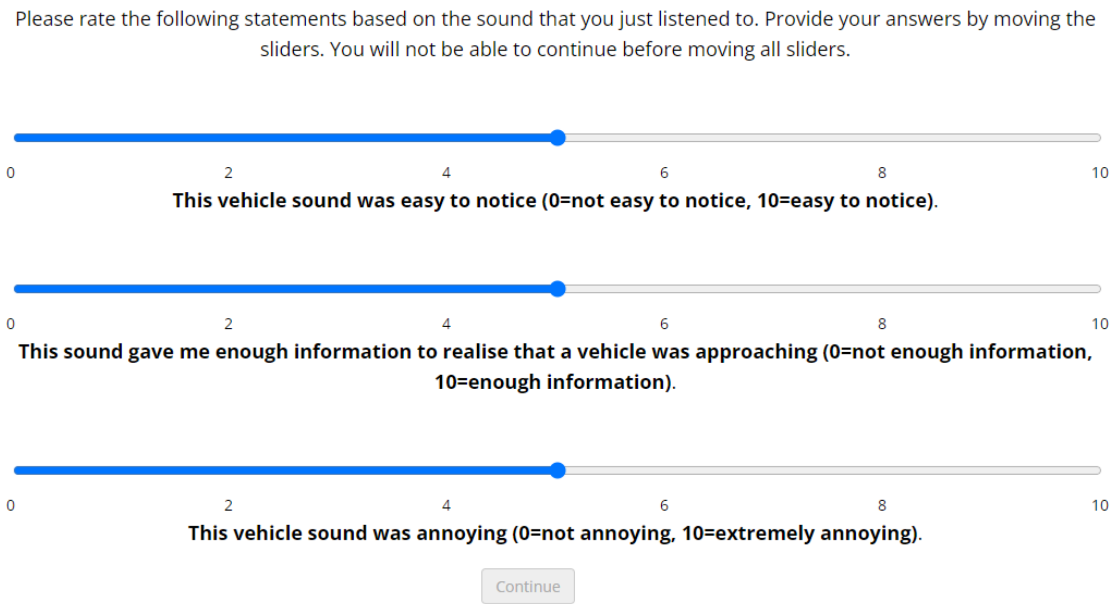

Figure 4. Set of questions that the participants received after listening to each sound.

In order to ensure attentive participation, six control phrases were injected, one in each batch: (1) "Oranges are orange", (2) "Lemons are yellow", (3) "Cherries are red", (4) "Apples are green", (5) "Blackberries are black", and (6) "Grapes are blue". They were generated using the British English Amy female voice available at [52]. The participants had to type the test phrase they listened to. The volume of the test phrases was normalized to the volume level of the stimuli. After the experiment, the response typed by the participant was automatically compared to the known response to determine whether participants were able to hear the sound and understand the task.

After every ten trials, participants were presented with text indicating how many of the 60 sounds they had completed, for example: "*You have now completed 10 sounds out of 60. When ready press 'C' to proceed to the next batch.*"

At the end of the experiment, the participants were given a unique code. They had to enter the code on the questionnaire as proof that they completed the experiment in order to receive their remuneration.

2.5. Data analysis

For each participant and sample, a performance score was computed as 100% minus the percentage of key presses in the 4–8 s interval. The performance score represents how many (or how early) participants released the key until the moment the virtual vehicle passed.

Moreover, the loudness for the 4–8 s interval was computed from the sound signal using ISO 532-1 (Zwicker loudness). Zwicker loudness is a method of measuring noise that takes into account characteristics of human hearing, such as, the dependence of sound transmission through the middle ear on frequency [53]. Swart and Bekker [54] compared Zwicker loudness with other psychoacoustic metrics and showed that it was suitable for distinguishing between similar vehicle sounds [54].

Pearson correlation coefficients were computed between performance scores, loudness levels, and the scores of the three questions presented after each stimulus ("easy to notice", "gave me enough information to realize that a vehicle was approaching", and "annoying"), averaged across all participants.

3. Results

Nine hundred ninety-five people participated between 16 and 18 September 2020. The study received a satisfaction rating of 4 on a scale from 1 ("very dissatisfied") to 5 ("very satisfied"). Participants who entered the same finishing code multiple times (n = 29), who indicated that they had not read the instructions (n = 20), who indicated that they were under 18 years old (n = 5), who completed the task in an impossible amount of time of less than 900 s (n = 180), who, due to a data exchange problem, had no data stored in the database (n = 399) or had incomplete data in the database (n = 419), who made 4 or more mistakes with test phrases (n = 390), who provided more than 5 answers with a difference of at least 9 points between the two repeated scenarios (n = 144), were removed. Additionally, if people completed the study more than once from the same IP address, only the first response was kept. In total, 885 participants were removed due to a technical issue with the data storage and the strict filtering described in Section 2.5, leaving 110 participants from 33 countries for the analysis.

The 110 participants had a mean age of 35.8 years (SD = 11.4 years). Of the 110 participants, 74 were male, 35 were female, and 1 preferred not to respond. The countries that were most represented were Venezuela (n = 31), India (n = 9), and Turkey (n = 7). The participants took, on average, 44.3 minutes to complete the study (SD = 16.0 min, median = 42.0 min).

