Take-over requests in highly automated driving
A crowdsourcing survey on auditory, vibrotactile, and visual displays

Bazilinskyy, P.; Petermeijer, S. M.; Petrovych, V.; Dodou, D.; de Winter, J. C.F.

DOI
10.1016/j.trf.2018.04.001

Publication date
2018

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Transportation Research Part F 56 (2018) 82–98

Contents lists available at ScienceDirect

Transportation Research Part F

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

Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays

P. Bazilinskyy [a,*], S.M. Petermeijer [a], V. Petrovych [b], D. Dodou [a], J.C.F. de Winter [a]

[a] Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands
[b] Swedish National Road and Transport Research Institute, Linköping, Sweden

ARTICLE INFO

Article history:
Received 6 February 2017
Received in revised form 13 March 2018
Accepted 4 April 2018
Available online 21 April 2018

Keywords:
Multimodal interfaces
Automated driving
Crowdsourcing

ABSTRACT

An important research question in the domain of highly automated driving is how to aid drivers in transitions between manual and automated control. Until highly automated cars are available, knowledge on this topic has to be obtained via simulators and self-report questionnaires. Using crowdsourcing, we surveyed 1692 people on auditory, visual, and vibrotactile take-over requests (TORs) in highly automated driving. The survey presented recordings of auditory messages and illustrations of visual and vibrational messages in traffic scenarios of various urgency levels. Multimodal TORs were the most preferred option in high-urgency scenarios. Auditory TORs were the most preferred option in low-urgency scenarios and as a confirmation message that the system is ready to switch from manual to automated mode. For low-urgency scenarios, visual-only TORs were more preferred than vibration-only TORs. Beeps with shorter interpulse intervals were perceived as more urgent, with Stevens' power law yielding an accurate fit to the data. Spoken messages were more accepted than abstract sounds, and the female voice was more preferred than the male voice. Preferences and perceived urgency ratings were similar in middle- and high-income countries. In summary, this international survey showed that people's preferences for TOR types in highly automated driving depend on the urgency of the situation.

1. Introduction

1.1. Highly automated driving and take-over requests

Now that partially automated driving systems are in serial production, it is foreseen that highly automated driving will be deployed on public roads in the next one or two decades (see Begg, 2014; Underwood, 2014 for predictive surveys). Highly automated driving, a term introduced in 2006 (Flemisch et al., 2006), is defined as a technology that takes over both longitudinal and lateral control, and in which the driver is no longer required to permanently monitor the machine (Gasser & Westhoff, 2012). In highly automated driving, the driver is permitted to take hands and feet off the steering wheel and pedals and may engage in non-driving tasks such as checking the phone, reading a book, or resting. When the automation reaches

* Corresponding author at: Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail address: p.bazilinskyy@tudelft.nl (P. Bazilinskyy).

https://doi.org/10.1016/j.trf.2018.04.001
1369-8478/

its operational limit in a given traffic situation, the automation issues a so-called take-over request (TOR), asking the driver to take back control of the vehicle (Gasser & Westhoff, 2012; Hoeger et al., 2011).

The level of urgency of the situation with which the automation cannot cope is a critical parameter of the take-over process. Several scenarios are imaginable, such as (in order of increasing urgency):

1. The automated vehicle is reaching a target highway exit.
2. The automated vehicle has to make a lane change because it is approaching a slow-moving vehicle.
3. There are construction works on the road ahead, and at least one of the lanes is closed.
4. A technical failure prevents the automated vehicle from working properly.
5. An accident has just happened right in front of the automated vehicle.

In the first scenario, the automation can issue the TOR long in advance, so that the driver has ample time to resume manual control, whereas in the last scenario the traffic situation has changed abruptly, leaving the driver with little time for taking over. Recent studies (Gold, Damböck, Lorenz, & Bengler, 2013; Gold, Lorenz, Damböck, & Bengler, 2013; Mok et al., 2015; You, Wang, Wang, Zhu, & Hansen, 2017) have quantified the effect of the urgency of the take-over (sometimes called "time budget"; Gold et al., 2013, "time buffer"; Gasser & Westhoff, 2012, or "lead time"; Society of Automotive Engineers [SAE], 2014) on the driver's response time and on the quality of the take-over. For example, Gold et al. (2013) found that for shorter time budgets after an audio-visual TOR, drivers responded faster, but the take-over was of lower quality (hard braking, swerving, and inappropriate full stops). Choosing the right display for providing a TOR is important, especially in urgent scenarios, where a delay of a few tenths of a second in brake reaction could mean the difference between colliding and not colliding. Note that in this article, the term 'display' does not necessarily refer to a visual instrument, but to "any instrument or device that presents information to any sense organ (visual, auditory, or other)" (Swain & Guttmann, 1983, p. 2–3).

1.2. Displays for take-over requests in highly automated driving

1.2.1. Visual displays

Manual driving is primarily a visual task (Green, 1999; Sivak & Owens, 1996). Traditionally, visual information about the vehicle state (e.g., speed, RPM) as well as warnings (e.g., low fuel, high engine temperature) are provided on the dashboard. It is well established that the appropriate use of colour, saliency, and spatial positioning according to the principles of moving part and proximity compatibility can make a visual display easy to understand (e.g., Fitts & Jones, 1947; Grether, 1949). Accordingly, visual TORs during highly automated driving could be issued by lighting up an icon/region on the dashboard (e.g., Flemisch, Kaussner, Petermann, Schieben, & Schöming, 2011; Naujoks, Mai, & Neukum, 2014) or, more innovatively, on a head-up display (Kim, Jeong, Yang, Oh, & Kim, 2017; Langlois & Soualmi, 2016; see Manca, De Winter, & Happee, 2015, for an overview).

While visual information presented on the dashboard during manual driving may be detectable using peripheral vision (Lamble, Laakso, & Summala, 1999), such information might be missed if the driver is engaged in a non-driving task. A promising approach for conveying visual messages is to use ambient TORs, for example by lighting up a LED strip or bar under the windshield, along the A-pillar, or around the cabin (e.g., Kelsch & Dziennus, 2015; Löcken, Heuten, & Boll, 2015; Meschtscherjakov, Döttlinger, Rödel, & Tscheligi, 2015; Pfromm, Cieler, & Bruder, 2013; Winkler et al., 2018), or by using spatially directed LED strips (Dettman & Bullinger, 2017). Such ambient TORs might be more easily detected by the distracted driver than localized visual cues.

1.2.2. Auditory displays

Auditory displays have the advantage of being "gaze-free" (Meng & Spence, 2015; Stokes, Wickens, & Kite, 1990), meaning that the stimulus does not have to be in the field of view of the driver in order to be detected (for an overview of the advantages of auditory over visual displays, see Sanders & McCormick, 1987, as summarized by Stanton & Edworthy, 1999). Furthermore, it has been argued that the distinctive meaning of multiple auditory warnings can be easily learned (Blattner, Sumikawa, & Greenberg, 1989; Gaver, 1986). For example, Patterson (1982); see also Patterson and Mayfield (1990) showed that listeners were able to learn between four and six auditory warnings in only a few practice trials.

Auditory displays are widely used in consumer cars for warning about hazards in the outside environment, such as approaching objects during parking (BMW, 2013; Toyota, 2014; Volkswagen, 2014) or vehicles on a collision path (Graham, 1999; Honda, 2014; Ramkissoon, 2001; for reviews, see Nees & Walker, 2011; Wickens & Seppelt, 2002). A TOR in a highly automated vehicle could be conveyed with similar abstract (i.e., non-verbal) warning sounds, such as beeps and tones, whereas voice messages could be used to transfer higher levels of semantics (Naujoks, Forster, Wiedemann, & Neukum, 2016; Politis, Brewster, & Pollick, 2015a). In complex tasks, abstract auditory warnings may give faster initial responses than voice messages, because the duration of the latter is longer and humans tend to wait until the completion of the entire voice message before taking meaningful action (e.g., Bate, 1969).

Whether voice gender is important in auditory warnings has been a subject of discussion for several decades. A previous online survey showed that the female voice was strongly favoured over the male voice when used as a TOR (Bazilinskyy & De Winter, 2015), and a questionnaire on satellite navigation devices showed that a female voice is more often used for providing directions than a male voice (Large & Burnett, 2013). However, the results of studies on perceived urgency of a male versus a female voice are inconclusive. Park and Jang (1999) found that a male voice was perceived as more urgent than a female voice, whereas others reported the opposite effect (e.g., Hellier, Edworthy, Weedon, Walters, & Adams, 2002; see Edworthy and Hellier (2003) for a review). Jang (2007) tested male and female voices of equal fundamental frequency and loudness and found that the male voice was perceived as more urgent. Jang argued that voice characteristics such as smoothness and timbre may explain the differences in perceived urgency between male and female voices. In a study investigating verbal cockpit warnings, Arrabito (2009) found that a male voice (either monotone or urgent) led to a larger number of correct responses and faster response times than a female voice. Nixon et al. (1998) assessed the intelligibility of the male and female voice in military aircraft cockpits with noise levels ranging between 95 dB and 115 dB and found that the female voice tended to be less intelligible than the male voice, especially at the highest noise level (Nixon et al., 1998; see Noyes, Hellier, & Edworthy, 2006 for a review on speech warnings).

Looming sounds, which are sounds that increase in intensity with the criticality of the situation, are another type of auditory displays that might be useful in automated driving. Studies measuring brain activity with event-related functional magnetic resonance imaging (fMRI) have indicated that a sound that increases in intensity is associated with activation of the right amygdala and left temporal regions, suggesting that looming sounds are an intrinsic warning cue (Bach et al., 2008). Moreover, according to the principle of 'auditory tau' (Shaw, McGowan, & Turvey, 1991), it is theoretically possible to estimate time-to-contact based on the looming intensity of an acoustic signal (Gray, 2011; Silva et al., 2017). In a driving simulator study, Gray (2011) reported that a looming auditory warning led to brake reaction times that were 130 ms faster than the brake reaction times for an auditory warning with constant intensity.

1.2.3. Vibrotactile displays

Compared to visual and auditory displays, tactile displays are underused in the automotive domain (Jones & Sarter, 2008), but the interest in employing vibrations to convey information to the driver is growing rapidly (e.g., Birrell, Young, & Weldon, 2013; De Groot, De Winter, García, Mulder, & Wieringa, 2011; Grah et al., 2015; Meng, Ho, Gray, & Spence, 2015; for a review, see Petermeijer, De Winter, & Bengler, 2016). A specific advantage of vibrotactile displays is that they can provide information in a private manner (Petermeijer et al., 2016), whereas a disadvantage is that they can capture only a limited amount of information compared to auditory displays (Lu, Wickens, Sarter, & Sebok, 2011) and may not be suitable for issuing multiple alerts. For example, Fitch, Hankey, Kleiner, and Dingus (2011) showed that while drivers could respond well to a display presenting three vibrational messages (conveyed by different combinations of pairs of four tactors at the seat pan), a seat pan display with seven distinctive vibrational alerts led to erroneous and delayed responses by the driver.

