Utrecht University

Testing the Effectiveness of Personal Names for Alerting Transitions in Semi-Automated Cars

Thesis (27,5 ECTS)

for the purpose of obtaining the degree of Master of Science
at Utrecht University

By
Jiaxin Xu

Student number	4640217
Master programme	Applied Cognitive Psychology
Department	Experimental Psychology
Faculty	Social and Behavioral Sciences
Supervisor	Dr. Christian P. Janssen
Assessors	Dr. Christian P. Janssen and Dr. P. Christiaan Klink
Date of submission	30 June 2022

Abstract

This thesis investigates how a speech alert that uses a driver's own name affects the speed, accuracy, and subjective experience of responding to transitions of control in the context of semi-automated driving. Preceding studies have shown that a personal name can be an effective stimulus to capture attention. However, studies that used names for alerts in automotive settings found mixed results when comparing the effectiveness between name and non-name conditions. In this thesis, an experiment with a simulated semi-automated car was conducted, in which participants were watching driving videos while playing a mobile game. Occasionally they heard speech alerts that started with either their name or a warning sound. They were required to immediately disengage from the game and respond to the alerts by pressing a button on the steering wheel or by pressing the brake. Results showed that participants' response times were significantly faster when hearing the alerts containing their names. There was no difference in subjective experience between conditions found. The results suggest that one's name is a useful addition to an in-car alert and warrants deeper investigation for future research.

Keywords: (semi-) automated cars, speech alerts, personal names, transitions of control, non-driving tasks, human-automation interaction

Table of Contents

1 Introduction

As cars become increasingly automated, human drivers' contribution to the cars' actions might be progressively decreasing (Köhn et al., 2019). At the upcoming middle levels of automation (i.e., SAE L3, SAE International, 2016), drivers might only occasionally intervene in driving while focusing more on non-driving tasks such as play and work. However, this multitasking scenario might severely impact driving safety: drivers might be absorbed in the non-driving tasks and thus not have sufficient attentional resources to handle a suddenly-occurred transition of control. At such a critical moment, auditory alerts are typically used to divert people's attention back to driving. Specifically, given humans' instinct to perceive languages, a speech-based alert potentially has a higher communicative ability to trigger faster and more accurate responses than other auditory elements such as an abstract sound (Lucas,1994; McKeown, 2005). A natural question to human-automation interaction (HAI) designers is then: how to design the speech alerts that enable efficient control transitions in (semi-) automated driving?

One promising way to improve the speech alerts in (semi-) automated driving is using one's personal name, as proposed by Maidhof (2020). This idea can be traced back to the early attention experiment by Cherry (1953) and the so-called "cocktail party effect" (CPE). According to CPE, people have the unique ability to notice their name despite multiple auditory components going on simultaneously (Cherry, 1953). Moray (1959) also observed that a personal name is one of only a few stimuli likely to be heard when presented to the unattended ear. Based on these observations, a driver's own name is hypothesized to be a strong cue for provoking their attentional processes, making them switch focus rapidly to the divided attention task of driving (Almén, 2002). Nevertheless, personal names have only been considered by a few studies in alerts in cars thus far.

A positive result can be found in the study by Maidhof (2020), in which the author compared the effectiveness of a driver's own name with a random (not own) name in alerting transitions in an automated setting. The results revealed that the participants responded significantly faster to the required transitions when they heard their own names in speech alerts (Maidhof, 2020). However, in other studies, when performance was compared between a personal name and other non-name auditory elements, such as a warning sound, no difference in response time was found (Giessen, 2021; Tobias et al., 2013). This inconsistency of results that whether using a personal name has benefits (Maidhof, 2020) or not (Giessen, 2021; Tobias et al., 2013) motivates this study.

In this study, I investigated how a speech alert that contained a driver's own name affected their takeover performance compared to an alert that contained a warning sound. I used a simulator-based experiment in which the participants were playing a mobile game while the car was performing automated driving (operationalized using videos). Occasionally the car issued a speech alert when there was an unexpected road hazard. At such a moment, the participants were asked to immediately disengage from the non-driving task and take corresponding actions such as pressing a button on the steering wheel or pressing the brake pedal. From this experiment, I analyzed participants' response time, response accuracy, and subjective experience.

The remainder of this thesis states the background on transitions of control in (semi-) automated cars, provides the theoretical rationale for in-car speech alerts and their design, explains why a personal name can be a promising addition to the alert, and further reviews the preceding studies using personal names for alerting transitions. After elaborating on the stimuli and experimental set-up, the results are reported. The findings are then discussed with reference to the previous works, the implications for theory and practice, the limitations, and possible directions for future research.