From the recruiting via acquaintances, 21 people participated between 4 December 2020 and 12 January 2021. Of these, 6 participants were excluded because they provided five or more answers with a difference of at least 9 points, leaving 15 participants. The remainder of the analysis will be conducted for the 110 crowdsourced and 15 additionally recruited participants together (N = 125).

Figure 5 shows the corresponding key-press percentages for two example signals from 30 auditory samples. A substantial difference in key presses can be distinguished, with the continuous tone being a stronger deterrent to cross (i.e., participants released the key earlier) compared to the beeping sound.

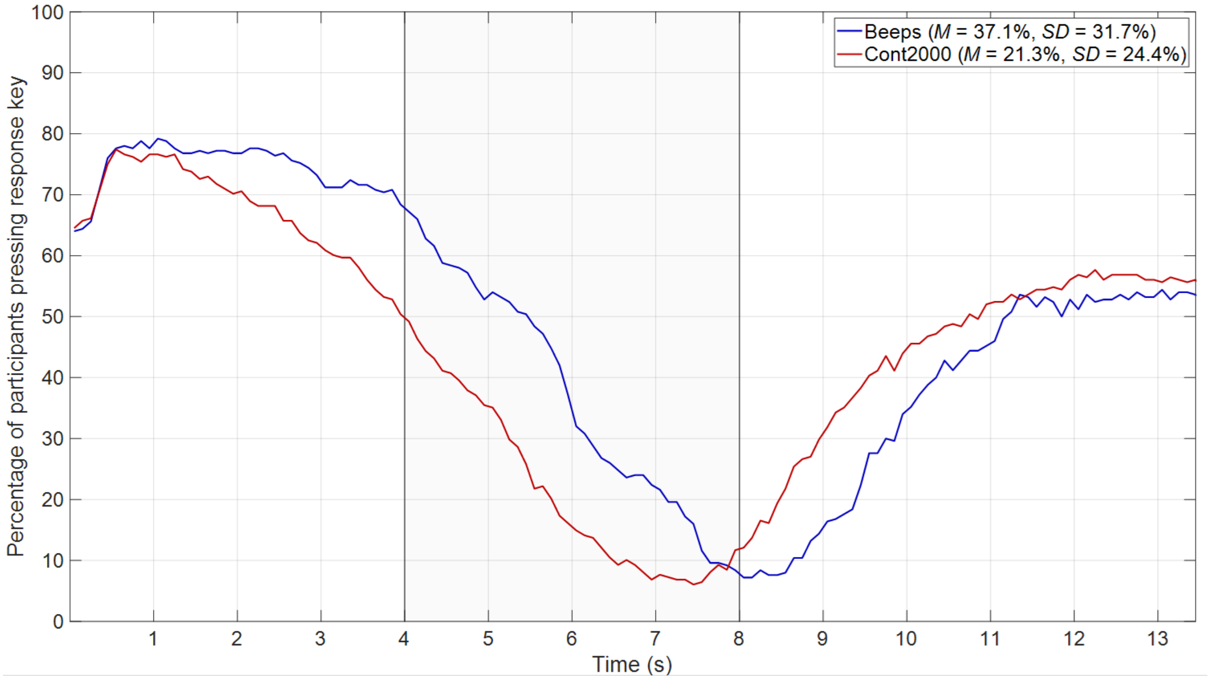

Figure 5. Key press percentage as a function of elapsed time for two selected samples. The legend shows the mean and *SD* of the performance score. The gray background indicates the time interval across which the performance score was computed.

Table 2 displays the correlation matrix among the performance score, the loudness for the 4–8 s interval, and the self-reported measures, with all scores being averaged over the 125 participants. Table 2 and Figure 6 show that sounds that yielded better performance scores were generally also louder (*r* = 0.81) and perceived as more annoying (*r* = 0.58). We also computed the correlation coefficients separately for the background noise stimuli (*n* = 15) and the no-background noise stimuli (*n* = 15). Table 2 reports the average of these two correlation coefficients in parenthesis. The correlation between performance and loudness remained present in the subgroups (*r* = 0.73)

From Figures 7 and 8, it can also be seen that continuous high-frequency tones yielded relatively high performance in the presence of background noise (74% for Cont2000bgn, and 70% for Cont1000bgn) even though they were not particularly easy to notice (4.71 and 4.64, respectively). Low-frequency tones, on the other hand, yielded relatively low performance (56% for Comb350bgn, 58% for Dscr350bgn, and 58% for Cont350bgn). It can also be seen that beeps and intermittent sounds, especially those of low tonal frequency, yielded relatively low performance scores, even in the absence of background noise (63% for Beeps, 62% for Dscr350).

Table 2

Means, standard deviations, and correlation coefficients of the variables for the auditory stimuli (n = 30).

