Synthetic Sounds for Automated Vehicles: Loud is Effective

Sounds for Automated Vehicles

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 a different approach by examining the efficacy of synthetic vehicle sounds in refraining 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 straight vehicle trajectory with 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 a key as long as they felt safe to cross, and 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. These results raise the question of whether electric vehicles emitting naturalistic sounds or non-annoying tones are optimally safe for pedestrians.

CCS CONCEPTS • Human-centered computing → Human computer interaction (HCI); HCI design and evaluation methods; Laboratory experiments; Interaction paradigms; Sound-based input / output

Additional Keywords and Phrases: Electric vehicles, Synthetic sound signals, Psychoacoustic annoyance, Road safety, Listening experiments

1 INTRODUCTION

More than 300,000 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 might help against 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 stop sales 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 could 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) so that the vehicle emits 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]. A common characteristic of both regulations is that they primarily focus on the A-weighted sound pressure level without further specification of the type of sound to be used.

Various types of artificial sounds for EVs have been studied, 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], but also 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] lead to 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 EV studies investigated the relationship between detectability and annoyance. Lee et al. [22] reported that sounds with amplitude and frequency modulation led to faster detection and lower perceived annoyance than saw-tooth signals, whereas Petiot et al. [27] optimized sounds by means of an interactive genetic algorithm and showed that it is possible to create sounds that, in the presence of background noise, are both highly detectable and pleasant.

The effect of sounds on participants' willingness to cross has not been extensively investigated. An exception is Emerson et al. [31], in which visually impaired participants were asked to indicate their willingness to cross by pressing and releasing a button in an on-road study. The authors reported that for the car model that did not switch to internal combustion until 48 km/h when accelerating from a stop, half of the approaches were not detected by the participants.

1.2 Sound design for automated vehicles

Next to EVs, auditory signals could also be useful in the form of auditory external human-machine interfaces (eHMIs) that communicate the state or intention of automated vehicles (AVs) to VRUs. The majority of eHMIs developed so far are visual, but a number of auditory eHMIs have 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 This study

The aim of this study is to examine what type of synthetic sounds could be used as auditory eHMI to inform pedestrians about an approaching vehicle. We tested pure tones, combined tones, intermittent tones, beeps, and engine/tire sounds, presented with and without background noise, and investigated their effect on perceptual measures (whether the sound is easy to notice, whether it gives enough information to realize a vehicle is approaching, and whether it is annoying), as well as willingness to cross the road as a behavioral measure. We expect detectability to increase with increasing loudness and tonal frequency, whereas annoyance is likely to increase with loudness [37-39].

2 METHOD

2.1 Sound stimuli

Thirty sound stimuli were synthetically generated. These sounds can be divided 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 (1), (2), and (3) were provided at four frequencies: 350 Hz, 500 Hz, 1000 Hz, and 2000 Hz. The combined tones (3) were the same as the continuous pure tones (1) but with two secondary tones at frequencies ±90 Hz from the main tone. The combined tones were included because they are expected to be perceived as more pitch-salient than pure tones [42]. The double-beep signal (4) consisted of eight pairs of 240-ms beeps with a 100-ms interval between beeps within 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 are 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 + tires + ICEV; 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. In the cases without background noise, no tire noise was added either to test the pure effect of the synthetic sound. The stimuli with background noise were less loud than the corresponding stimuli without background noise to assess a more realistic signal-to-noise ratio in practical applications (Figure 1). 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. 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 (12 x [500 ms emitting, 500 ms not emitting]), 350 Hz
6	Dscr500	Pure tone, intermittent (12 x [500 ms emitting, 500 ms not emitting]), 500 Hz
7	Dscr1000	Pure tone, intermittent (12 x [500 ms emitting, 500 ms not emitting]), 1000 Hz
8	Dscr2000	Pure tone, intermittent (12 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

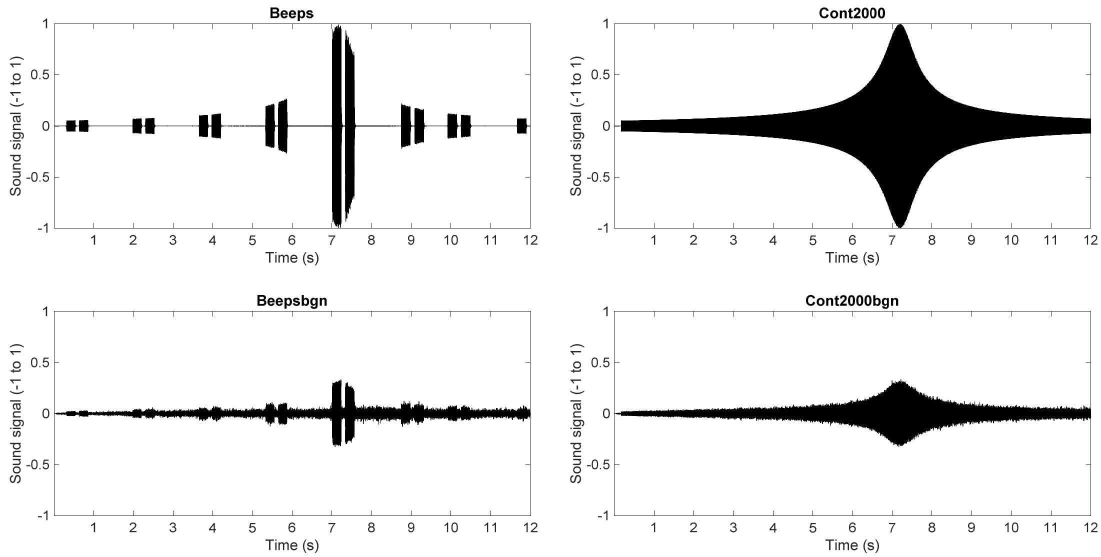

Figure 1: Sound signal as a function of the elapsed time for two example stimuli without background noise (upper half) and the same stimuli with background noise (bgn) (lower half)