1.1 Transitions of control in (semi-) automated driving

Developments in vehicle technology already realize functions to assist human drivers in multiple ways. In the foreseeable future, it is expected that automated cars will not only stay in conceptual prototypes but instead be available for industrialized productions (Pfleging et al., 2016). As stressed by authorities like the Society of Automotive Engineers, automation of cars (sometimes also referred to as autonomy of cars) is not a simple binary state (i.e., no automation to full automation) but rather a stepwise process that will be implemented in different cars in different ways progressively over time (SAE International, 2016). Thus, before automated systems can handle all driving tasks under all traffic situations (i.e., full automation or SAE L5), human drivers must take over control at the proper time (i.e., SAE L1-L4, which is referred to as "semi-automated car" in this thesis). For example, partial automation (SAE L2), which is already implemented in commercially available cars, can take care of many complex driving tasks, including lateral (e.g., steering) and longitudinal control (e.g., acceleration and braking) but require the driver to keep their hands always on the steering and must stay alert to any road events. At more advanced levels, such as the upcoming SAE L3 (conditional automation), cars though still have operational limits, can drive themselves for a relatively long time with reduced human input. In such a case, drivers might only be interrupted occasionally by their car while having more "easy moments" to engage in non-driving tasks (Carsten et al., 2012; Janssen et al., 2019; Llaneras et al., 2013; Merat et al., 2014; Zhang et al., 2019).

Based on a meta-review (de Winter et al., 2014), non-driving tasks are expected to be performed more frequently as the automation level rises. Some traditional non-driving activities such as interacting with a mobile phone (e.g., making calls, sending texts), or looking at the roadside have already been widely observed in naturalistic driving studies (Dingus et al., 2016; Klauer et al., 2014). Moreover, an online survey (Pfleging et al., 2016) found that drivers also have high expectations and desires for many other non-driving tasks, such as eating, drinking, daydreaming, and browsing the Internet. Moreover, gaming is also observed to be growing highly wanted during automated driving (Krome et al., 2015; Neubauer et al., 2014). Indeed, non-driving activities might in some sense have benefits, as they might help in reducing boredom and fatigue (Neubauer et al., 2014; Pfleging et al., 2016), maintaining vigilant and avoiding underload (Atchley & Chan, 2011; Clark et al., 2017; Young & Stanton, 2002). However, they can still be a serious cognitive distraction that threatens driving safety. From a cognitive science perspective, when humans have to perform multiple tasks at the same time, the limited attentional resources will inevitably be allocated between different information stimuli (Farmer et al., 2018; Friedenberg et al., 2021). When the ongoing non-driving tasks are highly cognitively demanding (e.g., playing a mobile game), the extent of processing of task-irrelevant stimuli (i.e., the driving tasks) will be significantly reduced (Hancock, 2017; van der Heiden et al., 2018). This likely can lead to an "out-of-the-loop" situation, in which a driver is less likely to monitor their car's status, losing situational awareness of both it and the road environment (Bailey & Scerbo, 2007; Martens & van den Beukel, 2013; Parasuraman & Manzey, 2010). At such moments, if the car suddenly needs human intervention, the attentional resources might not be sufficient enough to deal with the suddenly-increased task demand, which will likely lead to increased delays or decision errors in responding to critical incidents (de Winter et al., 2014; Young & Stanton, 2007).

Given the likelihood that drivers work on non-driving tasks and the accompanying serious risks, it is essential to interrupt the ongoing non-driving task and divert the driver's attention back to the driving before a necessary control transition (Martens & van den Beukel, 2013). Theoretically, interruptions are distinguished between exogenous (i.e., external alert) or endogenous (i.e., self-interruption) initiated (Gerber et al., 2020). As the automation level rises, it was anticipated that people would rely more on external alerts to trigger their input (Janssen et al., 2019). A likely ideal candidate for delivering external alerts is auditory signals (van der Heiden et al., 2021). With their significant benefit of the omnidirectional feature (i.e., can be detected from multiple directions regardless of the head or eye orientation), auditory alerts have a higher chance of being perceived in the dynamic environment compared to visual signals. Moreover, with the engineering improvements in sound delivery capability, auditory feedback has become a ubiquitous role in cars, such as seatbelt warnings, collision warnings, or proximity notifications for parking-assist sensors (van der Heiden et al., 2018). Therefore, investigating the auditory alerts in (semi-) automated cars is timely. In the time-sensitive and safety-critical moment of control transitions, more effective auditory alerts await further explorations by the HAI researcher and designers.