	Mean	*SD*	1	2	3	4
1 Performance score (%)	67.3	5.74				
2 Easy to notice (0–10)	5.03	0.28	0.52 (0.12)			
3 Gave enough information (0–10)	4.72	0.21	0.41 (0.07)	0.92 (0.89)		
4 Annoying (0–10)	4.21	0.42	0.58 (0.18)	0.91 (0.66)	0.78 (0.55)	
5 Loudness score (sones)	33.1	7.29	0.81 (0.73)	0.76 (0.31)	0.66 (0.28)	0.82 (0.38)

Note. Numbers in parentheses represent the average correlation computed for background noise (*n* = 15) and no background noise (*n* = 15).

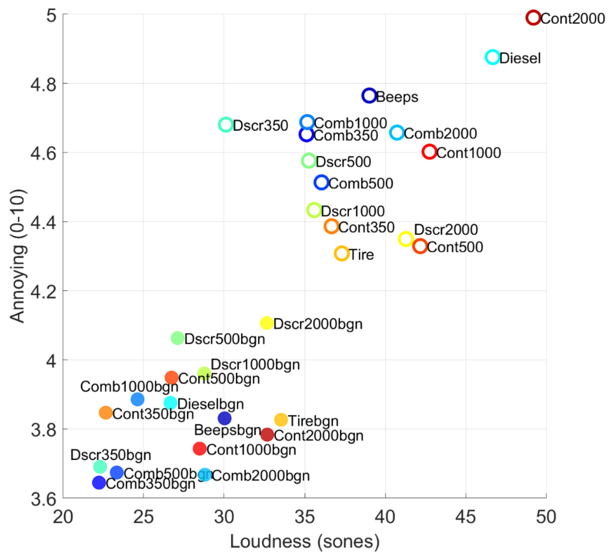

Figure 6. Scatter plot of perceived annoyance and computed loudness score of the 30 stimuli.

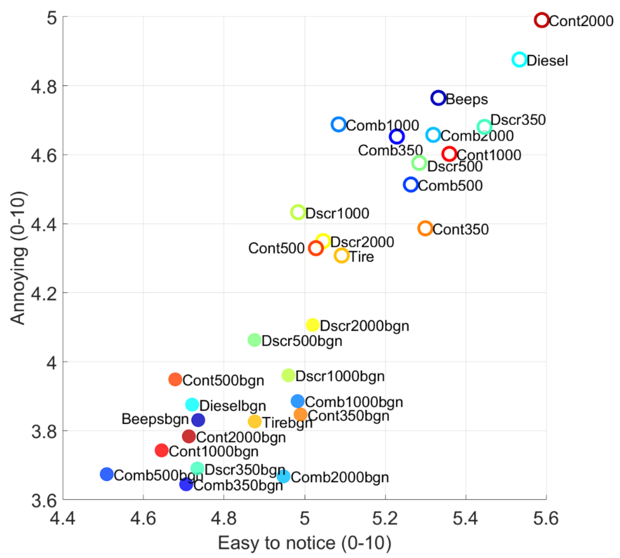

Figure 7. Scatter plot of perceived annoyance and perceived ‘easy to notice’ of the 30 stimuli.

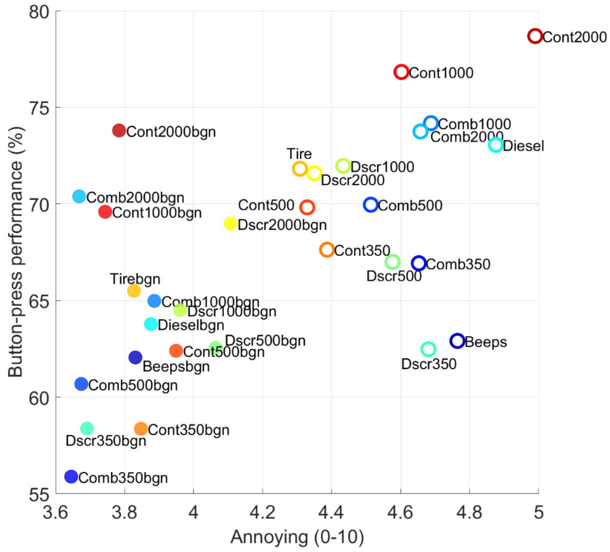

Figure 8. Scatter plot of performance (based on key-press inputs) and perceived annoyance of the 30 stimuli.

4. Discussion

This online study tested how vehicle sounds, naturalistic ones, as well as tones and beeps, influence participants' crossing intentions. In order to isolate the effect of sound, participants were not presented with any visual information, which may be representative of situations in which visual information is unavailable due to visual impairment or visual obstruction.

The results showed that loud sounds were the most effective in preventing participants from crossing the road, as demonstrated by a strong correlation between the performance score and the loudness of the stimulus (r = 0.81). Our findings also indicated that across all 30 stimuli, the higher-performing sounds were also the more annoying sounds. This supports Bisping's "trade-off hypothesis of pleasantness and power" [55] for in-vehicle sounds and according to which an increase in perceived pleasantness of a sound beyond a certain level has negative consequences on the perceived powerfulness of the car and vice versa (and see [56] for a similar negative correlation between valence and dominance/arousal of auditory stimuli). These results raise the question of whether current electric vehicles, which tend to emit naturalistic sounds or pleasant tones [28], are optimally effective in ensuring the safety of pedestrians.