For the driver to perceive vibrotactile warnings, the tactors and the human have to be in physical contact with each other. Recently, BMW and Citroën, among other car manufacturers, introduced a lane departure warning system that alerts the driver by vibrating the steering wheel (BMW, 2013; Spence & Ho, 2008). In highly automated mode, in which the driver is likely to have his hands off the steering wheel, alternative locations to provide vibrations to the driver should be considered, such as the seat back, seat pan, or seat belt (e.g., Schwalk, Kalogerakis, & Maier, 2015; Telpaz, Rhindress, Zelman, & Tsimhoni, 2015).

1.2.4. Multimodal displays

Auditory, visual, and vibrotactile displays can be combined into a multimodal display (e.g., Lee & Spence, 2008; Liu, 2001; Oviatt, 2003; Reeves et al., 2004). Multimodal displays can output more information per quantum of time, resulting in better task performance, compared to unimodal displays (Selcon, Taylor, & McKenna, 1995; Wickens, Hollands, Banbury, & Parasuraman, 2012; and see Burke et al., 2006, for a meta-analysis). For example, De Groot et al. (2013) showed that audio-visual displays were more effective than auditory-only displays for giving turn left/right instructions. A meta-analysis by Prewett, Elliott, Walvoord, and Coovert (2012) showed that task effectiveness (defined as an aggregate of error rate, task completion time, and reaction time effect sizes) was higher for visual-vibrotactile displays than for visual-only displays.

Multimodal displays incorporating all three modalities (i.e., visual, auditory, and vibrotactile) are rare in car driving (Petermeijer et al., 2016). In a series of driving simulator studies, Politis, Brewster, and Pollick (2014, 2015a, 2015b) investigated the effectiveness of tri-modal (visual, auditory, and vibrotactile) displays on driver behaviour and found that these led to increased perceived urgency and perceived alerting effectiveness as compared to unimodal displays. On the other hand, in a study investigating warnings for TORs in Adaptive Cruise Control (ACC), Lee, McGehee, Brown, and Marshall (2006) found that a combination of a visual warning, an auditory warning, a vibratory seat, and pulsation at the brake pedal led to 400 ms slower reaction times than a visual-auditory warning.

The difference in the effects of multimodal warnings in past research can be explained by the fact that the efficacy of a multimodal display depends on whether or not they are semantically, temporally, and spatially congruent, leading to redundancy (Diaconescu, Alain, & McIntosh, 2011; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). It should also be noted that the benefit of multimodal over unimodal displays is not necessarily due to sensory integration (i.e., the beneficial interaction of redundant signals) but could also be caused by one of the modalities substituting another modality that is unavailable or overloaded. As Hancock et al. (2015) noted: "multisensory audio-tactile cuing may be superior to audio cuing alone in noisy environments, but this effect is not necessarily contingent on multisensory integration of auditory with tactile cues, per se" (p. 7).

1.3. Perceived urgency as a function of the display characteristics and the operator's characteristics

Extensive work has been conducted on the relationship between display characteristics and evoked perceived urgency. As Baldwin et al. (2012), Hellier, Edworthy, and Dennis (1993), Hellier and Edworthy (1999), and Park and Jang (1999) showed, Stevens (1957) power law, which associates the objective magnitude of a physical stimulus with its perceived intensity, can be used to predict the perceived urgency of a warning. It has been found that perceived urgency increases with smaller interpulse intervals and that this effect holds for each of the modalities (visual, auditory, & vibrotactile) and combinations thereof (Haas & Casali, 1995; Haas & Edworthy, 1999; Van Erp, Toet, & Janssen, 2015). In Van Erp, Toet, and Janssen, who measured perceived urgency on a scale from 1 (*not urgent*) to 7 (*very urgent*) for 25 combinations of pulse and interpulse interval durations (both ranging between 100 and 1600 ms), perceived urgency increased with decreasing interpulse interval (e.g., for a pulse duration of 100 ms, perceived urgency increased from 2.3 for interpulse intervals of 1600 ms to 6.0 for interpulse intervals of 100 ms). Moreover, for the same interpulse interval, the highest perceived urgency was reported for pulses with a duration equal to the duration of the interpulse interval (i.e., symmetric pulse profiles). In verbal warnings, the semantics of the signal words is an additional parameter that affects perceived urgency, with words such as "deadly" and "danger" being perceived as more urgent than "warning", "caution", or "note" (Hellier et al., 2002).

1.4. Aim of the paper

The aim of the present study was to investigate which unimodal or multimodal display is the most preferred for issuing a TOR during highly automated driving in traffic scenarios of various levels of urgency. Although preference may not coincide with the effectiveness of a display (e.g., Scott & Gray, 2008), the former is important in designing automated driving systems, as dissatisfaction with the display may lead to disuse (Parasuraman & Riley, 1997). It was hypothesized that in low-urgency scenarios drivers opt for unimodal TORs via the traditional visual and/or auditory sensory channels, whereas in high-urgency scenarios a combination of auditory, visual, and vibrotactile TORs is preferred. Additionally, we expected that the shorter the duration of looming sounds and the shorter the intervals between beeps, the more urgent the warning is perceived to be.

We also explored whether TOR preferences correlate with trust in automation and self-reported driving violations. Trust in automation is an important predictor variable in the development of highly automated driving systems, as low levels of trust may lead to disuse of automation, whereas high levels of trust may lead to misuse of automation and compliance (Parasuraman & Riley, 1997). Self-reported driving violations have been included, because they are an important predictor of on-road driving speed (De Winter, Dodou, & Stanton, 2015), which in turn determines the criticality of driving situations and may therefore interact with TOR preferences.

Additionally, we aimed to gain additional insight into the results of Bazilinskyy and De Winter (2015), in which the female voice was preferred over the male voice for supporting TORs during highly automated driving. Bazilinskyy and De Winter included a textual question on this matter. In the current survey, we also produced a synthesized male and female voice and asked people to rate these voices.

Finally, we investigated whether preferences are consistent between countries with different income levels, in light of the view that income is likely to be a strong predictor of road safety (World Health Organization, 2015), and because previous cross-national questionnaire research suggests that respondents from higher-income countries are more likely to be critical towards aspects of automated driving (Bazilinskyy, Kyriakidis, & De Winter, 2015; Kyriakidis, Happee, & De Winter, 2015).

Nowadays, by means of the Internet, researchers can cost-effectively reach a large and diverse pool of participants (Gosling, Vazire, Srivastava, & John, 2004). Moreover, Internet and crowdsourcing platforms allow for the creation of media-rich surveys, containing audio snippets and videos. Media-rich surveys have been shown to be useful for investigating driving attitudes and opinions. For example, Eriksson, Solis Marcos, Kircher, Västfjäll, and Stanton (2015) used an online questionnaire with still images to investigate the type of information people would like to receive during a TOR as a function of the available time and the complexity of the traffic situation. The present survey was conducted with CrowdFlower, which is a platform that allows academic researchers to access an online community to complete a dedicated task, such as filling out a survey (Kyriakidis et al., 2015), categorizing data, or annotating images or videos (Cabrall et al., 2018). Researchers ("Customers" in CrowdFlower terminology) upload the tasks, which are then completed by respondents ("Contributors") in return for a small monetary reward.

2. Methods

2.1. Survey

A survey consisting of 67 questions was developed with CrowdFlower (www.crowdflower.com). Table S1 in the supplementary material shows the questions of the survey as well as the corresponding response options. Earlier surveys indicated that people who work on CrowdFlower surveys have mostly an undergraduate degree (Bazilinskyy & De Winter, 2015; Kyriakidis et al., 2015; see also Behrend, Sharek, Meade, and Wiebe (2011), for similar conclusions regarding Amazon Mechanical Turk). Therefore we used everyday language and avoided technical terms and definitions. A payment of $0.20 (USD) was offered to each respondent for completing the survey. Information about payment appears next to each task when a contributor browses through the list of available tasks. The research was approved by the TU Delft Human Research Ethics Committee.

2.2. Survey structure

At the beginning of the survey, contact information of the researchers was provided, and the purpose of the survey was described as "to explore the public opinion on the use of sound, vibration, and visual interfaces during highly automated car driving". Respondents were informed that the survey would take approximately 15 min of their time. Highly automated driving was then defined as "The automated driving car controls both speed and steering. The driver is not required to look at the road. If the automation cannot handle a situation, it provides a take-over request, and the driver must take over control". The respondents 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 to participate. Information about anonymity and voluntary participation was also given.

The questions were divided into six parts. First, general questions on the respondent's age, gender, age at which the respondent obtained their first license for driving a car or motorcycle, and driving habits and behaviour were posed (Q2–Q15). The questions about driving style (Q9–Q15) were based on the violations scale of the Driver Behaviour Questionnaire (DBQ) as presented by De Winter (2013) and used in previous CrowdFlower surveys (Bazilinskyy & De Winter, 2015; De Winter, Kyriakidis, Dodou, & Happee, 2015).

The second part of the survey (Q16–Q24) focused on general trust in automation, of which Q19–Q24 were previously used by Merritt, Heimbaugh, LaChapell, and Lee (2012) and De Winter and Hancock (2015).

The third part (Q25–Q46) consisted of questions on auditory TORs. Auditory examples were presented in the form of MP3 files stored on an external server. Before answering the questions, the respondents were asked to click on links directing to the recordings and listen to the sound samples. Questions checking whether the respondent actually listened to the samples were incorporated, where the respondents had to select which sound they had just listened to (Q25–Q29). The following auditory TORs were provided:

(1) Female voice "Please take over!".
(2) Male voice "Please take over!". The female and male voices saying "Please take over!" were created with the 'Free online voice generator' (http://onlinetonegenerator.com/voice-generator.html); *Google UK English Female* and *Google UK English Male* voices were used.
(3) Four pairs of beeps with long (2 s) interpulse intervals, 6 pairs of beeps with medium (1 s) interpulse intervals, 8 pairs of beeps with short (750 ms) interpulse intervals, and 11 pairs of beeps with very short (430 ms) interpulse intervals (each pair consisted of two 240-ms long beep tones separated by a 100-ms interpulse interval, with a frequency of 1840 Hz).
(4) Bell sound.
(5) Horn sound.
(6) Looming sounds with a duration of 1 s, 2.5 s, and 5 s. The digital volume of the three looming sounds increased from 0 at the beginning to 0.8 at the end of the sample, with the volume defined on a scale from 0 to 1. Volume increased linearly for the 1 s and 2.5 s samples and quadratically for the 5 s sample. All sound samples were pre-recorded and presented without context (e.g., the increase in volume was a function of time, and was not based on time-to-contact or any other parameter). The frequency of the looming sounds was 440 Hz. The looming sounds were generated with Audacity software.

Additionally, the respondents were asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether a TOR should be provided by means of the female voice (Q30), the male voice (Q31), the beeps (Q32), the bell sound (Q33), and the horn sound (Q34). Next, they were asked which of these five sound messages they considered as the most urgent (Q35) and the most annoying (Q36). Questions Q37 (multiple-choice question) and Q38 (free-response question) asked the respondents to indicate their opinion on why the female voice is often seen as the most preferred type of auditory message to be used in cars. In questions Q39–Q45, the respondents were asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether they considered each of the provided beep sounds and looming sounds as urgent, and in question Q46 the respondents' opinion was polled on whether sounds are a good way to get their attention back to the road, again using a scale from *Disagree strongly* to *Agree strongly*.