1.2 In-car speech alerts and their design

The way of designing auditory alerts can vary widely, for example, through auditory icons (representative sounds of an object or event, e.g., Gaver, 1986), earcons (short musical sounds, e.g., Blattner et al., 1989), or spearcons (algorithmically compressed speeches without pitch change, e.g., Walker et al., 2006). Of all those possible approaches, speech is considered one of the most prominent ones (Šabić et al., 2019) and has a widespread prevalence in the HAI contexts, especially with the rapid advancement of conversational artificial intelligence. In fact, speech is a distinctive identification (Barthes, 1977) and a unique way of human communication (Nass & Brave, 2005). It appears to be human's instinct to use speeches to understand the world and shape their behaviors and reactions accordingly (Apple et al., 1979; Scherer, 1978). Non-speech auditory signals (e.g., an abstract sound) usually rely on a relatively complex learning process, as they require people first to establish the relationship between them to a particular event or context. In contrast, plain language (even if produced by machines) can automatically trigger an association to a previous semantic knowledge (Noyes et al., 2006; Richie et al., 2018;

Wogalter et al., 2002), making the information transfer more direct, immediate, and elaborate (e.g. Bazilinskyy & de Winter, 2015). Therefore, by using speech, auditory alerts can be empowered with higher communicative ability, resulting in faster and more accurate responses (Lucas, 1994; McKeown, 2005).

Using speech in alerts has already received attention in the driving domain (Gold et al., 2015; Iqbal et al., 2011; Mok et al., 2015; Politis et al., 2015; van der Heiden et al., 2017). For example, Iqbal et al. (2011) reported that speech alerts that notified critical road conditions were effective in lowering driving errors (compared to the trials without speech alerts) while the participants engaged in phone conversations while driving. Politis et al. (2015) investigated how speech alone or merged with visual or tactile cues affected takeover performance while the drivers were busy with a distraction task (a concentration memory game). The results showed that the speech combined with other modalities led to minor lateral deviation (i.e., better performance) than the speech alone.

Other than comparing speech and non-speech alerts, studies also focused on specific perceptual properties of speech itself, the first of which is the conveyance of perceived urgency (Edwards et al., 2021; Hellier et al., 2002). Given many observations that drivers take actions significantly faster in highly urgent situations (e.g., Mok et al., 2015; Politis et al., 2015), the safety-critical actions (e.g., transitions of control) might be effectively initiated by auditory warnings that have high-level perceived urgency. Word semantics has been observed to affect perceived urgency significantly. For example, the word "Danger" is observed to be perceived as more urgent than others like "Caution", "Notice" or "Warning", and can lead to faster reactions in the simulated driving (Baldwin & Moore, 2002; Baldwin, 2011). Urgent intonation style can also have a better alerting effect (Arrabito, 2009; Edworthy et al., 2003; Hellier et al., 2002; Wong et al., 2019). For example, Wong et al. (2019) found that more assertively-spoken commands (e.g., “Watch out! Brake immediately!”) was more effective in diverting drivers’ attention back to driving from non-driving tasks compared to the neutrally-spoken commands (e.g., “Please apply the brakes”). Furthermore, (though still controversial) speech spoken by females might convey higher urgency compared to that spoken by males (Edworthy et al., 2003; Hellier et al., 2002; Machado, Duarte, Teles, Reis, & Rebelo, 2012).

Another important perceptual property of speech alerts is pleasantness (Bazilinskyy & de Winter, 2017). If an alert becomes displeasing or annoying, it would likely be disabled or ignored (Eichelberger & McCartt, 2014; Parasuraman & Riley, 1997). An alert with inappropriate urgency can also come with a "startle effect", which evokes responses that, while timely, cause a lot of errors (Bliss & Acton, 2003). To this end, consideration for designing effective speech alerts might be not blindly exaggerating the importance of perceived urgency but instead creating a delicate balance between perceived urgency and pleasantness. One promising approach to making the alerts more pleasant is through a personalization approach (Hasenjäger & Wersing, 2017). As reported by Orth et al (2017), the speech assistant that was personalized to provide recommendations has received a more apparent preference. Moreover, the speech alerts using voices that match the listener's accent (e.g., using an American accent for American people) (Dahlbäck et al., 2007) and personality (e.g., using an extroverted voice for extroverted people)(Nass & Lee, 2000) were also observed to be more attractive and acceptable.

Based on the above suggestions, the current study seeks to explore a both urgency-transmitting and well-accepted auditory element, which can be applied in speech alerts to support efficient transitions of control in (semi-) automated cars. That is, to be specific, a driver's personal name.

1.3 Personal names in speech alerts

The idea of using a personal name to attract attention can be traced back to the early attention experiment by Cherry (1953) which aimed at investigating how people selectively attend to one message amid other noises. In the experiment, participants were given different messages to each ear and asked to repeat aloud the message that was heard in a specified ear. It was found that participants could detect their names from the unattended channel (the channel they were not shadowing). This suggests that human brains possess the unique ability to focus one's auditory attention on a specific stimulus while filtering out a variety of others, which was defined as the "cocktail party effect"(CPE) (Cherry, 1953). Neville Moray did similar research in 1959. In the experiment, participants were asked to attend to a message they heard in one single ear while ignoring the message given to the other ear. The ignored message occasionally comprised the participants' own names. The results showed that unattended messages preceded by participants' names were recognized more frequently than those without the names (Moray, 1959). This study was later modified by Wood and Cowan (1995), and the conclusion was further replicated. Moreover, another early attention experiment by

Treisman and Riley (1969) also suggested that people are permanently primed to detect personally significant words, like their names, and theorized that they might require less perceptual information than other stimuli to trigger identifications. Taken together, one's name seems to be a potent attentional attractor, even in the presence of other competing stimuli (though exceptions exist, e.g., Conway et al., 2001).