Beeps and intermittent sounds, particularly those of low tonal frequency, yielded relatively poor performance, even when there was no background noise. One possible explanation is that the inter-beep intervals we used were relatively long(1000 ms). Previous studies that have investigated shorter intervals found that perceived urgency increases as the inter-pulse interval decreases [42, 57], following Stevens' power law [42]. In practice, slow beeps are typically used for relatively slowly evolving situations, such as a truck reversing, whereas fast beeps indicate an approaching hazard [58]. Moreover, the duration of the beeps themselves was short, which inhibited speed estimation. During the non-emitting intervals, especially if they are long, no new information is conveyed, leading to a lag in information processing.

A limitation of this study is that only a small range of relatively simple signals were tested. Moreover, only non-verbal stimuli were tested. While verbal messages have rich semantic content, non-verbal ones can convey information faster and more effectively in the presence of noise [59]. In addition, more complex sounds, such as music, were not included. Such sounds could draw attention [35] while being less loud and less annoying than tones or beeps. However, music might be less accepted than engine sounds [15] or verbal messages [34]. A second limitation is that, although we asked participants to adjust their volume so that a reference sound (a song) was audible, we did not have control over the quality of the sound equipment the participants used. Lastly, the ecological validity of our experiment is limited because the sounds were tested in a computer environment and in a predictable scenario of an approaching vehicle. In real-world sitations, road users divide their attention over a large number audiovisual stimuli, some of them even distracting from the crossing task, which would potentially impact the detectability of auditory signals emitted from approaching cars.

It is important to note that our findings do not necessarily imply that electric vehicles or eHMIs for automated vehicles should be made louder (and, by extension, more annoying). Doing so is one of the possibilities, but not the only way of informing vulnerable road users. Another way to increase the signal-to-noise ratio (as demonstrated in this study) would be to reduce the level of noise. For example, it may be expected that with the increasing electrification of vehicle fleets and the corresponding decline of combustion engine vehicles, artificial engine sounds may be more readily heard. Other options would be to continue research into sounds that are easily detectable yet not annoying, a topic that has been explored before in the literature [27].

To conclude, this study found a strong relationship between the loudness of a sound and its effectiveness in preventing participants from crossing the road. The present findings provide a useful reminder that loudness may be the primary (yet not the only) factor to consider in synthetic sound design for EVs and AVs. The study also suggests that intermittent beeps may need to be avoided as they may impede the pedestrian's ability to perceive the speed and distance of the approaching vehicle.

5. Supplementary material

The questionnaire used in Appen, sound stimuli, song for volume adjustment, test phrases, anonymized data, and MATLAB code are accessible at https://www.dropbox.com/sh/gbho9jyea9z0wgd/AADNFSJCC4N-HewqEU8lXIFGa?dl=0. The code for the experiment setup and the analysis are available at https://github.com/bazilinskyy/sound-ev-crowdsourcing.

References

[1] World Health Organization (2013) Pedestrian safety. A road safety manual for decision-makers and practitioners. http://apps.who.int/iris/bitstream/handle/10665/79753/9789241505352_eng.pdf
[2] DaSilva MP, Smith JD, Najm WG (2003) Analysis of pedestrian crashes. Report No. DOT-VNTSC-NHTSA-02-02. John A. Volpe National Transportation Systems Center, Cambridge, MA.
[3] Habibovic A, Davidsson J (2012) Causation mechanisms in car-to-vulnerable road user crashes: Implications for active safety systems. Accid Anal Prev 49:493-500. https://doi.org/10.1016/j.aap.2012.03.022
[4] Governors Highway Safety Association (2022) Pedestrian traffic fatalities by state. 2021 Preliminary data (January - December). https://www.ghsa.org/resources/Pedestrians22
[5] Bálint A, Labenski V, Köbe M, Vogl C, Stoll J, Schories L, Amann L, Baroda Sudhakaran G, Huertas Leyva P, Pallacci T, Östling M, Schmidt D., Schindler R (2021) Use case definitions and initial safety-critical scenarios. Report No. D2.6. Project SAFE-UP. https://www.safe-up.eu/resources
[6] Dow J (2021) Norway bans gas car sales in 2025, but trends point toward 100% EV sales as early as April. https://electrek.co/2021/09/23/norway-bans-gas-cars-in-2025-but-trends-point-toward-100-ev-sales-as-early-as-april
[7] Morgan S (2018) Denmark to ban petrol and diesel car sales by 2030. ttps://www.euractiv.com/section/electric-cars/news/denmark-to-ban-petrol-and-diesel-car-sales-by-2030
[8] Lambert F (2017) The Dutch government confirms plan to ban new petrol and diesel cars by 2030. https://electrek.co/2017/10/10/netherlands-dutch-ban-petrol-diesel-cars-2030-electric-cars
[9] GOV.UK (2020) Government takes historic step towards net-zero with end of sale of new petrol and diesel cars by 2030. https://www.gov.uk/government/news/government-takes-historic-step-towards-net-zero-with-end-of-sale-of-new-petrol-and-diesel-cars-by-2030
[10] Liu S, Fitzharris M, Oxley J, Edwards C (2018) The impact of electric / hybrid vehicles and bicycles on pedestrians who are blind or have low vision. Accident Research Centre. Monash University.
[11] Parizet E, Ellermeier W, Robart R (2014) Auditory warnings for electric vehicles: Detectability in normal-vision and visually-impaired listeners. Appl Acoust 86:50-58. https://doi.org/10.1016/j.apacoust.2014.05.006
[12] Federal Register (2018) Federal Motor Vehicle Safety Standard No. 141, Minimum sound requirements for hybrid and electric vehicles, section E. https://www.federalregister.gov/documents/2018/02/26/2018-03721/federal-motor-vehicle-safety-standard-no-141-minimum-sound-requirements-for-hybrid-and-electric