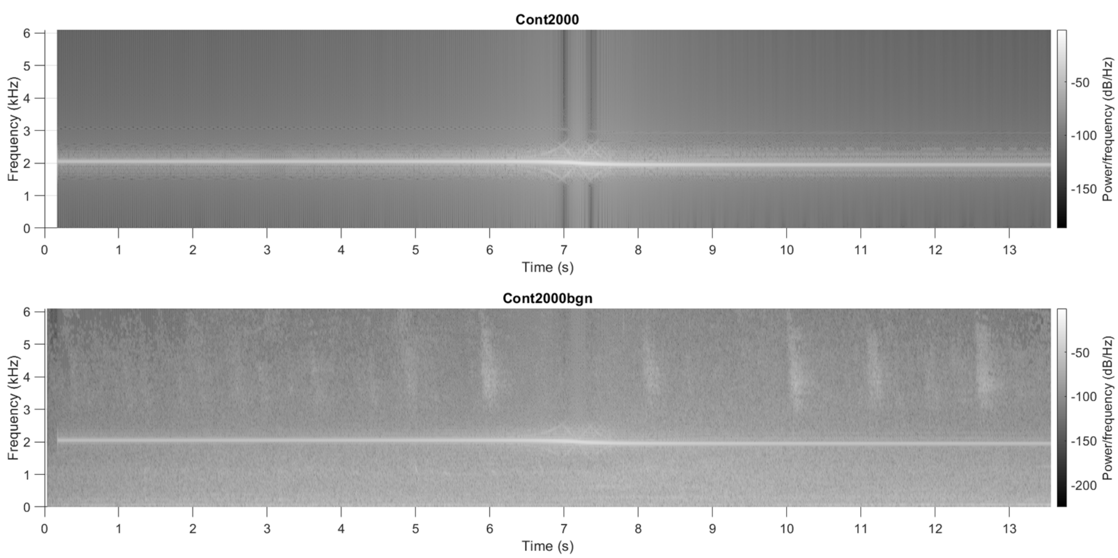

Figure 2: Spectrogram of stimuli Cont2000 (up) and Cont2000bgn (bottom)

2.2 Moving source effects

For a stationary observer, the perception of a sound emitted by a moving source changes over time due to the relative motion of the source. Consider a sound source with velocity vector $\boldsymbol{V}$, position vector $\boldsymbol{r}$ with respect to a stationary observer, and relative angle $\boldsymbol{\theta}$ (Figure 3). Assuming a quiescent propagation medium (i.e., no wind), the amplitude of the received sound signal by the observer at a certain instant t, $\chi(t)$, can be expressed as [45]:

$$\chi(t) = \frac{p(\tau_e)}{4\pi\|\boldsymbol{r}\|\left(1-M\cos(\theta)\right)^2} \qquad (1)$$

where $p(\tau e)$ is the sound signal emitted at the source position and at the emission time τe, and M is the Mach number, defined as $M = \|\boldsymbol{V}\| / c$, where c is the speed of sound in the medium, here considered 343 m/s. The term between parentheses in the denominator of Eq. (1) represents the convective amplification of sound.

The observer time t is the sum of the emission time τe and the travel time of the sound between the source and the observer [46]:

$$t = \tau_e + \frac{\|\boldsymbol{r}\|}{c} \qquad (2)$$

The effect of the source motion causes the well–known frequency shift due to the Doppler effect [47]:

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

On the basis of Eqs. (1) and (3), the sound emissions of a source moving towards the observer are perceived as a signal with a higher frequency and amplitude due to the convective amplification of the sound waves and vice versa.

The denominator of Eq. (1) contains the effect of geometric spreading of sound with the term $\|\boldsymbol{r}\|$. However, for the propagation of sound through air, the atmospheric absorption of sound is also considered a frequency- and weather-dependent term [48, 49]. For the development of the stimuli, an ambient temperature of 20ºC, an ambient pressure of 101.325 Pa, and relative humidity of 60% were considered.

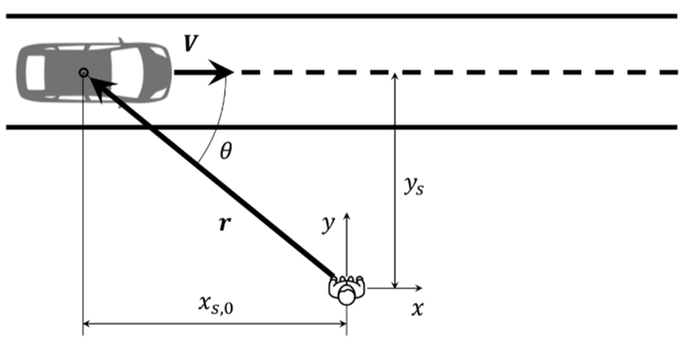

Figure 3: Schematic representation of the EV approaching a stationary pedestrian and the corresponding geometric configurations for the moving source effect

The current study considers a relatively simple sound source and observer geometry. First, a two-dimensional arrangement is employed, i.e., the observer and the sound source are in the same plane, which means that the reflection of the sound from buildings or other structures is neglected [50]. Furthermore, the sound source is considered as a point source with a constant velocity V of 30 km/h (i.e., 8.33 m/s, $M \approx 0.025$) in the positive x direction, i.e., the source has a linear trajectory along a straight line at a distance $ys = 3$ m from the observer (Figure 2). The initial position of the sound source at $\tau e = 0$ is defined at $\boldsymbol{r}(0) = (xs{,}0, ys)$ and at a general instant τe as $r(\tau e) = (xs{,}0 + V\,\tau e, ys)$. For the generation of the sound stimuli, the source moved from $xs,0 = -60$ m to $xs = 60$ m. Consequently, a signal length of 14.4 s was generated in all cases.

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. It was not permitted to complete the experiment more than once with 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 old. Information about anonymity and voluntary participation was provided as well.

The participants first answered demographic questions 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 experiment was created using a modified version of the framework based on jsPsych [51] that was used in a previous study on the measurement of reaction times to auditory, visual, and multi-modal stimuli [52], 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 [53, 54]. The sounds were pre-loaded before the start of the experiment to prevent delays during the experiment.

The participants were asked to sit in a quiet room and use their computer's built-in loudspeaker system. To adjust the volume properly, they had to listen to a song [55] with a normalized low volume to ensure that the users adjusted their volume high enough to hear the stimuli properly. They were asked not to change the volume level till the end of the experiment.