The fourth part of the survey (Q47–Q52) focused on visual TORs. Respondents were presented with illustrations of (1) a green icon on the dashboard, (2) a strip of lights at the bottom of the windshield, (3) a head-up display with a green icon, and (4) a brighter dashboard (Fig. 1), and were asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether a take-over request should be provided by means of each of these four visual displays (Q47–Q50). Next, they were asked whether they would like the automation to take over control when they were not looking at the road for over 5 s (Q51), and whether visual messages are a good way to get their attention back to the road (Q52), both questions on a scale from *Disagree strongly* to *Agree strongly*.

The fifth part of the survey (Q53–Q57) posed questions on vibrotactile TORs. Fig. 2 shows the illustrations that were provided to the respondents to indicate four proposed locations of vibrations, namely: (1) the seat back, (2) the seat pan, (3) the seat belt, and (4) the steering wheel. The respondents were again asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether a take-over request should be provided by means of each of these four vibrotactile displays (Q53–Q56). In Question Q57 the respondents' were asked whether vibrations are a good way to get their attention back to the road on a scale from *Disagree strongly* to *Agree strongly*.

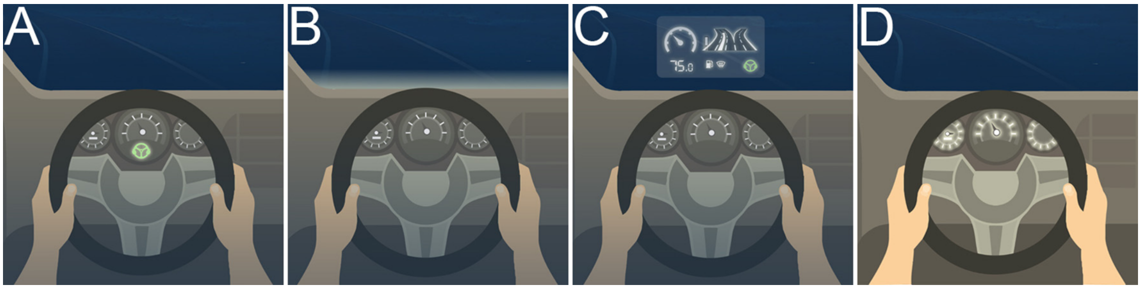

Fig. 1. Illustrations for visual take-over requests. (A) A green icon on the dashboard (Q47); (B) A strip of lights at the bottom of the windshield (Q48); (C) A head-up display with a green icon (Q49); (D) A brighter dashboard (Q50). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

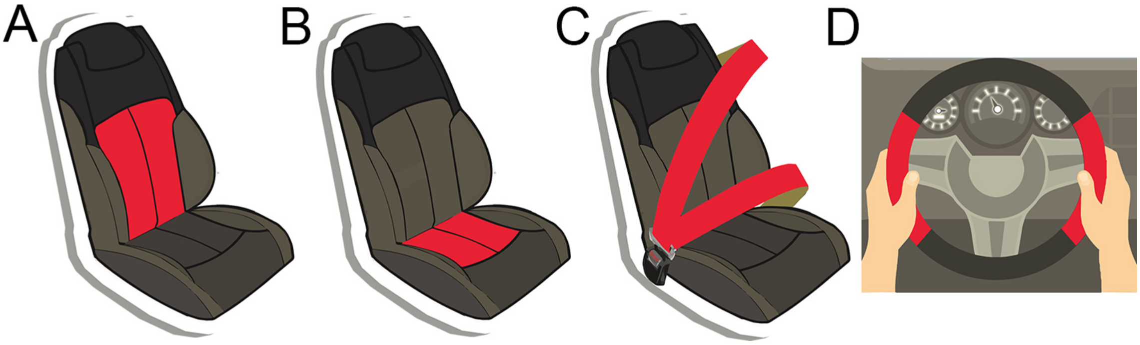

Fig. 2. Illustrations for vibrotactile take-over requests. (A) Vibrations in the seat back (Q53); (2) Vibrations in the seat pan (Q54); (C) Vibrations in the seat belt (Q55); (D) Vibrations in the steering wheel (Q56).

The last part of the survey (Q58–Q65) presented five take-over scenarios of various levels of urgency: (1) construction works, (2) exit highway, (3) changing lanes, (4) automation failure, and (5) traffic accident ahead. For each scenario, the respondents were asked with which display they would like to receive a TOR as well as get informed that the automation was ready to take back control again (Response options: sound message; vibrations; visual message; sound message and vibrations (in any order); sound message and visual message (in any order); visual message and vibrations (in any order); sound message, visual message, and vibrations (in any order)). The scenarios were supplemented with illustrations (Fig. 3). Finally, to poll the public view on the future of transportation, the respondents were asked in which year they think that most cars will be able to drive fully automatically in their country of residence (Q66).

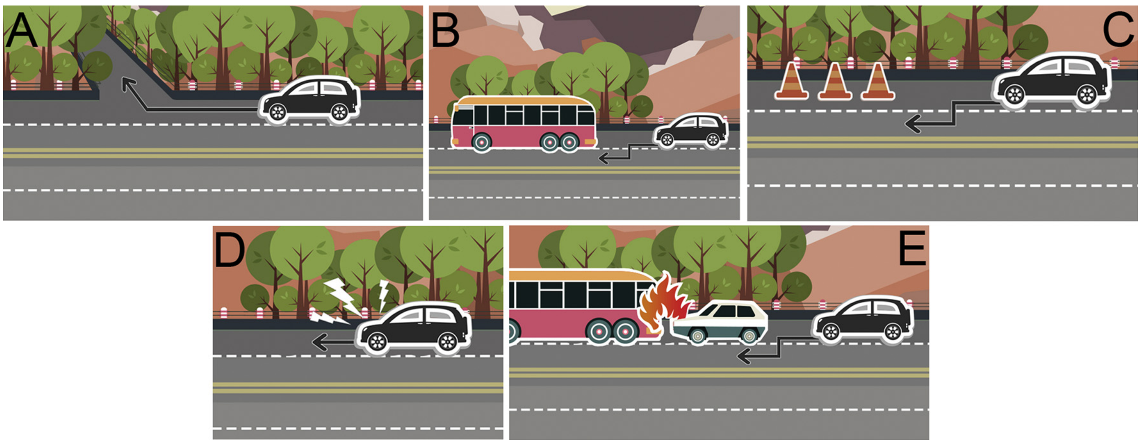

Fig. 3. Illustrations for take-over scenarios, presented in order of increasing urgency. (A) ExitHighway (Q60); (B) ChangeLanes (Q61, Q62); (C) ConstWorks (Q58, Q59); (D) Failure (Q63); (E) Accident (Q64, Q65).

The respondents had to complete all questions, but each question had an "I prefer not to respond" option. The last question (Q67) was the only optional one, asking respondents to "provide any suggestions, which could help engineers to build safe and enjoyable automated cars".

2.3. Configuration of CrowdFlower

In CrowdFlower, the researcher can specify in which countries the target contributors reside. We allowed contributors from all countries to participate in the survey. CrowdFlower contributors are further classified in various performance categories, depending on how trustworthy their former contributions have been. The 'Highest speed–Level 1' contributors were selected. This level of experience accounts for 60% of completed work on CrowdFlower (CrowdFlower, 2015), allowing to reach a diverse group of users of the platform. Completing the survey more than once from the same IP address was not permitted. A payment of $0.20 (USD) was offered to each respondent for completing the survey. We collected 3000 surveys, at a total cost of $798.00. The study was preceded by a pilot test with 10 respondents. These respondents were not included in the analysis. The pilot test did not lead to any changes in the survey.

2.4. Statistical analysis

Descriptive statistics (i.e., mean, median, standard deviation, skewness, and numbers of responses) were calculated for each of the variables. The skewness was calculated as the third central moment divided by the cube of the standard deviation. A Spearman correlation matrix of selected predictor variables (gender, age, driving and accident history, mean of the DBQ violations items Q9–Q15, a trust score calculated as the average of z-transformed responses to Q19–Q24, year in which the respondent thought that most cars would be able to drive fully automatically in his/her country of residence (Q66), and survey completion time) and outcome variables (all ordinal variables related to auditory, visual, and vibrotactile messages) was created at the level of respondents. A distinction was made between respondents located in middle-income countries (gross national income [GNI] per capita: $1026–$12,475) and respondents located in high-income countries (GNI per capita: $12,476 or higher; The World Bank, 2016). We did not perform statistical tests for each possible comparison because the sample size was large, and our concern rests more with practical significance than with statistical significance. As an indication, at our sample size of 1692, a correlation coefficient of 0.05 is statistically significant from zero ($p = 0.04$). For all questions regarding the use of male and female voice, differences between male and female respondents were investigated using Fisher's exact test. 95% confidence intervals around the means per item were calculated assuming a normal distribution, whereas 95% confidence intervals of proportions of respondents were calculated assuming a binomial distribution. At the national level, correlations were calculated between GNI per capita in 2013 (The World Bank, 2015) and the outcome variables, only taking into account those countries with 25 or more respondents (see also Kyriakidis et al., 2015).

3. Results

3.1. Number of respondents and respondent satisfaction

In total, 3000 respondents located in 102 countries were surveyed in this study. The responses were collected between 31 March 2015, 14:01 and 1 April 2015, 4:49 (GMT). The survey received an overall satisfaction rating of 4.4 on a scale from 1 (very dissatisfied) to 5 (very satisfied). The respondents took on average 16.9 min to complete the survey ($SD = 11.1$ min, median = 13.8 min).

3.2. Data filtering

Respondents who indicated they had not read the instructions ($N = 18$) and respondents who indicated they were under 18 and thereby did not adhere to the survey instructions ($N = 5$) were excluded. Additionally, respondents who selected incorrect auditory sounds in Q25–Q29 were excluded ($N_{Q25} = 204$, $N_{Q26} = 281$, $N_{Q27} = 513$, $N_{Q28} = 597$, $N_{Q29} = 508$). We also excluded 25 respondents whose country of residence was not identified by CrowdFlower. Finally, we excluded respondents who chose the "I prefer not to respond" or "I did not hear the recording clearly" options in one or more of the multiple-choice questions, as well as participants who did not report their age or age of obtaining a driving license ($N = 711$; which implies that retained participants all reported to have a driving license). In total, 1308 surveys were removed, leaving 1692 completed surveys for further analysis. For the question "In which year do you think that most cars will be able to drive fully automatically in your country of residence?", 35 of the 1692 answers contained years before 2014 and were excluded from the analysis. Descriptive statistics for all variables are shown in Table S2 of the supplementary material. Table S3 shows histogram counts for all variables.