[13] European Commission (2017) Commission Delegated Regulation (EU) 2017/1576 of 26 June 2017 amending Regulation (EU) No 540/2014 of the European Parliament and of the Council as regards the Acoustic Vehicle Alerting System requirements for vehicle EU-type approval (Text with EEA relevance, §III.3.a(i). https://eur-lex.europa.eu/eli/reg_del/2017/1576/oj

[14] United Nations (2017) Agreement concerning the adoption of harmonized technical United Nations regulations for wheeled vehicles, equipment and parts which can be fitted and/or be used on wheeled vehicles and the conditions for reciprocal recognition of approvals granted on the basis of these United Nations regulations (Revision 3, including the amendments which entered into force on 14 September 2017), Addendum 137: UN Regulation No. 138, §6.2.8. https://unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2017/R138r1e.pdf

[15] Wogalter MS, Lim RW, Nyeste PG (2014) On the hazard of quiet vehicles to pedestrians and drivers. Appl Ergon 45:1306-1312. https://doi.org/10.1016/j.apergo.2013.08.002

[16] Yamauchi K, Menzel D, Takada M, Nagahata K, Iwamiya SI, Fastl H (2015) Psychoacoustic examination of feasible level of additional warning sound for quiet vehicles. Acoust Sci Technol 36:120-125. https://doi.org/10.1250/ast.36.120

[17] Singh S, Payne SR, Jennings PA (2014) Toward a methodology for assessing electric vehicle exterior sounds. IEEE Trans Intell Transp Syst 15:1790-1800. https://doi.org/10.1109/tits.2014.2327062

[18] Hsieh MC, Chen HJ, Tong ML, Yan CW (2021) Effect of environmental noise, distance and warning sound on pedestrians' auditory detectability of electric vehicles. Int J Environ Res Public Health 18:9290. https://doi.org/10.3390/ijerph18179290

[19] Fleury S, Jamet É, Roussarie V, Bosc L, Chamard JC (2016) Effect of additional warning sounds on pedestrians' detection of electric vehicles: An ecological approach. Accid Anal Prev 97:176-185. https://doi.org/10.1016/j.aap.2016.09.002

[20] Misdariis N, Cera A, Levallois E, Locqueteau C (2012) Do electric cars have to make noise? An emblematic opportunity for designing sounds and soundscapes. Paper presented at Acoustics 2012, Nantes, France, 23-27 April 2012

[21] Frlić Sekulić N, Blagojević I, Popović V, Janković S (2019) Development of pedestrian alert system for use in electric vehicles. Tehnički Vjesnik 26:1614-1619. https://doi.org/10.17559/TV-20180730144329

[22] Lee SK, Lee SM, Shin T, Han M. (2017) 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 18:699-705. https://doi.org/10.1007/s12239-017-0069-6

[23] Yamauchi K, Menzel D, Fastl H, Takada M, Nagahata K, Iwamiya SI (2011) Cross-cultural study on feasible sound levels of possible warning signals for quiet vehicles. INTER-NOISE NOISE-CON Congr Conf Proc 4:3236-3241

[24] Kournoutos N, Cheer J (2019) An environmentally adaptive warning sound system for electric vehicles. INTER-NOISE NOISE-CON Congr Conf Proc 259:683-691

[25] Poveda-Martinez P, Peral-Orts R, Campillo-Davo N, Nescolarde-Selva J, Lloret-Climent M, Ramis-Soriano J (2017) Study of the effectiveness of electric vehicle warning sounds depending on the urban environment. Appl Acoust 116:317-328. https://doi.org/10.1016/j.apacoust.2016.10.003

[26] Petiot JF, Kristensen BG, Maier AM (2013) How should an electric vehicle sound? User and expert perception. Int Des Eng Tech Conf Comput Inform Eng Conf 55928:V005T06A028. https://doi.org/10.1115/DETC2013-12535

[27] Petiot JF, Legeay K, Lagrange M (2019) Optimization of the sound of electric vehicles according to unpleasantness and detectability. Proc Des Soc: Int Conf Eng Des 1:3949-3958. https://doi.org/10.1017/dsi.2019.402

[28] Singh S, Payne SR, Jennings PA (2014) Toward a methodology for assessing electric vehicle exterior sounds. IEEE Trans Intell Transport Syst 15:1790-1800. https://doi.org/10.1109/TITS.2014.2327062