The participants had to respond to 60 sounds presented in blocks. Participants randomly either listened to sounds with background noise (30 trials) followed by trials without background noise (30 trials) or vice versa. Each sound was presented twice per block, presented in a random order that differed for each participant. After every ten trials, the participants were shown the following text: "*You have now completed 10 (alternatively 20, 30, 40 or 50) sounds out of 60. When ready press 'C' to proceed to the next batch.*"

The participants were asked to start holding the 'F' key before each sound stimulus. Pressing the key would start the sound playback. The participants had to release the key when they believed it became unsafe to cross and press again when it was safe to cross (they were allowed to press and release multiple times during each 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 moving all sliders (Figure 4). Participants did not receive feedback on their responses.

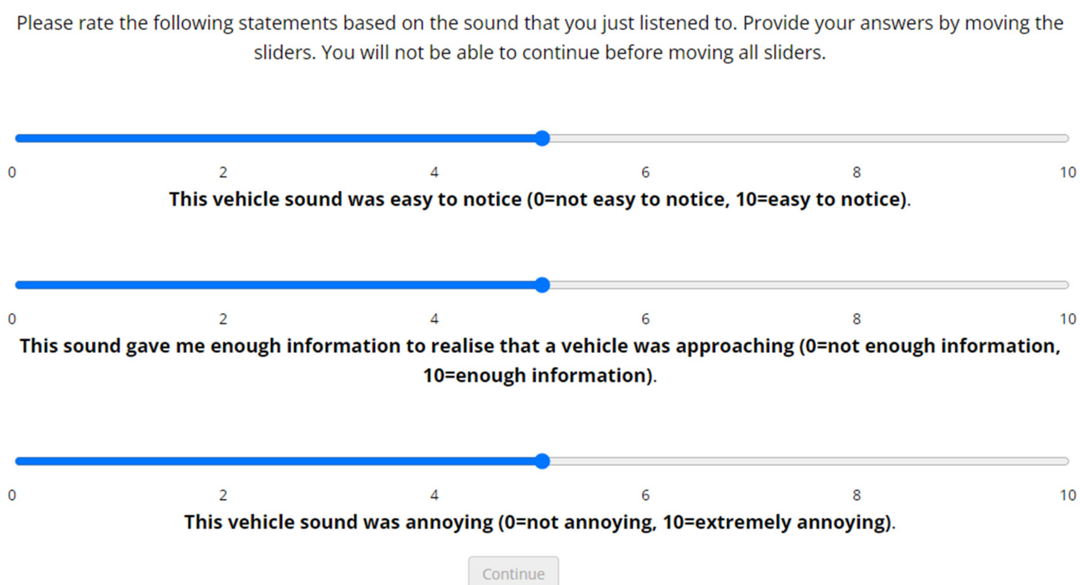

Figure 4: Set of questions that the participants received after listening to each sound

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 with the British English Amy female voice available at [56]. 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.

At the end of the experiment, the participants were shown a unique code. They were required to enter the code on the questionnaire to prove that they completed the experiment to receive their remuneration.

2.5 Data filtering

Participants who entered the same finishing code multiple times (n = 29), 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 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. If people completed the study more than once from the same IP address, only the first response was kept.

2.6 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 measurement method of noise that takes into account characteristics of human hearing (e.g., the dependence of sound transmission through the middle ear on frequency) [57]. The method has been proven able to distinguish between similar vehicle sounds [58].

Pearson correlation coefficients were computed between the performance scores, loudness, 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 over the numbers of participants.

3 RESULTS

Via the crowdsourcing study, 1,000 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"). In total, 885 participants were removed because of a database error and the strict filtering described in 2.5, leaving 110 participants from 33 countries for the analysis.

The 110 participants had a mean age of 35.8 years (standard deviation, SD = 11.4 years). Of the 110 participants, 74 were male, 35 were female, and 1 person 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 recruitment 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, where the continuous tone was a stronger deterrent to cross (i.e., participants released the key earlier) than the beeping sound.

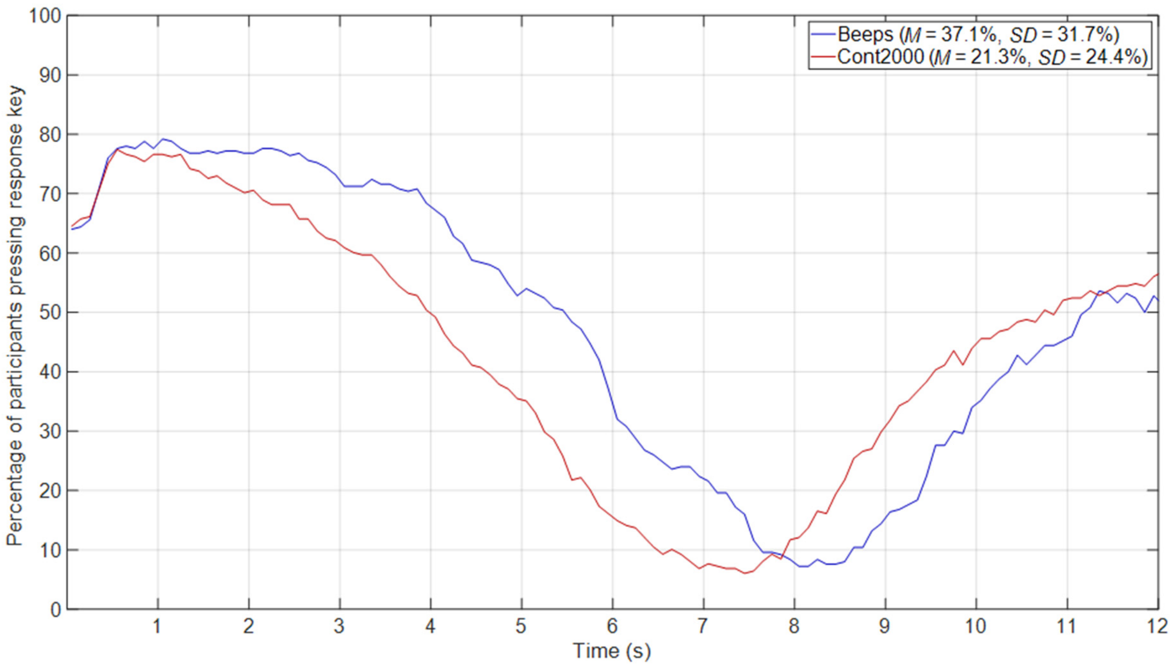

Figure 5: Key press percentage as a function of elapsed time for two selected samples. The legend shows the mean and SD of the performance score

Table 2 shows the correlation matrix between the performance score, the loudness for the 4–8 s interval, and the self-reported measures, all averaged over the 125 participants. From Table 2 and Figure 6, it can be seen that, generally, sounds that yielded a better performance score were also perceived as loud ($r = 0.81$) and annoying ($r = 0.58$).