Note that by the listwise deletion of respondents with missing data (i.e., "I prefer not to respond" or "I did not hear the recording clearly" in one or more questions), the sample size across items becomes constant, and the analysis and presentation of our results are simplified at the expense of some statistical power. The mean responses to the questions were hardly affected by including or excluding these respondents. Specifically, the mean absolute difference of the mean responses across all 41 five-point ordinal items between the used sample ($N = 1692$) and the sample with missing data (Ns for the 41 items between 1991 and 2026) was 0.02.

3.3. Sample characteristics

The mean age of the sample was 32.9 ($SD = 10.4$), and the mean age of obtaining a license for a car or motorcycle was 20.1 years ($SD = 3.8$; $N = 1686$, after excluding 6 participants who reported an unrealistic licensing age below 14 years). Of the 1692 respondents, 1220 respondents were male and 472 were female. 1127 respondents reported that a private vehicle was their primary transportation mode. 632 respondents reported driving every day, and 469 respondents reported driving 4 to 6 days a week. 444 respondents reported that they had driven between 5001 and 15,000 km in the last 12 months, and 329 respondents reported that they had driven between 1001 and 5000 km. The respondents were located in 91 countries. The countries with the largest number of respondents were India ($n = 169$), Spain ($n = 83$), the United States ($n = 82$), Venezuela ($n = 75$), Canada ($n = 71$), and Italy ($n = 68$). Of the 1692 respondents, 565 were located in middle-income countries (GNI per capita: \$1026–\$12,475), and 1127 were located in high-income countries (GNI per capita: \$ 12,476 or higher; The World Bank, 2016). In the middle-income countries, 80% of the respondents were male, whereas the corresponding value was 68% for the high-income countries. In addition to this country-gender association, there was also a strong positive correlation between the mean age of the respondents and the GNI (Spearman $\rho = 0.75$ across the 23 countries with 25 or more respondents).

3.4. Respondents' opinion on auditory, visual, and vibrotactile TORs in highly automated driving

The respondents provided their opinion on whether auditory, visual, and vibrotactile messages are a good way to get their attention back to the road (Q46, Q52, and Q57, respectively), all questions on a five-point Likert scale from *Disagree strongly* to *Agree strongly*. The respondents appeared to be equally positive for all three modalities, with means of 4.06, 3.95, and 4.07 for auditory, visual, and vibrotactile messages, respectively.

Furthermore, the respondents indicated what type of messages should be used for TORs (Q30–Q34 for auditory messages, Q47–Q50 for visual messages, and Q53– Q56 for vibrotactile messages), all questions on a five-point Likert scale from *Disagree strongly* to *Agree strongly*. The results in Fig. 4 show that among visual displays, the head-up display with a green icon received the highest ratings ($M = 3.76$), followed by the green icon on the dashboard ($M = 3.69$). The respondents were less favourable towards the brighter dashboard ($M = 3.06$) and the strip of lights at the bottom of the windshield ($M = 2.84$). The respondents expressed relatively similar acceptance of all four types of vibrotactile messages presented in the survey: vibration of the steering wheel ($M = 3.66$), vibration of the seat back ($M = 3.34$), vibration of the driver's seat pan ($M = 3.19$), and vibration of the seat belt ($M = 3.13$). Large differences were observed among the provided auditory messages, with spoken messages being more accepted than abstract sounds. Specifically, the female and male voices were the most preferred auditory options ($M = 3.60$ and $M = 3.34$, respectively), whereas a horn sound was the least preferred auditory option ($M = 2.27$).

In Q35, the respondents were asked to select which of the proposed auditory messages they considered the most urgent, whereas in Q36 they were asked to report which auditory message they considered the most annoying. Fig. 5 shows the results for both questions. The male voice was considered the most urgent auditory message ($N = 518$; 31% of the male respondents and 30% of the female respondents, $p = 1$), followed by the female voice ($N = 495$; 29% of the male respondents

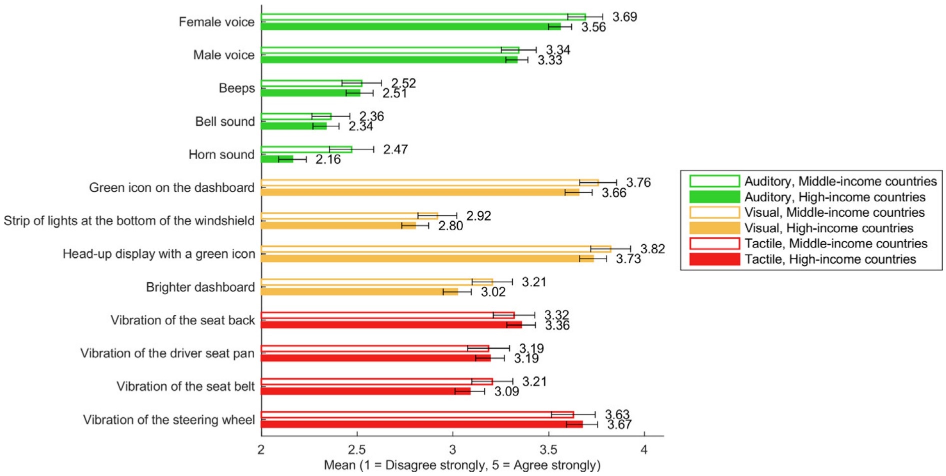

Fig. 4. Respondents' agreement regarding whether a take-over request should be provided by means of specific auditory, visual, or vibrotactile messages (Q30–Q34; Q47–Q50; Q53–Q56). A distinction is made between respondents from middle- and high-income countries ($n = 565$ and 1127, respectively). The number next to each bar is the mean on the scale from 1 to 5. Error bars denote 95% confidence intervals.

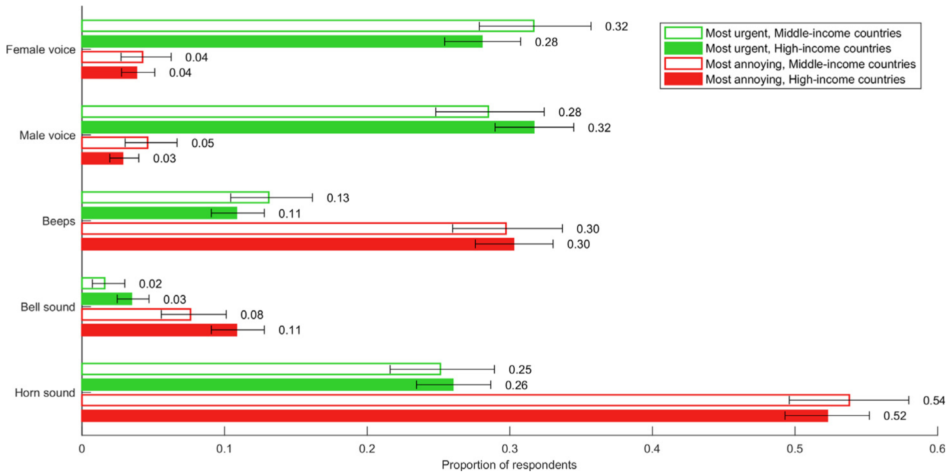

Fig. 5. Proportion of respondents who selected a particular auditory message as the most urgent (Q35) and the most annoying (Q36) after listening to these auditory messages. A distinction is made between respondents from middle- and high-income countries ($n = 565$ and 1127, respectively). Error bars denote 95% confidence intervals.

and 29% of the female respondents, $p = 1$) and the horn sound ($N = 435$). The horn sound was considered the most annoying of the auditory messages ($N = 893$).

In Q39–Q42 the respondents were asked to rate on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether they considered each of the provided beeps as urgent. A monotonic relationship between interpulse interval and perceived urgency was observed, with the beeps with long interpulse intervals seen as the least urgent (Q39; $M = 2.28$) and the beeps with very short interpulse intervals as the most urgent (Q42; $M = 4.20$). Fig. 6 shows the mean perceived urgency rate as a function of the beep rate for the four provided beep messages. A strong linear relationship is observed in logarithmic space ($r = .999$; exponent = 0.66). Among the looming sounds, the sound of 1 s duration was considered the least urgent (Q43; $M = 2.52$), followed by the looming sound of 2.5 s (Q44; $M = 2.92$) and 5 s (Q45; $M = 3.17$).

The survey included a question (Q37) asking respondents to indicate their opinion on why a female voice is often seen as the most preferred type of auditory message to be used in cars. Most participants indicated that a female voice is more pleasant (51%; 854 out of 1686)—significantly more male respondents (55%) than female respondents (40%; $p < 0.001$), followed by less commanding and easier to agree with ($N = 354$)—significantly more female respondents (29%) than male respondents (18%; $p < 0.001$). A female voice was seen as preferred in noisy environments by 285 respondents (16% of the male respondents vs. 18% of the female respondents; $p = 0.278$), and for driving in a car with males by 160 respondents (9% of the male respondents vs. 10% of the female respondents; $p = 0.645$). The respondents were given the opportunity to provide a textual comment, if they chose the option "Other". One hundred fifty-two comments were collected; from these, 79 carried no meaningful information/were unclear, and 3 were not written in English. Nine respondents mentioned that the female voice had acoustic advantages over the male voice (as in being clearer, having higher pitch). Others repeated that the female voice

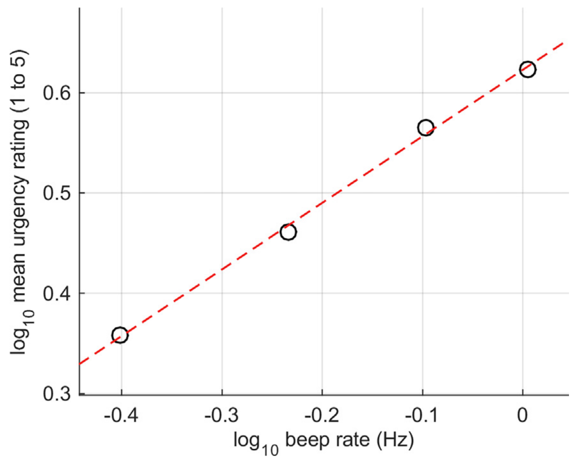

Fig. 6. Mean perceived urgency as a function of the beep rate for the four provided beep messages (Q39–Q42).

is more pleasant ($N = 6$; an option already provided in Q37), calming or relaxing ($N = 4$), comforting, trustworthy, or familiar ($N = 3$), attracting attention ($N = 3$), soft ($N = 2$), authoritative ($N = 1$), believable ($N = 1$), or sensual ($N = 1$). Other explanations included that most drivers are men ($N = 3$), that some drivers consider their cars feminine ($N = 3$), and that women play a protective role in people's life ($N = 2$).

3.5. Respondents' opinion on TORs as a function of the level of urgency

Fig. 7 shows the opinion of the respondents on TORs in the five provided scenarios. A combination of all three types of messages (auditory, visual, and vibrotactile) was the preferred type of TOR for the scenarios of the highest levels of urgency, that is, Accident (Q64; $N = 774$) and Failure (Q63; $N = 687$). An auditory message was selected as the most preferred type of TOR for the scenarios of the lowest levels of urgency, that is, ExitHighway (Q60; $N = 472$) and ChangeLanes (Q61; $N = 386$). A combination of an auditory and a vibrotactile message was the most favourable TOR for the medium-urgency scenario ConstWorks (Q58; $N = 446$).