[29] Jacobsen G, Song W, MaDonald E (2016) Predicting detectability and annoyance of EV warning sounds using partial loudness. INTER-NOISE NOISE-CON Congr Conf Proc 253:886-895

[30] Yamauchi K, Sano T, Hasegawa S, Tamura F, Takeda Y (2014) Detectability and hearing impression of additional warning sounds for electric or hybrid vehicles. INTER-NOISE and NOISE-CON Congr Conf Proc 249:5279-5285

[31] Emerson RW, Naghshineh K, Hapeman J, Wiener W (2011) A pilot study of pedestrians with visual impairments detecting traffic gaps and surges containing hybrid vehicles. Transport Res F: Traffic Psychol Behav 14:117-127. https://doi.org/10.1016/j.trf.2010.11.007

[32] Dey D, Habibovic A, Löcken A, Wintersberger P, Pfleging B, Riener A, Martens M, Terken J (2020) Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design

principles of automated vehicles' external human-machine interfaces. Transport Res Interdiscip Perspect 7, 100174. https://doi.org/10.1016/j.trip.2020.100174
[33] Mahadevan K, Somanath S, Sharlin E (2018). Communicating awareness and intent in autonomous vehicle-pedestrian interaction. Paper presented at CHI Conference on Human Factors in Computing Systems, Montréal, Canada, 21-26 April 2018. https://doi.org/10.1145/3173574.3174003
[34] Deb S, Strawderman LJ, Carruth DW (2018) Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. Transport Res F: Traffic Psychol Behav 59:135-149. https://doi.org/10.1016/j.trf.2018.08.016
[35] Florentine E, Ang MA, Pendleton SD, Andersen H, Ang Jr MH (2016) Pedestrian notification methods in autonomous vehicles for multi-class mobility-on-demand service. Paper presented at the Fourth International Conference on Human Agent Interaction, Biopolis, Singapore, 4-7 October 2016, 387-392. https://doi.org/10.1145/2974804.2974833
[36] Moore D, Currano R, Sirkin D (2020) Sound decisions: How synthetic motor sounds improve autonomous vehicle-pedestrian interactions. Paper presented at the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Virtual Event, 21-22 September 2020, 94-103. https://doi.org/10.1145/3409120.3410667
[37] Fidell S, Teffeteller S (1981) Scaling the annoyance of intrusive sounds. J Sound Vib 78:291-298. https://doi.org/10.1016/s0022-460x(81)80039-9
[38] Fidell S, Teffeteller S, Horonjeff R, Green DM (1979) Predicting annoyance from detectability of low-level sounds. J Acoust Soc Am 66:1427-1434. https://doi.org/10.1121/1.383536
[39] Berglund B, Berglund U, Lindvall T (1976) Scaling loudness, noisiness, and annoyance of community noises. J Acoust Soc Am 60:1119-1125. https://doi.org/10.1121/1.381212
[40] Randall's Rest & Relaxation (2014) 12 hours of cummins diesel engine sound idling = dodge ram truck engine noise for sleeping. https://www.youtube.com/watch?v=2Y33bTIAA-E
[41] Zero Tailpipe (2018) What sound does Hyundai iONIQ Electric make? Noisy from outside at speed? https://www.youtube.com/watch?v=X0wpizkkH_Q
[42] Subedi JK, Yamaguchi H, Matsumoto Y, Ishihara M (2005) Annoyance of low frequency tones and objective evaluation methods. J Low Freq Noise Vib Active Control 24:81-95. https://doi.org/10.1260/0263092054531000
[43] Bazilinskyy P, Petermeijer SM, Petrovych V, Dodou D, De Winter JCF (2018) Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. Transport Res F: Traffic Psychol Behav 56:82-98. https://doi.org/10.1016/j.trf.2018.04.001
[44] Articulated Sounds (2017) Quiet streets ambience sound library. https://www.youtube.com/watch?v=6C-W_7BZBxQ
[45] Morse PM, Ingard KU (1968) Theoretical acoustics. McGraw-Hill, New York, p 74
[46] Ouyang K, Lu S, Zhang S, Zhang H, He Q, Kong F (2015) Online Doppler effect elimination based on unequal time interval sampling for wayside acoustic bearing fault detecting system. Sens 15:21075-21098. https://doi.org/10.3390/s150921075
[47] De Leeuw J (2014) jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behav Res Methods 47:1-12. https://doi.org/10.3758/s13428-014-0458-y
[48] Bazilinskyy P, De Winter JCF (2018) Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony. Hum Factors 60:1192-1206. https://doi.org/10.1177/0018720818787126
[49] Bazilinskyy P, Kooijman L, Dodou D, De Winter JCF (2021) How should external Human-Machine Interfaces behave? Examining the effects of colour, position, message, activation distance, vehicle yielding, and visual distraction among 1,434 participants. Appl Ergon 95:103450. https://doi.org/10.1016/j.apergo.2021.103450
[50] Sripada A, Bazilinskyy P, De Winter J (2021) Automated vehicles that communicate implicitly: examining the use of lateral position within the lane. Ergon 64:1416-1428. https://doi.org/10.1080/00140139.2021.1925353
[51] Royalty Free Music - No Copyright Music (2018) Creative minds - Bensound. https://www.youtube.com/watch?v=zFBwdFnezG0
[52] TTSMP3 (2022) Free text-to-speech and text-to-mp3 for US English. https://ttsmp3.com
[53] Zwicker E, Scharf B (1965) A model of loudness summation. Psychol Rev 72:3-26. https://doi.org/10.1037/h0021703
[54] Swart DJ, Bekkers A (2016) Interior and Motorbay sound quality evaluation of full electric and hybrid-electric vehicles based on psychoacoustics. INTER-NOISE NOISE-CON Congr Conf Proc 253:7029-7039.