Table 1: Sound stimuli included in the experiment

		Mean	*SD*	1	2	3	4
1	Performance score (%)	67.3	5.74				
2	Easy to notice (0–10)	5.03	0.28	0.51			
3	Gave enough information (0–10)	4.72	0.21	0.41	0.92		
4	Annoying (0–10)	4.21	0.42	0.58	0.91	0.78	
5	Loudness score (sones)	33.1	7.29	0.81	0.76	0.66	0.82

From Figures 7 and 8, it can also be recognized that a continuous high-frequency tone 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 of low tonal frequency, yielded relatively low performance, even without background noise (63% for Beeps, 62% for Dscr350).

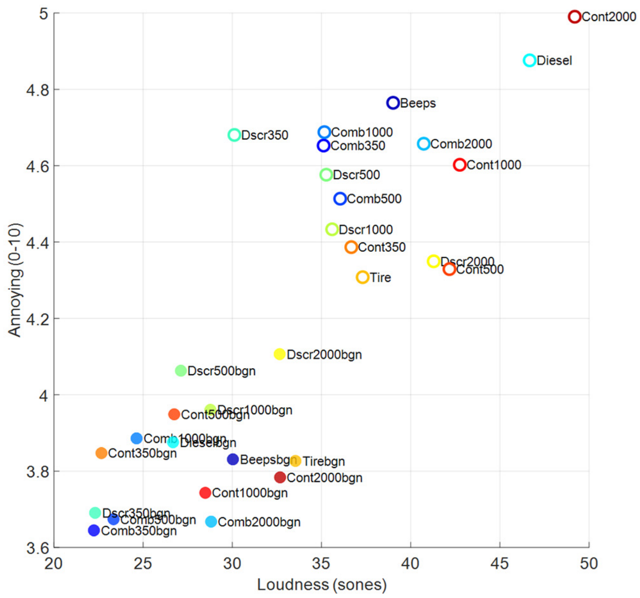

Figure 6: Scatter plot of perceived annoyance and loudness of the 30 stimuli

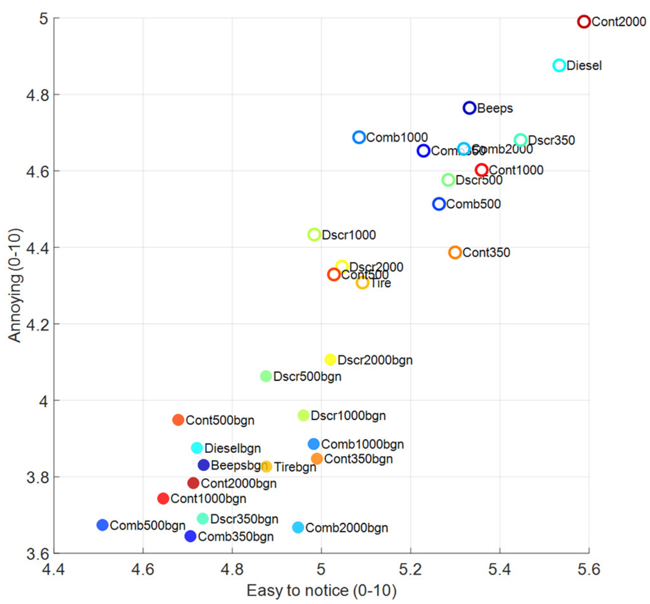

Figure 7: Scatter plot of perceived annoyance and perceived ‘easy to notice’ of the 30 stimuli

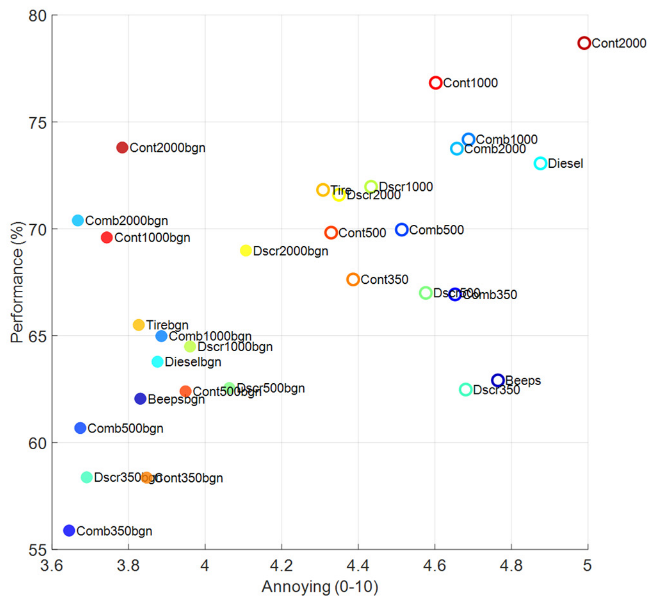

Figure 8: Scatter plot of performance (based on key-press inputs) and perceived annoyance of the 30 stimuli

4 DISCUSSION

This online study tested how vehicle sounds, naturalistic ones as well as tones and beeps, affect participants' crossing intentions. Participants were not presented with any visual information to test the effect of sound only, a situation that may be representative of cases in which visual information is unavailable due to visual impairment or visual obstruction.

The results showed that loud sounds were the most effective in refraining participants from crossing the road. The correlation between the performance score and the loudness of the stimulus was strong ($r = 0.81$). Our findings also indicate that high-performing sounds are annoying. This resonates with the so-called "trade-off hypothesis of pleasantness and power" that was reported by Bisping [59] 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 [60] 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 non-annoying tones [28], are optimally safe for pedestrians.