Fig. 8 shows the respondents' opinion on the types of messages they would like to receive for confirmation that the system is ready to switch back from manual to automated mode for the scenarios ConstWorks (Q59), ChangeLanes (Q62), and Accident (Q65). Auditory messages were the most preferred option for all three scenarios (ConstWorks: $N = 634$; ChangeLanes: $N = 625$; Accident: $N = 604$), followed by a combination of an auditory and a visual message (ConstWorks: $N = 318$; ChangeLanes: $N = 329$; Accident: $N = 346$). There were no distinct differences between the three scenarios (ChangeLanes, ConstWorks, Accident), which is logical, because the confirmation message indicates that the situation was over.

3.6. Correlation analysis

Table S4 in the supplementary material contains a Spearman correlation matrix of selected predictor and outcome variables at the level of respondents. The correlations between the predictor variables and the outcome variables related to auditory (Audio6–10, Beeps1–4, Looming1–3, Audio15), visual (Visual1–6), and vibrotactile (Tactile1–5) TORs were overall small to moderate, ranging between −0.12 and 0.26. The mean of correlations between the trust score (Trust) and the variables that related to auditory, visual, and vibrotactile TORs were small: 0.04, 0.11, and 0.09, respectively. The highest correlation ($\rho = 0.24$) with Trust was found for Visual5, which was polling the people's acceptance of the automation taking over control after 5 s of a driver not looking at the road. The correlations of Trust with Audio15, Visual6, and Tactile5, which represent the opinion of people about whether auditory, visual, and vibrotactile messages, respectively, are a good way to get their attention back to the road, were 0.17, 0.18, and 0.16, respectively. The correlations between the DBQ violations score and the acceptance of the proposed messages were overly small, ranging between −0.06 and 0.06. A correlation of 0.27 was observed between the number of accidents over the last 3 years and the DBQ violations score, in line with the correlations found in previous crowdsourcing surveys ($\rho = 0.28$, $N = 1862$ in De Winter et al., 2015; $\rho = 0.24$, $N = 1205$ in Bazilinskyy & De Winter, 2015).

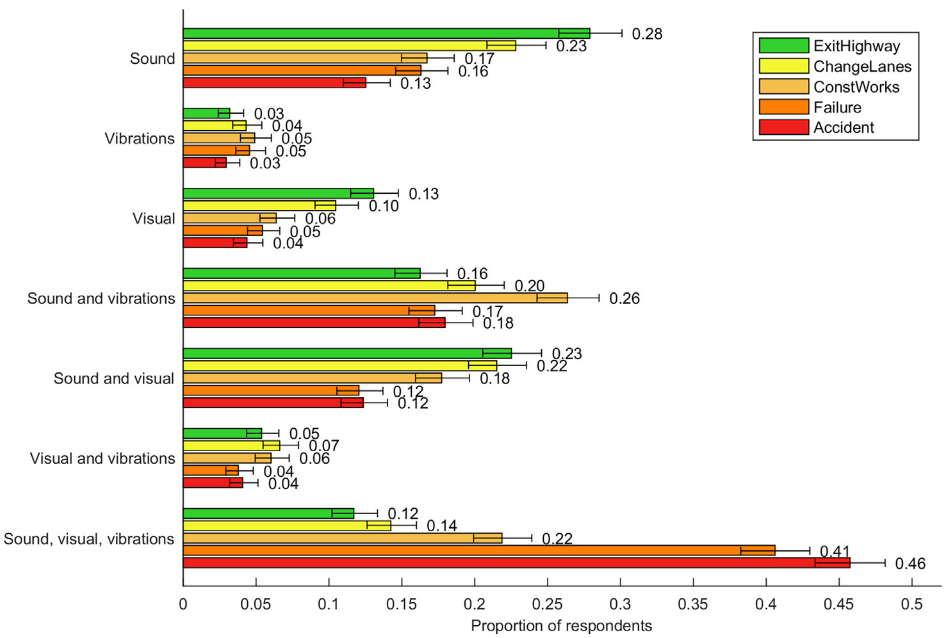

Fig. 7. Proportion of respondents who selected particular take-over requests in the scenarios ExitHighway (Q60), ChangeLanes (Q61), ConstWorks (Q58), Failure (Q63), and Accident (Q64). The scenarios are presented in order of increasing urgency, from low to high. Error bars denote 95% confidence intervals.

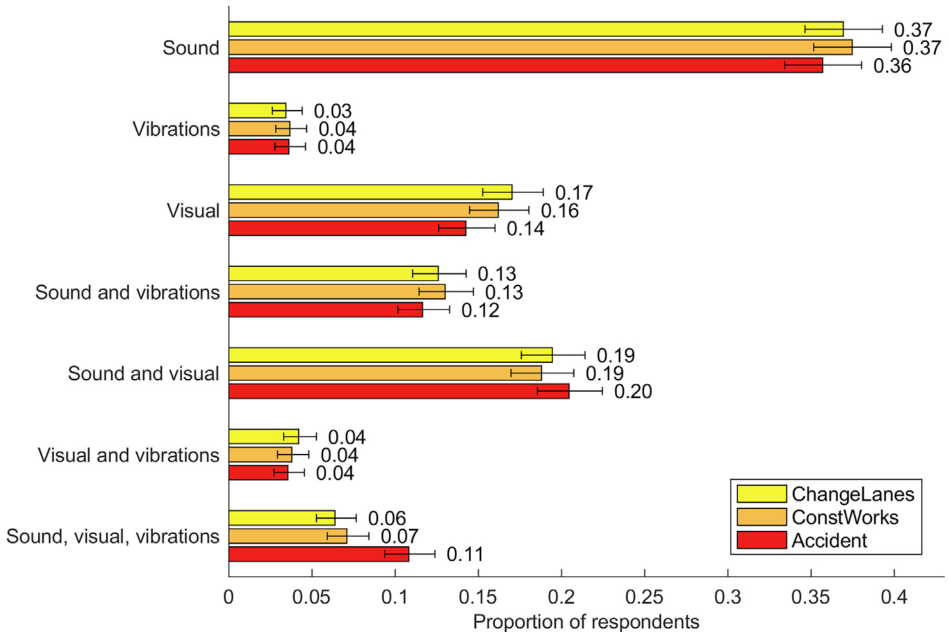

Fig. 8. Proportion of respondents who selected particular messages for confirmation that the system is ready to switch back from manual to automated mode in the scenarios ChangeLanes (Q62), ConstWorks (Q59), and Accident (Q65). Error bars denote 95% confidence intervals.

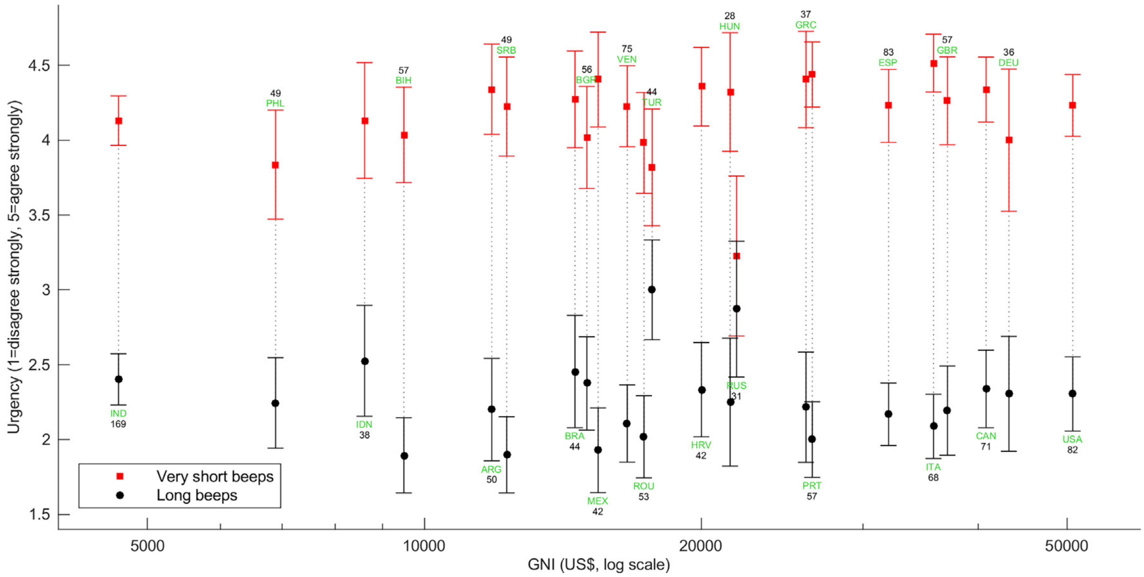

Fig. 9. The reported urgency of long and very short beeps as a function of the gross national income (GNI) per capita. Error bars denote 95% confidence intervals. The numbers below or above each country abbreviation indicate the sample size per country. Country abbreviations are according to ISO 3166-1 alpha-3.

3.7. National comparisons

Figs. 4 and 5 show that the respondents' preferences and ratings of urgency and annoyance were similar in middle- and high-income countries. Horn honking tended to be more preferred in middle- than in high-income countries, which is in line with De Winter and Dodou (2016) where it was found that lower-income countries exhibit more horn honking. Fig. 9 shows the perceived urgency of long and very short beeps per country, as a function of GNI. It can be seen that in all countries long beeps were perceived as less urgent than very short beeps, with no interpretable trends with respect to the countries' GNI.

4. Discussion

In this study 3000 respondents from 102 countries (1692 respondents from 91 countries after filtering) were asked to state their opinion on various types of auditory, visual, and vibrotactile messages for TORs in general and in five scenarios of various levels of urgency during highly automated driving. In three of the scenarios, the respondents were also asked to indicate what type of messages they would prefer to receive to indicate that the system is ready to switch back from manual to automated mode.

4.1. Respondents' opinion on unimodal messages in highly automated driving

The respondents were moderately positive towards all three proposed (auditory, vibrotactile, and visual) types of messages for supporting TORs. Among the provided auditory messages, large differences in preference were observed, with verbal messages being more accepted than abstract sounds. Additionally, the female voice was ranked higher in preference than the male voice, in line with findings from a previous online survey (Bazilinskyy & De Winter, 2015). The respondents indicated their opinion regarding why the female voice may be seen as a preferred modality of feedback in cars. The most selected answer (by 51% of the respondents) was "A female voice is more pleasant". Our results further showed that male and female respondents preferred the female voice for different reasons: males in particular found a female voice pleasant, whereas females in particular found a female voice less commanding and easier to agree with than a male voice.

The difference in the perceived urgency of the male versus female voice in our study was small, with the male and female voices being seen as the most urgent type of auditory warning by 518 and 495 out of the 1692 respondents, respectively (see Q35). Note that no background noise was present in the auditory recordings used in our survey. It has been shown that a male voice is more audible in environments with high noise levels (Nixon et al., 1998). In our survey, in the question asking respondents to indicate their opinion on why the female voice is often seen as the most preferred type of auditory message to be used in cars, 17% ($N = 285$) of the respondents considered that the female voice was easier to understand in a noisy environment.