[55] Bisping R (1995) Emotional effect of car interior sounds: Pleasantness and power and their relation to acoustic key features. SAE Technical Pap Ser 104: 2207-2213. https://doi.org/10.4271/951284
[56] Yang W, Makita K, Nakao T, Kanayama N, Machizawa MG, Sasaoka T, Sugata A, Kobayashi R, Hiramoto R, Yamawaki S, Iwanaga M, Miyatani M (2018) Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E). Behav Res Methods 50:1415-1429. https://doi.org/10.3758/s13428-018-1027-6
[57] Haas EC, Casali JG (1995) Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. Ergon 38:2313-2326. https://doi.org/10.1080/00140139508925270
[58] Watts GR, Godfrey NS, Savill TA (2001) Development of audible locating signals for use at pedestrian crossings. Appl Acoust 62:15-27. https://doi.org/10.1016/S0003-682X(00)00027-X
[59] Edworthy, J (1994) The design and implementation of non-verbal auditory warnings. Appl Ergon 25:202-210. https://doi.org/10.1016/0003-6870(94)90001-9

A APPENDIX

In the analysis described in the paper, strict filtering of participants was applied. We repeated the analysis with less strict criteria. More specifically, from the crowdsourced sample, only participants who indicated that they did not read the instructions (*n* = 20), who indicated that they were under 18 years old (*n* = 5), who completed the task in an implausibly short amount of time of less than 900 s (*n* = 180), or who, due to a data exchange problem, had no data stored in the database (*n* = 419) were removed. In total, 450 of 995 participants were removed, resulting in 550 participants from 50 countries remaining for the analysis.

The mean age of the 550 participants was 35.3 years (*SD* = 11.4 years). Of the 550 participants, 355 were male, 190 were female, and 5 preferred not to respond. The countries with the most participants were Venezuela (*n* = 306), USA (*n* = 37), and Russia (*n* = 20). On average, it took the participants 50.04 minutes to complete the study (*SD* = 19.7 min, median = 44.4 min).

From the recruitment via acquaintances, all participants were included. The results presented in this appendix are based on 550 crowdsourced and the 21 additionally recruited participants together (*N* = 571). It can be seen that the results are similar to those obtained while applying strict filtering.