Beeps and intermittent sounds, especially of low tonal frequency, yielded relatively low performance, even in the absence of background noise. A possible explanation is that the inter-beep intervals we used were relatively large (1000 ms). Other studies investigated shorter intervals and found that perceived urgency increases with increasing pulse level and decreasing inter-pulse interval [42, 61], following Stevens' power law [42]. Slow beeps are typically used for relatively slowly evolving situations, such as a truck reversing, whereas fast beeps indicate an approaching hazard. Moreover, the duration of the beeps themselves is short, which inhibits speed estimation. At the same time, during the non-emitting intervals, particularly if these are large, no new information is provided, creating a lag in information processing.

A limitation of this study is that only a small range of relatively simple signals was tested. Moreover, only non-verbal stimuli were tested because, while verbal messages have rich semantic content, non-verbal ones can convey information faster and more effectively in the presence of noise [63]. Nevertheless, a direct comparison of verbal and non-verbal sounds would deserve further investigation.

In conclusion, we found a strong relationship between the loudness of a sound and its effectiveness in preventing participants from crossing the road. This finding has consequences for the quest for pleasant vehicle sounds.

5 SUPPLEMENTARY MATERIAL

The questionnaire used in Appen, sound stimuli, song for volume adjustment, test phrases, anonymized data, and MATLAB code will be uploaded after acceptance to a public repository. The codes for the experiment setup and the analysis will be available at GitHub.

REFERENCES

[1] World Health Organization. 2013. *Pedestrian safety. A road safety manual for decision-makers and practitioners*. Retrieved October 12, 2022 from http://apps.who.int/iris/bitstream/handle/10665/79753/9789241505352_eng.pdf

[2] Marco P. DaSilva, John D. Smith, and Wassim G. Najm. 2003. *Analysis of pedestrian crashes.* Report No. DOT-VNTSC-NHTSA-02-02. John A. Volpe National Transportation Systems Center, Cambridge, MA. Retrieved October 12, 2022 from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2003/DOTHS809585.pdf

[3] Azra Habibovic and Johan Davidsson. 2012. Causation mechanisms in car-to-vulnerable road user crashes: Implications for active safety systems. *Accident Analysis & Prevention 49* (November 2012), 493-500. DOI: 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).* Retrieved October 12, 2022 from https://www.ghsa.org/sites/default/files/2022-05/Pedestrian%20Traffic%20Fatalities%20by%20State%20-%202021%20Preliminary%20Data%20%28January-December%29.pdf

[5] András Bálint, Volker Labenski, Markus Köbe, Carina Vogl, Johan Stoll, Lars Schories, Lena Amann, Ganesh Baroda Sudhakaran, Pedro HuertasLeyva, Thomas Pallacci, Martin Östling, Daniel Schmidt, D., and Ron Schindler. 2021. *Use case definitions and initial safety-critical scenarios.* Report No. D2.6. Project SAFE-UP.

[6] Jameson Dow. 2021. Norway bans gas car sales in 2025, but trends point toward 100% EV sales as early as April. (September 2021). Retrieved October 9, 2022 from https://electrek.co/2021/09/23/norway-bans-gas-cars-in-2025-but-trends-point-toward-100-ev-sales-as-early-as-april

[7] Sam Morgan. 2018. Denmark to ban petrol and diesel car sales by 2030. (October 2018). Retrieved October 9, 2022 from https://www.euractiv.com/section/electric-cars/news/denmark-to-ban-petrol-and-diesel-car-sales-by-2030

[8] Fred Lambert. 2017. The Dutch government confirms plan to ban new petrol and diesel cars by 2030. (October 2017). Retrieved October 9, 2022 from 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. (November 2020). Retrieved October 9, 2022 from 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] Sara Liu, Michael Fitzharris, Jennie Oxley, and Chris Edwards. 2018. The impact of electric / hybrid vehicles and bicycles on pedestrians who are blind or have low vision. Accident Research Centre. Monash University.

[11] Etienne Parizet, Wolfgang Ellermeier, and Ryan Robart. 2014. Auditory warnings for electric vehicles: Detectability in normal-vision and visually-impaired listeners. *Applied Acoustics 86* (December 2014), 50-58. DOI: 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.* Retrieved October 9, 2022 from 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). Retrieved October 9, 2022 from 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. Retrieved October 9, 2022 from https://unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2017/R138r1e.pdf

[15] Michael S. Wogalter, Raymond W. Lim, and Patrick G. Nyeste. 2014. On the hazard of quiet vehicles to pedestrians and drivers. *Applied Ergonomics 45*, 5 (September 2014), 1306-1312. DOI: https://doi.org/10.1016/j.apergo.2013.08.002

[16] Katsuya Yamauchi, Daniel Menzel, Masayuki Takada, Koji Nagahata, Shin-ichro Iwamiya, and Hugo Fastl. 2015. Psychoacoustic examination of feasible level of additional warning sound for quiet vehicles. *Acoustical Science and Technology 36*, 2, 120-125. DOI:

https://doi.org/10.1250/ast.36.120

[17] Sneha Singh, Sarah R. Payne, and Paul A. Jennings. 2014. Toward a methodology for assessing electric vehicle exterior sounds. *IEEE Transactions on Intelligent Transportation Systems 15*, 4 (August 2014), 1790-1800. DOI: https://doi.org/10.1109/tits.2014.2327062

[18] Min-Chih Hsieh, Hung-Jen Chen, Ming-Le Tong, and Cheng-Wu Yan. 2021. Effect of environmental noise, distance and warning sound on pedestrians' auditory detectability of electric vehicles. *International Journal of Environmental Research and Public Health 18*, 17, Article 9290 (September 2021), 16 pages. DOI: https://doi.org/10.3390/ijerph18179290

[19] Sylvain Fleury, Éric Jamet, Vincent Roussarie, Laure Bosc, and Jean-Christophe Chamard. 2016. Effect of additional warning sounds on pedestrians' detection of electric vehicles: An ecological approach. *Accident Analysis & Prevention 97* (December 2016), 176-185. DOI: https://doi.org/10.1016/j.aap.2016.09.002

[20] Nicolas Misdariis, Andrea Cera, Eugenie Levallois, and Christophe Locqueteau. 2012. Do electric cars have to make noise? An emblematic opportunity for designing sounds and soundscapes. *Acoustics 2012*, April 23-27, 2012, Nantes, France. Retrieved October 10, 2022 from https://hal.archives-ouvertes.fr/hal-00810920