The respondents perceived the horn signal as the most annoying type of auditory warning, which is in agreement with the survey results by Bazilinskyy and De Winter (2015). We observed a strong linear relationship between the perceived urgency ratings and the beep rate ($r = 0.999$; exponent = 0.66), which verifies previous experimental studies showing that the relationship between perceived urgency and interpulse duration follows Stevens' power law (exponent = 0.61 in Hellier & Edworthy, 1989; between 0.47 and 0.77 in four experiments presented in Gonzalez, Lewis, & Baldwin, 2012; but see also Hellier et al., 1993, for a higher exponent of 1.35). The confirmation of this relationship by means of survey data is an indication that crowdsourcing offers an attractive solution for psychophysics research, which is traditionally carried out in the lab (Stevens & Boring, 1947). Note, however, that generalizations should be made with caution, as the respondents in our survey did not conduct a true magnitude estimation task on a ratio scale, but were asked to report their perceived urgency on a scale from *Disagree strongly* to *Agree strongly*. Further discussion on ratio scales versus ordinal scales in psychophysics research is provided by Stevens and Galanter (1957) and Walker (2002).

The looming warning of the longest duration (5 s) was perceived as the most urgent of the three looming warnings provided. This was unexpected, as we reasoned that a high rate of volume increase (i.e., a looming sound of short duration rapidly building up to a maximum volume) indicates a danger that approaches more rapidly. Note, however, that the looming warnings in our survey lacked context. That is, the respondents were not exposed to visual input from the environment (e.g., time-to-contact or other hazards). Moreover, it is possible that the long duration/slow volume increase in the 5 s looming was experienced as more arousing than the fast looming, in which the volume increase was difficult to perceive because the signal was of short duration (1 s). To shed more light on the effect of looming sounds on the perceived urgency, we recommend driving simulator research in which both (1) the duration of the looming sound and (2) the actual build-up of intervehicular conflict are systematically investigated in a two-factor design.

Among the visual messages provided in the part of the questionnaire investigating the respondents' general opinion on feedback modalities, the head-up display with a green icon and the green icon on the dashboard were the most preferred. When presented within the context of specific TOR scenarios, a combination of auditory and visual messages was moderately accepted for TORs of low and medium urgency (exiting the highway and changing lanes). For low-urgency scenarios, visual-only messages were more often selected than vibration-only messages, which could be explained by the fact that visual information can remain in view for a prolonged amount of time (e.g., as a state indicator), whereas auditory (and vibrotactile) warnings are preferred in cases in which the operator does not need to refer to the message later in time (Stanton & Edworthy, 1999; see also Wickens et al. (2012)).

All four vibrotactile messages received similar ratings in the part of the questionnaire on general opinion (i.e., without offering the context of a specific scenario). Although in highly automated driving the human is not required to keep the hands on the steering wheel, providing vibrations via the steering wheel was rated as most preferred. It has to be noted that the survey did not mention whether the driver could take the hands off the steering wheel. Respondents did not prefer vibrotactile warnings as a single modality or in combination with visual warnings for any of the five TOR scenarios.

The correlations between the trust score and the acceptance of the proposed types of messages were small to moderate. The respondents with high trust scores indicated that they would prefer the automation to be able to take over control after

5 s of a driver not looking at the road ($\rho = 0.24$), which is an expected result, because adaptive automation ("automation design where tasks are dynamically allocated between the human operator and computer", Byrne & Parasuraman, 1996, p. 249) requires trust in automation (Parasuraman, Sheridan, & Wickens, 2008).

The preferences of the respondents for the use of messages for supporting take-over requests and the perception of auditory messages as the most urgent and the most annoying were similar in middle- and high-income countries, indicating that our findings are generalizable across different countries. These findings are in line with Klein et al. (2014), who found that effect sizes in typical psychological experiments are about the same regardless of whether the experiment had been executed in the United States or not.

4.2. Respondents' opinion on TORs as a function of the level of urgency

Five scenarios, with different levels of urgency, were presented in our survey (highway exit, changing lane, construction works, automation failure, and traffic accident ahead). For each of the five scenarios, the respondents were asked to state the modalities that they preferred for TORs, whereas for three of the scenarios (i.e., construction works, changing lane, and traffic accident), the respondents were also asked to choose their preferred modality for receiving a confirmation message that the system is ready to switch back from manual to automated mode. The combination of auditory, visual, and vibrotactile messages was the most preferred option for supporting TORs in high-urgency scenarios. This survey adds to the existing knowledge by showing that people have a preference for multimodal warnings when the situation becomes more critical, in line with the theories on the benefits of multimodal over unimodal displays described in the introduction (see also Haas & Casali, 1995; Selcon et al., 1995; Petermeijer, Abbink, Mulder, & De Winter, 2015). Future behavioural research should investigate whether tri-modal TORs are indeed more effective than audio-visual ones.

An auditory message was selected as the most preferred option for low-urgency scenarios. Also for confirmation that the system is ready to switch back from manual to automated mode, were auditory messages the most preferred option for all three provided scenarios, followed by a combination of auditory and visual messages. These results indicate that the respondents wanted different types of messages for urgent TORs and for low-urgency TORs/confirmation messages: multimodal displays were preferred for urgent TORs, whereas traditional audio-visual displays were preferred for warnings in low-urgency situations and as state indicators.

4.3. Limitations and future work

We applied a strict screening by excluding all participants who failed one or more test questions. We also excluded participants who skipped a question or who did not report having a driver's license. Accordingly, we expect that our data are of good quality and that our findings are robust. For example, we expect that the observation that mean perceived urgency increases as a function of the beep rate is replicable and generalizable. However, other findings may be contingent on context, as participants had to imagine concepts of displays in a highly automated vehicle based on textual, visual, and auditory descriptions. The results could be different if participants were physically interacting with such displays. This is particularly true for vibrotactile displays, which are not often used in the automotive domain (Meng & Spence, 2015), meaning that it is likely that the respondents had not experienced vibrotactile displays in the context of driving before (note, however, that humans nowadays are familiar with receiving vibrotactile notifications from their mobile devices).

A related limitation is that the survey merely contained descriptions of TOR scenarios and that the respondents had to imagine being distracted before judging which of the proposed messages would be the most preferred to draw their attention back to the driving scene. For example, participants had to imagine that visual warnings could be ineffective if the driver is engaged in a visually demanding non-driving task such as reading. Behavioural research in simulators or on actual roads should be conducted to acquire knowledge on the effectiveness of the warnings in scenarios that involve visual distractions or competing stimuli, such as a visually demanding outside environment, task-intrinsic vibrations in the driver seat (e.g., 'road rumble'), and sounds from into vehicle devices and other road users (e.g., emergency vehicles). It is possible to extend the present crowdsourcing research towards an interactive environment in which participants are shown videos or animations of traffic scenes and have to respond to actual sounds and visual warnings. Although the present survey was of low perceptual fidelity (i.e., participants did not actually experience the interfaces in a driving context), it offered high scope and statistical power, with a large number of participants worldwide being surveyed with identical instructions, questions, sounds, and images. However, the participants' hardware and volume settings were not under our control.

Furthermore, we note that the TORs which we investigated are only a selection of all possible design solutions. For example, only one male voice and one female voice were provided, which means that results may be contingent on the voice generator, accent, pitch, etc. (for a similar discussion on TORs see Bazilinskyy & De Winter, 2017, in which participants rated a number of male and female voices). Also the five traffic situations are merely a selection of possible scenarios in actual automated driving. For example, transitions of control may be mandatory in case the driver is unable to take over (Lu, Happee, Cabrall, Kyriakidis, & De Winter, 2016), and there may be situations where the car has to bring itself to a minimum risk condition (Hoeger et al., 2008; Gasser & Westhoff, 2012).

A final point of attention is whether CrowdFlower respondents are representative of the (expectedly high-income) future stakeholders of highly automated driving technology. The representativeness of crowdsourced samples is the topic of ongoing research, and it has become clear that the active CrowdFlower and Amazon Mechanical Turk populations are relatively small (<10,000 persons), consisting of people who have evolved into specialized research participants and who may spend a large share of their time behind the computer (Chandler, Mueller, & Paolacci, 2014; Stewart et al., 2015). Of the 1692 respondents, 67% were located in high-income countries and the remaining 33% were located in middle-income countries. Hence, a large number of the participants were from countries where highly automated vehicles will likely be initially introduced to public roads, and where warning and assistance systems such as Adaptive Cruise Control, Lane Keeping Assist, and parking sensors, are already common (Bishop, 2005; Bazilinskyy & De Winter, 2015). However, a previous study showed that only 8% of the CrowdFlower participants reported a gross annual income of EUR 44,000 or more (Kyriakidis et al., 2015). Thus, it is unlikely that the respondents themselves represent the typical early adopters of automated driving technology.

4.4. Conclusions

In conclusion, this study allowed us to access a large and diverse population to gain a first impression about display design for highly automated driving. The survey showed that people's preferences for the type of messages they would like to receive in highly automated driving depend on the urgency of the situation ahead and on whether the message concerns a TOR or a confirmation that the system is ready to switch back from manual to automated mode. Specifically: (1) For high-urgency situations, multimodal warnings were the most preferred option, (2) For low-urgency situations and for receiving confirmation that the system is ready to switch back from manual to automated mode, auditory messages were the most preferred option, (3) For low-urgency scenarios, visual-only TORs were more preferred than vibration-only TORs, (4) Among messages represented by beeps, beeps with shorter interpulse intervals were perceived as more urgent, consistent with Stevens' power law, (5) Among the five provided sounds, spoken messages were more accepted than abstract sounds, and the female voice was more preferred than the male voice.

With a large sample of 1692 respondents, our survey validates previous experimental findings and theoretical considerations with respect to preferred modalities for issuing TORs during highly automated driving as a function of the urgency level of the TOR. The survey may be launched again in the future among high-income users who are likely to purchase automated cars. The present results are of preliminary and abstract nature, and do not immediately allow for specific design recommendations concerning TORs in future automated vehicles. Simulator-based and on-road studies need to be performed, and designers need to be involved, before being able to transfer the results to TORs in real vehicles.

Supplementary material

Supplementary material may be found at https://doi.org/10.4121/uuid:e3908ec5-d086-4737-8d4a-d4046dbbc53c.

Acknowledgments

We would like to express our special gratitude to Daria Nikulina for designing the illustrations used in the survey. The research presented in this paper is being conducted in the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.trf.2018.04.001.

References

Arrabito, G. R. (2009). Effects of talker sex and voice style of verbal cockpit warnings on performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 51*, 3–20. https://doi.org/10.1177/0018720808333411.

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., ... Seifritz, E. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18*, 145–150. https://doi.org/10.1093/cercor/bhm040.