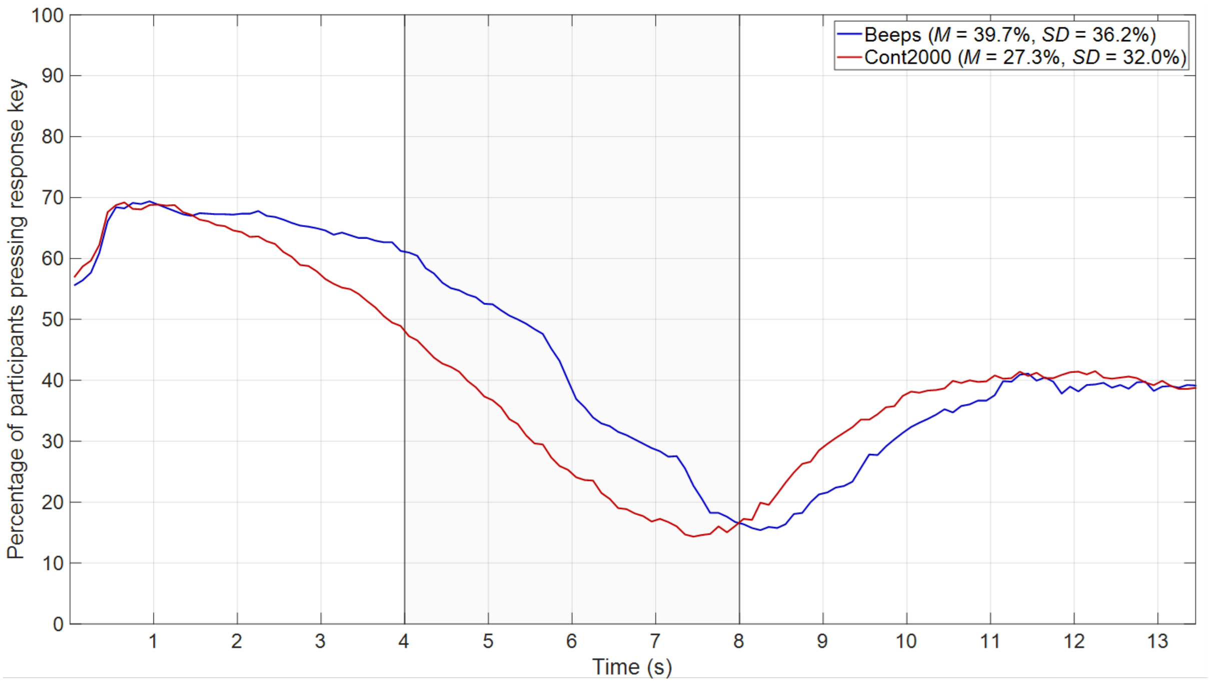

Figure A1. Key press percentage as a function of elapsed time for two selected samples. The legend shows the mean and SD of the performance score.

Table A1
Means, standard deviations, and correlation coefficients of the variables for the auditory stimuli (n = 30).

	Mean	*SD*	1	2	3	4
1 Performance score (%)	64.9	4.12				
2 Easy to notice (0–10)	5.56	0.18	0.44 (0.00)			

3	Gave enough information (0–10)	5.35	0.16	0.41 (0.04)	0.97 (0.95)		
4	Annoying (0–10)	4.22	0.33	0.52 (-0.03)	0.91 (0.84)	0.84 (0.79)	
5	Loudness score (sones)	33.1	7.29	0.78 (0.70)	0.69 (0.18)	0.63 (0.19)	0.78 (0.12)

Note. Numbers in parentheses represent the average correlation computed for background noise (n = 15) and no background noise (n = 15).

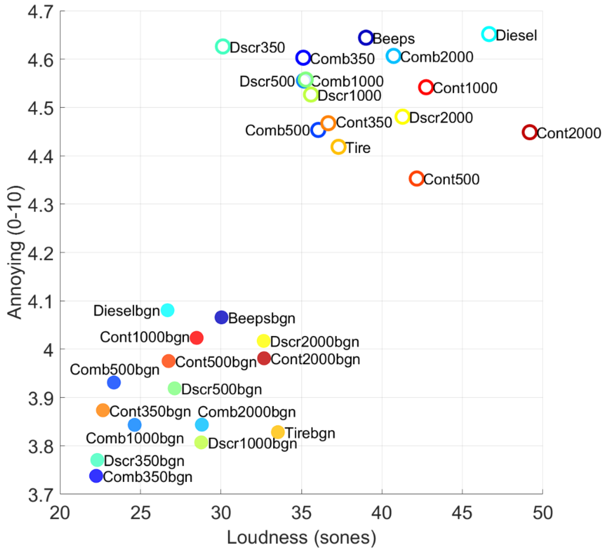

Figure A2. Scatter plot of perceived annoyance and computed loudness score of the 30 stimuli.

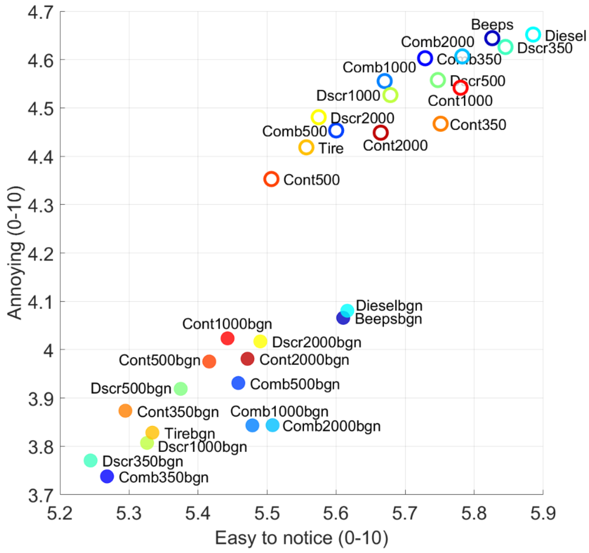

Figure A3. Scatter plot of perceived annoyance and perceived ‘easy to notice’ of the 30 stimuli.

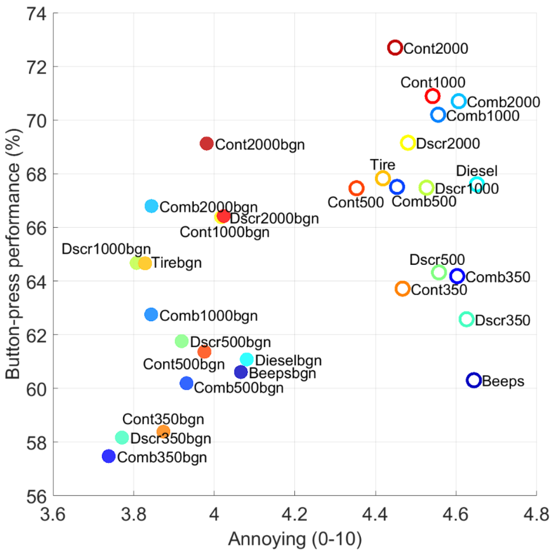

Figure A4. Scatter plot of performance (based on key-press inputs) and perceived annoyance of the 30 stimuli.