[21] Nicola Frlić Sekulić, Ivan Blagojević, Vladimir Popović, Dragan Stamenković, and Slobodan Janković. 2019. Development of pedestrian alert system for use in electric vehicles. *Tehnički Vjesnik 26*, 6, 1614-1619. DOI: https://doi.org/10.17559/TV-20180730144329

[22] Sang Kwon Lee, Seung Min Lee, Taejin Shin, and Manug Han. 2017. Objective evaluation of the sound quality of the warning sound of electric vehicles with a consideration of the masking effect: Annoyance and detectability. *International Journal of Automotive Technology 18*, 4, 699-705. https://doi.org/10.1007/s12239-017-0069-6

[23] Katsuya Yamauchi, Daniel Menzel, Hugo Fastl, Masayuki Takada, Koji Nagahata, and Shin-ichiro Iwamiya. 2011. Cross-cultural study on feasible sound levels of possible warning signals for quiet vehicles. *INTER-NOISE*, September 4-7, 2011, Osaka, Japan.

[24] Nikolaos Kournoutos and Jordan Cheer. 2019. An environmentally adaptive warning sound system for electric vehicles. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 259*, 9, June 16-19, 2019, Madrid, Spain, Institute of Noise Control Engineering, 683-691.

[25] Pedro Poveda-Martínez, Ramón Peral-Orts, Nuria Campillo-Davo, Josue Nescolarde-Selva, Miguel Lloret-Climent, and Jaime Ramis-Soriano. 2017. Study of the effectiveness of electric vehicle warning sounds depending on the urban environment. *Applied Acoustics 116* (January 2017), 317-328. DOI: https://doi.org/10.1016/j.apacoust.2016.10.003

[26] Jean-François Petiot, Bjørn G. Kristensen, and Anja M. Maier. 2013. How should an electric vehicle sound? User and expert perception. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference 55928*, Article DETC2013-12535, August 4–7, 2013, Portland, Oregon, American Society of Mechanical Engineers, 12 pages. https://doi.org/10.1115/DETC2013-12535

[27] Jean-François Petiot, Killian Legeay, and Mathieu Lagrange. 2019. Optimization of the sound of electric vehicles according to unpleasantness and detectability. *Proceedings of the Design Society: International Conference on Engineering Design 1*, 1, August 5-8, 2019, Delft, The Netherlands, Cambridge University Press, 3949-3958. https://doi.org/10.1017/dsi.2019.402

[28] Sneha Singh, Sarah R. Payne, and Paul A. Jennings. 2014. Toward a methodology for assessing electric vehicle exterior sounds. *IEEE Transactions on Intelligent Transportation Systems 15*, 4 (June 2014), 1790-1800. https://doi.org/10.1109/TITS.2014.2327062

[29] Gustav Jacobsen, Wookeun Song, and Ewen Macdonald. 2016. Predicting detectability and annoyance of EV warning sounds using partial loudness. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 253*, 7, August 21-24, 2016, Hamburg, Germany, Institute of Noise Control Engineering, 886-895.

[30] Katsuya Yamauchi, Takaichi Sano, Shin Hasegawa, Fumio Tamura, and Yuichiro Takeda. 2014. Detectability and hearing impression of additional warning sounds for electric or hybrid vehicles. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 249*, 2, November 16-19, 2014, Melbourne, Australia, Institute of Noise Control Engineering, pp. 5279-5285.

[31] Robert Wall Emerson, Koorosh Naghshineh, Julie Hapeman, and William Wiener. 2011. A pilot study of pedestrians with visual impairments detecting traffic gaps and surges containing hybrid vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour 14*, 2 (March 2011), 117-127. DOI: https://doi.org/10.1016/j.trf.2010.11.007

[32] Debargha Dey, Azra Habibovic, Andreas Löcken, Philipp Wintersberger, Bastian Pfleging, Andreas Riener, Marieke Martens, and Jacques Terken. 2020. Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles. *Transportation Research Interdisciplinary Perspectives 7*, Article 100174 (September 2020), 24 pages. DOI: https://doi.org/10.1016/j.trip.2020.100174

[33] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. *CHI Conference on Human Factors in Computing Systems*, April 21-26, 2018, Montréal, Canada, Article 429. DOI: https://doi.org/10.1145/3173574.3174003

[34] Shuchisnigdha Deb, Lesley J. Strawderman, and Daniel W. Carruth. 2018. Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transportation Research Part F: Traffic Psychology and Behaviour 59*, Part A (November 2018), 135-149. DOI: https://doi.org/10.1016/j.trf.2018.08.016

[35] Evelyn Florentine, Mark Adam Ang, Scott Drew Pendleton, Hans Andersen, and Marcelo H. Ang. 2016. Pedestrian notification methods in autonomous vehicles for multi-class mobility-on-demand service. *Fourth International Conference on Human Agent Interaction*, October 4-7, 2016, Biopolis, Singapore, 387-392. DOI: https://doi.org/10.1145/2974804.2974833

[36] Dylan Moore, Rebecca Currano, and David Sirkin. 2020. Sound decisions: How synthetic motor sounds improve autonomous vehicle-pedestrian interactions. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, September 21 - 22, 2020, Virtual Event, 94-103. DOI: https://doi.org/10.1145/3409120.3410667

[37] S. Fidell and S. Teffeteller. 1981. Scaling the annoyance of intrusive sounds. *Journal of Sound and Vibration 78*, 2 (September 1981), 291-298. DOI: https://doi.org/10.1016/s0022-460x(81)80039-9

[38] Sanford Fidell, Sherri Teffeteller, Richard Horonjeff, and David M. Green. 1979. Predicting annoyance from detectability of low-level sounds. *The*

Journal of the Acoustical Society of America 66, 5 (November 1979), 1427-1434. DOI: https://doi.org/10.1121/1.383536

[39] Birgitta Berglund, Ulf Berglund, and Thomas Lindvall. 1976. Scaling loudness, noisiness, and annoyance of community noises. *The Journal of the Acoustical Society of America 60*, 5, 1119-1125. DOI :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. (May 2014). Retrieved October 9, 2022 from https://www.youtube.com/watch?v=2Y33bTlAA-E