Baldwin, C. L., Eisert, J. L., Garcia, A., Lewis, B., Pratt, S. M., & Gonzalez, C. (2012). Multimodal urgency coding: auditory, visual, and tactile parameters and their impact on perceived urgency. *Work-Journal of Prevention Assessment and Rehabilitation, 41*, 3586–3591. https://doi.org/10.3233/WOR-2012-0669-3586.

Bate, A. J. (1969). *Cockpit warning systems comparative study* (Final Report AMRL-TR, 68–193). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command.

Bazilinskyy, P., & De Winter, J. C. F. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science, 3*, e1520. https://doi.org/10.7717/peerj-cs.13.

Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. *Applied Ergonomics, 64*, 56–64. https://doi.org/10.1016/j.apergo.2017.05.001.

Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2015). An international crowdsourcing study into people's statements on fully automated driving. In *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)* (pp. 2534–2542). Las Vegas, NV. https://doi.org/10.1016/j.promfg.2015.07.540.

Begg, D. (2014). A 2050 vision for London: What are the implications of driverless transport. Retrieved from https://trid.trb.org/view/1319762.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*, 800–813. https://doi.org/10.3758/s13428-011-0081-0.

Birrell, S. A., Young, M. S., & Weldon, A. M. (2013). Vibrotactile pedals: provision of haptic feedback to support economical driving. *Ergonomics, 56*, 282–292. https://doi.org/10.1080/00140139.2012.760750.

Bishop, R. (2005). Intelligent vehicle R&D: A review and contrast of programs worldwide and emerging trends. *Annales Des Télécommunications, 60*, 228–263. https://doi.org/10.1007/BF03219820.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4*, 11–44.

BMW. (2013). The BMW 3 Series Sedan. Owner's manual. Retrieved from https://carmanuals2.com/get/bmw-335i-2013-owner-s-manual-55036.

Burke, J. L., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coovert, M. D., ... & Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. In *Proceedings of the 8th International Conference on Multimodal Interfaces* (pp. 108–117). New York: ACM. https://doi.org/10.1145/1180995.1181017.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*, 249–268. https://doi.org/10.1016/0301-0511(95)05161-9.

Cabrall, C. D. D., Lu, Z., Kyriakidis, M., Manca, L., Dijksterhuis, C., Happee, R., & De Winter, J. C. F. (2018). Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing. *Accid. Anal. Prev., 114*, 25–33. https://doi.org/10.1016/j.aap.2017.08.036.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*, 112–130. https://doi.org/10.3758/s13428-013-0365-7.

CrowdFlower. (2015). Job settings|Guide to: Basic job settings page. Retrieved from https://success.crowdflower.com/hc/en-us/articles/201855719-Job-Settings-Guide-to-Basic-Job-Settings-Page.

De Groot, S., De Winter, J. C. F., García, J. M. L., Mulder, M., & Wieringa, P. A. (2011). The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*, 50–62. https://doi.org/10.1177/0018720810393241.

De Groot, S., De Winter, J. C. F., Mulder, M., Kuipers, J., Mulder, J. A., & Wieringa, P. A. (2013). The effects of route-instruction modality on driving performance in a simulator. In S. de Groot (Ed.), *Exploiting the possibilities of simulators for driver training* [Doctoral dissertation]. The Netherlands: Delft University of Technology.

De Winter, J. C. F. (2013). Predicting self-reported violations among novice license drivers using pre-license simulator measures. *Accident Analysis and Prevention, 52*, 71–79. https://doi.org/10.1016/j.aap.2012.12.018.

De Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? In *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)* (pp. 5334–5341). Las Vegas, NV. https://doi.org/10.1016/j.promfg.2015.07.641.

De Winter, J. C. F., & Dodou, D. (2016). National correlates of self-reported traffic violations across 41 countries. *Personality and Individual Differences, 98*, 145–152. https://doi.org/10.1016/j.paid.2016.03.091.

De Winter, J. C. F., Dodou, D., & Stanton, N. A. (2015). A quarter of a century of the DBQ: Some supplementary notes on its validity with regard to accidents. *Ergonomics, 58*, 1745–1769. https://doi.org/10.1080/00140139.2015.1030460.

De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. In *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)* (pp. 2518–2525). Las Vegas, NV. https://doi.org/10.1016/j.promfg.2015.07.514.

Dettmann, A., & Bullinger, A. C. (2017). Spatially distributed visual, auditory and multimodal warning signals–A comparison. In D. de Waard, A. Toffetti, R. Wiczorek, A. Sonderegger, S. Röttger, P. Bouchner, T. Franke, S. Fairclough, M. Noordzij, & K. Brookhuis (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference* (pp. 185–199).

Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology, 106*, 2896–2909. https://doi.org/10.1152/jn.00303.2011.

Edworthy, J., & Hellier, E. J. (2003). Urgency in speech warnings. In R. G. de Jong, T. Houtgast, E. A. M. Fransten, & W. F. Hofman (Eds.), *Proceedings of the 8th International Congress on Noise as a Public Health Problem* (pp. 65–69). Rotterdam, the Netherlands: Foundation ICBEN.

Eriksson, A., Solis Marcos, I., Kircher, K., Västfjäll, D., & Stanton, N. A. (2015). The development of a method to assess the effects of traffic situation and time pressure on driver information preferences. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (pp. 3–12). Springer International Publishing.

Fitch, G. M., Hankey, J. M., Kleiner, B. M., & Dingus, T. A. (2011). Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*, 278–290. https://doi.org/10.1016/j.trf.2011.02.001.

Fitts, P. M., & Jones, R. E. (1947). *Psychological aspects of instrument display: 1. Analysis of 270 "pilot-error" experiences in reading and interpreting aircraft instruments* (Memorandum Report TSEAA-694-12A). Wright-Patterson Air Force Base, OH: Air Materiel Command.

Flemisch, F., Kaussner, A., Petermann, I., Schieben, A., & Schöming, N. (2011). *HAVE-IT. Highly automated vehicles for intelligent transport. Validation of concept on optimum task repartition* (Deliverable D.33.6). Regensburg, Germany: Continental Automotive GmbH.

Flemisch, F. O., Kelsch, J., Schieben, A., & Schindler, J. (2006). Stücke des Puzzleshochautomatisiertes Fahren: H-Metapher und H-Mode. In C. Stiller & M. Maurer (Eds.), *Workshop ahrerassistenzsysteme*. Karlsruhe.

Gasser, T., & Westhoff, D. (2012). *Bast-study: Definitions of automation and legal issues in germany*. Irvine, CA, USA: TRB Road Vehicle Automation Workshop.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction, 2*, 167–177.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*, 1938–1942. https://doi.org/10.1177/1541931213571433.

Gold, C., Lorenz, L., Damböck, D., & Bengler, K. (2013). Partially automated driving as a fallback level of high automation. 6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren. TÜV SÜD Akademie GmbH.

Gonzalez, C., Lewis, B. A., & Baldwin, C. L. (2012). Revisiting pulse rate, frequency and perceived urgency: have relationships changed and why? In *Proceedings of the 18th International Conference on Auditory Display* (pp. 44–51). Atlanta, GA.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93–104. https://doi.org/10.1037/0003-066X.59.2.93.

Grah, T., Epp, F., Wuchse, M., Meschtscherjakov, A., Gabler, F., Steinmetz, A., & Tscheligi, M. (2015). Dorsal haptic display: a shape-changing car seat for sensory augmentation of rear obstacles. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 305–312). ACM. https://doi.org/10.1145/2799250.2799281.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics, 42*, 1233–1248. https://doi.org/10.1080/001401399185108.

Gray, R. (2011). Looming auditory collision warnings for driving. *Human Factors, 53*, 63–74. https://doi.org/10.1177/0018720810397833.

Green, P. (1999). *Visual and task demands of driver information systems* (Technical Report UMTRI-98-16). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Grether, W. F. (1949). Instrument reading. I. The design of long-scale indicators for speed and accuracy of quantitative readings. *Journal of Applied Psychology, 33*, 363–372. https://doi.org/10.1037/h0058374.

Haas, E. C., & Casali, J. G. (1995). Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics, 38*, 2313–2326. https://doi.org/10.1080/00140139508925270.

Haas, E. C., & Edworthy, J. (1999). The perceived urgency and detection time of multitone auditory signals. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 129–149). Aldershot: Ashgate Publishing Limited.

Hancock, P. A., Lawson, B., Cholewiak, R., Elliott, L. R., Van Erp, J. B., Mortimer, B. J., ... & Redden, E. S. (2015). Tactile cuing to augment multisensory human-machine interaction. *Ergonomics in Design: The Quarterly of Human Factors Applications, 23*, 4–9. https://doi.org/10.1177/1064804615572623.