[41] Zero Tailpipe. 2018. What sound does Hyundai iONIQ Electric make? Noisy from outside at speed? (September 2018). Retrieved October 9, 2022 from https://www.youtube.com/watch?v=X0wpizkkH_Q

[42] P. A. Cariani and B. Delgutte. 1996. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology 76*, 3 (September 1996), 1698-1716. DOI: https://doi.org/10.1152/jn.1996.76.3.1698

[43] Pavlo Bazilinskyy, Sebastiaan M. Petermeijer, Veronika Petrovych, Dimitra Dodou, and Joost C. F. de Winter. 2018. Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour 56* (July 2018), 82-98. DOI: https://doi.org/10.1016/j.trf.2018.04.001

[44] Articulated Sounds. 2017. Quiet streets ambience sound library. (July 2017). Retrieved October 9, 2022 from https://www.youtube.com/watch?v=6C-W_7BZBxQ

[45] P. Sijtsma. 2006. *Beamforming on moving sources.* Report No. NLR–TP–2006–733. National Aerospace Laboratory (NLR), Amsterdam, The Netherlands.

[46] Roberto Merino-Martínez, Mirjam Snellen, and Dick G. Simons. 2016. Functional beamforming applied to imaging of flyover noise on landing aircraft. *Journal of Aircraft 53*, 6 (June 2016), 1830-1843. DOI: https://doi.org/10.2514/1.c033691

[47] G. P. Howell, A. J. Bradley, M. A. McCormick, and J. D. Brown. 1986. De-Dopplerization and acoustic imaging of aircraft flyover noise measurements. *Journal of Sound and Vibration 105*, 1 (February 1986), 151-167. DOI: https://doi.org/10.1016/0022-460x(86)90227-0

[48] Roberto Merino-Martínez. 2018. *Microphone Arrays for Imaging of Aerospace Noise Sources.* Ph.D. Dissertation. Delft University of Technology, Delft, The Netherlands.

[49] Thomas Rossing. 2007. *Springer Handbook of Acoustics* (2nd ed.). Springer Science & Business Media.

[50] Keith Attenborough, Kai Ming Li, and Kirill Horoshenkov. 2006. *Predicting Outdoor Sound* (1st ed.). CRC Press.

[51] Joshua R. De Leeuw. 2014. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods 47*, 1, 1-12. DOI: https://doi.org/10.3758/s13428-014-0458-y

[52] Pavlo Bazilinskyy and Joost de Winter. 2018. Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony. *Human Factors 60*, 8, 1192-1206. DOI: https://doi.org/10.1177/0018720818787126

[53] Pavlo Bazilinskyy, Lars Kooijman, Dimitra Dodou, and Joost C. F. de Winter. 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. *Applied Ergonomics 95* (September 2021), 103450. DOI: https://doi.org/10.1016/j.apergo.2021.103450.

[54] Anirudh Sripada, Pavlo Bazilinskyy, and Joost de Winter. 2021. Automated vehicles that communicate implicitly: examining the use of lateral position within the lane. *Ergonomics 64*, 11, 1416-1428. DOI: https://doi.org/10.1080/00140139.2021.1925353

[55] Royalty Free Music - No Copyright Music. 2018. Creative minds - Bensound. (March 2018). Retrieved October 9, 2022 from https://www.youtube.com/watch?v=zFBwdFnezG0

[56] TTSMP3. 2022. Free text-to-speech and text-to-mp3 for US English. Retrieved October 9, 2022 from https://ttsmp3.com

[57] Eberhard Zwicker and Bertram Scharf. 1965. A model of loudness summation. *Psychological Review 72*, 1 (January 1965), 3-26. DOI: https://doi.org/10.1037/h0021703

[58] D. J. Swart and A. Bekkers. 2016. Interior and Motorbay sound quality evaluation of full electric and hybrid-electric vehicles based on psychoacoustics. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 253*, 1, August 21-24, 2016, Hamburg, Germany, Institute of Noise Control Engineering, 7029-7039.

[59] Rudolf Bisping. 1995. Emotional effect of car interior sounds: Pleasantness and power and their relation to acoustic key features. *SAE Technical Paper Series 104*, 6 (May 1995), 2207-2213. DOI: https://doi.org/10.4271/951284

[60] Wanlu Yang, Kai Makita, Takashi Nakao, Noriaki Kanayama, Maro G. Machizawa, Takafumi Sasaoka, Ayako Sugata, Ryota Kobayashi, Ryosuke Hiramoto, Shigeto Yamawaki, Makoto Iwanaga, and Makoto Miyatani. 2018. Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E). *Behavior Research Methods 50*, 4, 1415-1429. DOI: https://doi.org/10.3758/s13428-018-1027-6

[61] Ellen C. Haas and John G. Casali. 1995. Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics 38*, 11, 2313-2326. DOI: https://doi.org/10.1080/00140139508925270

[62] Judy Edworthy. 1994. The design and implementation of non-verbal auditory warnings. *Applied Ergonomics 25*, 4 (August 1994), 202-210. DOI: 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. In that case, 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 impossible 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 participants were removed, leaving 550 participants from 50 countries for the analysis.

The 550 participants had a mean age of 35.3 years (standard deviation, SD = 11.4 years). Of the 550 participants, 355 were male, 190 were female, and 5 persons preferred not to respond. The countries that were most represented were Venezuela (n = 306), USA (n = 37), and Russia (n = 20). The participants took, on average, 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 with those obtained while applying strict filtering.