Hellier, E., & Edworthy, J. (1989). Quantifying the perceived urgency of auditory warnings. *Canadian Acoustics, 17*, 3–11.
Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics, 30*, 167–171. https://doi.org/10.1016/S0003-6870(97)00013-6.
Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*, 693–706. https://doi.org/10.1177/001872089303500408.
Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*, 1–17. https://doi.org/10.1518/0018720024494810.
Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemisch, F., Krueger, H. P., ... & Pagle, K. (2008). Highly automated vehicles for intelligent transport: HAVEit approach. In Proceedings of the 15th World Congress on Intelligent Transport Systems. New York, NY.
Hoeger, R., Zeng, H., Hoess, A., Kranz, T., Boverie, S., Strauss, M., et al. (2011). *The future of driving – HAVEit* (Final Report, Deliverable D61.1). Regensburg, Germany: Continental Automotive GmbH.
Honda. (2014). 2014 ACCORD 4D online reference owner's manual. Retrieved from https://carmanuals2.com/get/honda-accord-2014-owner-s-manual-80195.
Jang, P. S. (2007). Designing acoustic and non-acoustic parameters of synthesized speech warnings to control perceived urgency. *International Journal of Industrial Ergonomics, 37*, 213–223. https://doi.org/10.1016/j.ergon.2006.10.018.
Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 90–111. https://doi.org/10.1518/001872008X250638.
Kelsch, J., & Dziennus, M. (2015). Joint driver-automation system design: Gradual action-oriented ambient stimuli. *Presented at the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*. Las Vegas, NV.
Kim, N., Jeong, K., Yang, M., Oh, Y., & Kim, J. (2017). Are you ready to take-over? An exploratory study on visual assistance to enhance driver vigilance. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1771–1778). New York, NY: ACM. https://doi.org/10.1145/3027063.3053155.
Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr, Bahník, Š., Bernstein, M. J., ... Cemalcilar, Z. (2014). Investigating variation in replicability. *Social Psychology, 45*, 142–152. https://doi.org/10.1027/1864-9335/a000178.
Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014.
Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics, 42*, 807–815. https://doi.org/10.1080/001401399185306.
Langlois, S., & Soualmi, B. (2016). Augmented reality versus classical HUD to take over from automated driving: An aid to smooth reactions and to anticipate maneuvers. In *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, (pp. 1571–1578). IEEE. https://doi.org/10.1109/ITSC.2016.7795767.
Large, D. R., & Burnett, G. E. (2013). Drivers' preferences and emotional responses to satellite navigation voices. *International Journal of Vehicle Noise and Vibration, 9*, 28–46. https://doi.org/10.1504/IJVNV.2013.053815.
Lee, J. H., & Spence, C. (2008). Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1* (pp. 185–192). British Computer Society.
Lee, J., McGehee, D., Brown, T., & Marshall, D. (2006). Effects of adaptive cruise control and alert modality on driver performance. *Transportation Research Record: Journal of the Transportation Research Board, 1980*, 49–56. https://doi.org/10.3141/1980-09.
Liu, Y.-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics, 44*, 425–442. https://doi.org/10.1080/00140130010011369.
Löcken, A., Heuten, W., & Boll, S. (2015). Supporting lane change decisions with ambient light. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 204–211). New York, NY: ACM. https://doi.org/10.1145/2799250.2799259.
Lu, Z., Happee, R., Cabrall, C. D. D., Kyriakidis, M., & De Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 183–198. https://doi.org/10.1016/j.trf.2016.10.007.
Lu, S. A., Wickens, C. D., Sarter, N. B., & Sebok, A. (2011). Informing the design of multimodal displays: a meta-analysis of empirical studies comparing auditory and tactile interruptions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*, 1170–1174. https://doi.org/10.1177/1071181311551244.
Manca, L., De Winter, J. C. F., & Happee, R. (2015). Visual displays for automated driving: A survey. In *Paper presented at the Workshop on Adaptive Ambient In-Vehicle Displays and Interactions*. Nottingham, UK.
Meng, F., Ho, C., Gray, R., & Spence, C. (2015). Dynamic vibrotactile warning signals for frontal collision avoidance: Towards the torso versus towards the head. *Ergonomics, 58*, 411–425. https://doi.org/10.1080/00140139.2014.976278.
Meng, F., & Spence, C. (2015). Tactile warning signals for in-vehicle systems. *Accident Analysis and Prevention, 75*, 333–346. https://doi.org/10.1016/j.aap.2014.12.013.
Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2012). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*, 520–534. https://doi.org/10.1177/0018720812465081.
Meschtscherjakov, A., Döttlinger, C., Rödel, C., & Tscheligi, M. (2015). ChaseLight: ambient LED stripes to control driving speed. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 212–219). New York, NY: ACM. https://doi.org/10.1145/2799250.2799279.
Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: unstructured transition timing for distracted drivers of automated vehicles. In *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Canary Islands, Spain. https://doi.org/10.1109/ITSC.2015.396.
Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. In N. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in Human Aspects of Transportation: Part I* (pp. 431–438). AHFE Conference.
Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2016). *Speech improves human-automation cooperation in automated driving*. Aachen: Workshop Automotive HMI. https://doi.org/10.18420/muc2016-ws08-0007.
Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of Human Factors and Ergonomics, 7*, 58–99. https://doi.org/10.1177/1557234X11410396.
Nixon, C. W., Morris, L. J., McCavitt, A. R., McKinley, R. L., Anderson, T. R., McDaniel, M. P., et al (1998). Female voice communications in high levels of aircraft cockpit noises–part I: Spectra, levels, and microphones. *Aviation, Space, and Environmental Medicine, 69*, 675–683.
Noyes, J. M., Hellier, E., & Edworthy, J. (2006). Speech warnings: A review. *Theoretical Issues in Ergonomics Science, 7*, 551–571. https://doi.org/10.1080/14639220600731123.
Oviatt, S. (2003). Multimodal interfaces. In J. Jacko & A. Sears (Eds.), *Handbook of human-computer interaction* (pp. 286–304). Hillsdale, NJ: Erlbaum.
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*, 230–253. https://doi.org/10.1518/001872097778543886.
Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*, 140–160. https://doi.org/10.1518/155534308X284417.
Park, K. S., & Jang, P. S. (1999). Effects of synthesized voice warning parameters on perceived urgency. *International Journal of Occupational Safety and Ergonomics, 5*, 73–95. https://doi.org/10.1080/10803548.1999.11076412.
Patterson, R. D. (1982). *Guidelines for auditory warning systems on civil aircraft (CAA paper 82017)*. London: Civil Aviation Authority.

Patterson, R. D., & Mayfield, T. F. (1990). Auditory warning sounds in the work environment [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 327*, 485–492. https://doi.org/10.1098/rstb.1990.0091.
Petermeijer, S. M., Abbink, D. A., Mulder, M., & De Winter, J. C. F. (2015). The effect of haptic support systems on driver performance: a literature survey. *IEEE Transactions on Haptics, 8*, 467–479. https://doi.org/10.1109/TOH.2015.2437871.
Petermeijer, S. M., De Winter, J. C. F., & Bengler, K. J. (2016). Vibrotactile displays: a survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transporation Systems, 17*, 897–907. https://doi.org/10.1109/TITS.2015.2494873.
Pfromm, M., Cieler, S., & Bruder, R. (2013). Driver assistance via optical information with spatial reference. In *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems* (pp. 2006–2011). The Hague, the Netherlands. https://doi.org/10.1109/ITSC.2013.6728524.
Politis, I., Brewster, S. A., & Pollick, F. (2014). Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 4067–4076). New York, NY: ACM. https://doi.org/10.1145/2556288.2556988.
Politis, I., Brewster, S., & Pollick, F. (2015a). To beep or not to beep?: Comparing abstract versus language-based multimodal driver displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3971–3980). New York, NY: ACM. https://doi.org/10.1145/2702123.2702167.
Politis, I., Brewster, S., & Pollick, F. (2015b). Language-based multimodal displays for the handover of control in autonomous cars. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 3–10). New York: ACM. https://doi.org/10.1145/2799250.2799262.
Prewett, M. S., Elliott, L. R., Walvoord, A. G., & Coovert, M. D. (2012). A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42*, 123–132, https://doi.org10.1109/TSMCC.2010.2103057.
Ramkissoon, R. N. (2001). *The use of auditory cues to provide dynamic alerts to drivers* [Doctoral dissertation]. Massachusetts Institute of Technology.
Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., ... Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM, 47*, 57–59. https://doi.org/10.1145/962081.962106.
Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design*. New York: McGraw-Hill.
Schwalk, M., Kalogerakis, N., & Maier, T. (2015). Driver support by a vibrotactile seat matrix – Recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving. In *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics* (AHFE) (pp. 1427–1434). Las Vegas, NV. https://doi.org/10.1016/j.promfg.2015.07.507.
Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors, 50*, 264–275. https://doi.org/10.1518/001872008X250674.
Selcon, S. J., Taylor, R. M., & McKenna, F. P. (1995). Integrating multiple information sources: Using redundancy in the design of warnings. *Ergonomics, 38*, 2362–2370. https://doi.org/10.1080/00140139508925273.
Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time to contact. *Ecological Psychology, 3*, 253–261. https://doi.org/10.1207/s15326969eco0303_4.
Silva, R. M., Lamas, J., Silva, C. C., Coello, Y., Mouta, S., & Santos, J. A. (2017). Judging time-to-passage of looming sounds: Evidence for the use of distance-based information. *PLoS ONE, 12*, e0177734. https://doi.org/10.1371/journal.pone.0177734.
Sivak, M., & Owens, D. A. (1996). The information that drivers use: is it indeed 90% visual? *Perception, 25*, 1081–1090. https://doi.org/10.1068/p251081.
Society of Automotive Engineers [SAE] (2014). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (Standard No. J3016)*. SAE International. Retrieved from http://standards.sae.org/j3016_201401/.
Spence, C., & Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics, 1*, 121–129. https://doi.org/10.1109/TOH.2008.14.
Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and displays: An overview. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 3–30). Aldershot: Ashgate Publishing Limited.
Stevens, S. S. (1957). On the psychophysical law. *The Psychological Review, 64*, 153–181. https://doi.org/10.1037/h0046162.
Stevens, S. S., & Boring, E. G. (1947). The new Harvard psychological laboratories. *American Psychologist, 2*, 239–243. https://doi.org/10.1037/h0060913.
Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54*, 377–411. https://doi.org/10.1037/h0043680.
Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., et al (2015). The average laboratory samples a population of 7300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10*, 479–491.
Stokes, A., Wickens, C., & Kite, K. (1990). *Display technology-human factors concepts* (Technical Report SAE R-102). Warrendale, PA: Society of Automotive Engineers Inc.
Swain, A. D., & Guttmann, H. E. (1983). *Handbook of human reliability analysis with emphasis on nuclear power plant applications* (Final Report NUREG/CR-1278). Albuquerque, NM: Sandia National Laboratories.
Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*, 400–410. https://doi.org/10.1016/j.tics.2010.06.008.
Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic seat for automated driving: preparing the driver to take control effectively. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 23–30). New York, NY: ACM. https://doi.org/10.1145/2799250.2799267.
The World Bank. (2015). Indicators [data]. Retrieved from http://data.worldbank.org/indicator/.
The World Bank. (2016). Country and lending groups. Retrieved from http://data.worldbank.org/about/country-and-lending-groups.
Toyota (2014). *4RUNNER 2014*. Retrieved from: Owner's manual,. https://carmanuals2.com/get/toyota-4runner-2014-owner-s-manual-89510.
Underwood, S. E. (2014). Automated vehicles forecast vehicle symposium opinion survey. *Paper presented at the Automated Vehicles Symposium 2014*. San Francisco, CA.
Van Erp, J. B. F., Toet, A., & Janssen, J. B. (2015). Uni-, bi-and tri-modal warning signals: Effects of temporal parameters and sensory modality on perceived urgency. *Safety Science, 72*, 1–8. https://doi.org/10.1016/j.ssci.2014.07.022.
Volkswagen. (2014). Park assist. Retrieved from http://www.volkswagen.co.uk/technology/parking-and-manoeuvring/park-assist.
Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied, 8*, 211–221. https://doi.org/10.1037/1076-898X.8.4.211.
Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2012). *Engineering psychology and human performance. 4th ed.*. Psychology Press.
Wickens, C. D., & Seppelt, B. (2002). *Interference with driving or in-vehicle task information: The effects of auditory versus visual delivery (Technical Report AHFD-02-18/GM-02-3)*. Urbana-Champaign: University of Illinois at Urbana-Champaign, Aviation Human Factors Division.
Winkler, S., Powelleit, M., Kazazi, J., Vollrath, M., Krautter, W., Korthauer, A., ... & Bendewald, L. (2018). HMI strategy–warnings and interventions. In: K. Bengler, J. Drüke, S. Hoffmann, D. Manstetten, & A. Neukum (Eds.), *UR:BAN Human factors in traffic. Approaches for safe, efficient and stress-free urban traffic* (pp. 75–103). Wiesbaden: Springer Vieweg. https://doi.org/10.1007/978-3-658-15418-9_5.
World Health Organization (2015). *WHO global status report on road safety 2015*. WHO.
You, F., Wang, Y., Wang, J., Zhu, X., & Hansen, P. (2017). Take-over requests analysis in conditional automated driving and driver visual research under encountering road hazard of highway. In: I. Nunes (Ed.), *Advances in human factors and systems interaction. AHFE 2017. Advances in Intelligent Systems and Computing* (pp. 230–240). Cham: Springer. https://doi.org/10.1007/978-3-319-60366-7_22.