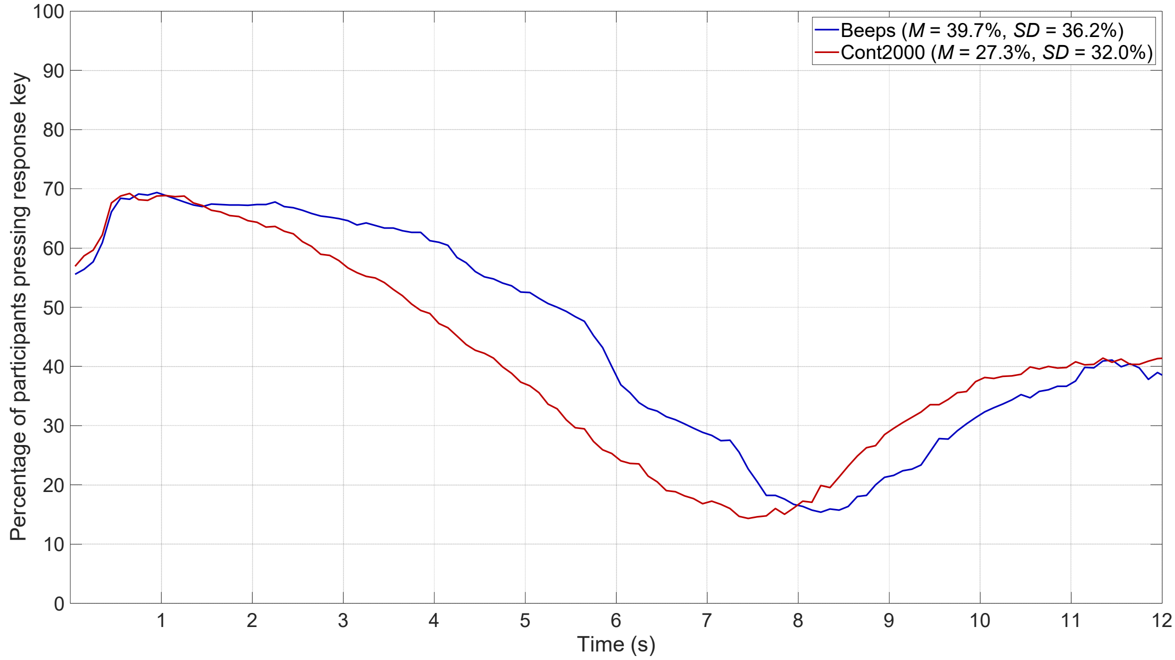

Figure A1: Key press percentage as a function of elapsed time for two selected samples. The legend shows the mean and SD of the performance score

Table A1: Sound stimuli included in the experiment

		1	2	3	4
1	Performance score (%)				
2	Easy to notice (0–10)	0.44			
3	Gave enough information (0–10)	0.41	0.97		
4	Annoying (0–10)	0.52	0.91	0.84	
5	Loudness score (sones)	0.78	0.69	0.63	0.78

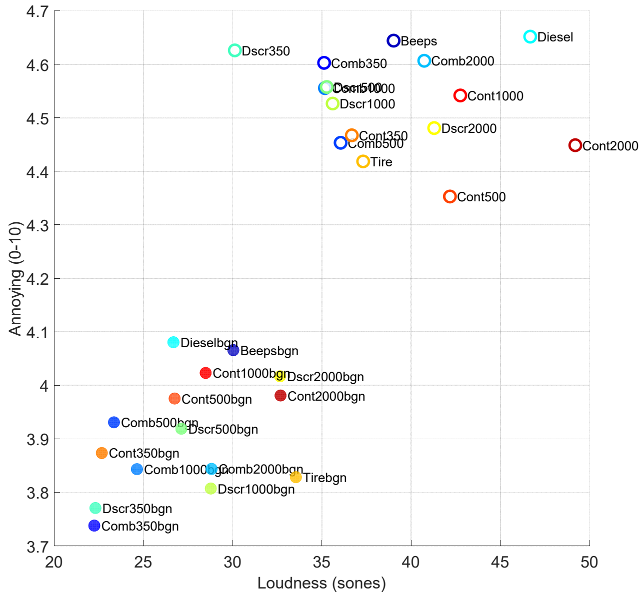

Figure A2: Scatter plot of perceived annoyance and loudness of the 30 stimuli

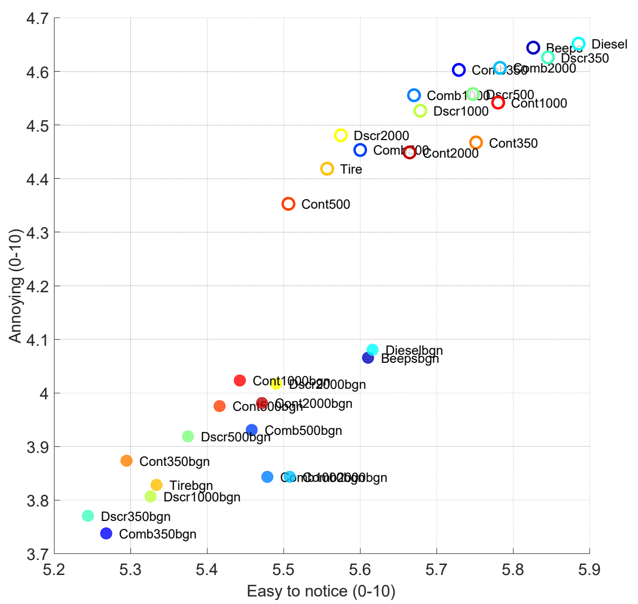

Figure A3: Scatter plot of perceived annoyance and perceived ‘easy to notice’ of the 30 stimuli

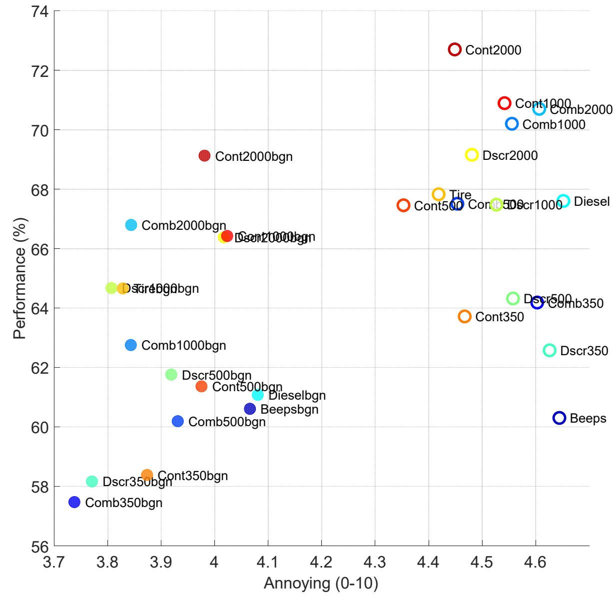

Figure A4: Scatter plot of performance (based on key-press inputs) and perceived annoyance of the 30 stimuli