Later cognitive neuroscience works also highlighted the unique ability of human brains to perceive personal names (Folmer & Yingling, 1997; Gray et al., 2004; Holeckova et al., 2006; Holeckova et al., 2008; Müller & Kutas, 1996; Nitz, 2006; Perrin et al., 1999; Perrin et al., 2006). Studies showed that typically developing infants can orient to the sound of their own name (Mandel et al., 1995), recognize it in noisy environment (Newman, 2005), and even distinguish it from names with only a different first phoneme (Mandel-Emer et al., 2003) by the age of 5 months. Research using event-related potentials (ERPs) showed that P300 was augmented for self-relevant stimuli (such as a name) relative to control stimuli, which indicated that the attention resources were highly allocated to the self-relevance information (Gray et al., 2004). Moreover, personal names were even found to trigger specific brain responses during sleep (Perrin et al., 1999; Portas et al., 2000) and even while in vegetative and coma conditions (Fischer et al., 2008; Fischer et al., 2010; Perrin et al., 2006; Schnakers et al., 2008).

Given the human's instinct of processing their name, it could be conceivable that a personal name can be helpful in alerts for drawing attention and reinforcing the alerting effectiveness. Researchers in different fields have been trying to explore this possibility. For example, in residential contexts, a child's first name has been used in a speech smoke alarm to awaken sleeping children (Smith et al., 2019; Smith et al., 2006); In the construction industry, a personal name was considered in alarm sound for reducing accidents caused by blind spots (Chae & Kang, 2021). In the driving context, to the best of our knowledge, only four studies have been done on this topic thus far.

In a study that aimed at investigating the alert effectiveness during manual driving, Almén tested four different conditions, including an auditory alert with a name, vibration as a tactile alert, the combination of both, and a control condition without alerts. Although no difference in participants' response was found between these manipulations, the response time when the alerts were presented by a combination of the name and vibration was observed to be slightly faster (Almén, 2002).

Tobias et al. (2013) later tested the attention-getting effect of a personal name against a warning tone in an automated driving set-up where participants engaged in a non-driving task (i.e., inputting a number sequence into a number pad). The results showed that participants who heard their name reacted marginally faster than those who heard the tone, and 77% of participants showed positive interest in having their names in alerts. The authors suggested that a name can be seen as a promising cue in diverting drivers' attention back to driving. However, this study only used standalone names without additional messages in the alerts. The author proposed a suggestion that future studies should consider combining a personal name with other directional cues (e.g., "Sarah! Steer Right!") to potentially enhance the alerting effectiveness (Tobias et al., 2013).

Further, Maidhof (2020) addressed the above limitation by testing whether there is a benefit to combining the personal names with assertive commands that explicitly expressed the driving situations and required actions (e.g., "Caterina! There is traffic coming up. Watch out! Brake right now"). In Maidhof (2020)'s study, participants watched video clips of driving scenarios while playing a mobile game in a driving simulator, and sometimes had to respond to speech commands (adapted from Wong et al., 2019) by making proper responses such as pressing the brake and indicating left or right. The results showed that the alerts that contained the driver's own name resulted in a significantly faster response than those that contained a random (not own) name (Maidhof, 2020). However, there were still limitations. The response time observed was noticeably slower than Wong et al. (2019) who used almost identical experimental set-up and stimuli. The potential cause of the difference might be that the critical incidents happened (in the video) 5s after the signal words emanated in the speech, whereas in Wong et al. (2019)'s study the incidents happened quite immediately after the onset of the alert. This relatively long response window might have made participants feel calmer and thus have resulted in a delayed response (Maidhof, 2020). Moreover, the length of the two-sentence commands (including a scenario command and an executive command) might also be too long and varied, which might limit the comparability among trials. Furthermore, the driving videos were filmed in a left-hand-traffic country (i.e., the UK), which went against the right-hand-traffic habits of the participants. The uncontrolled variables such as the driving speed and traffic conditions in the driving videos might also have influenced the participants' performance.

Recently, following the idea of Tobias et al. (2013) and Maidhof (2020), Giessen (2021) tested the effectiveness of using personal names compared to a warning tone using the same experimental set-up with the same driving videos and speech commands as Maidhof (2020). However, the study revealed conflicting results: there was no significant effect on neither response time nor response accuracy between conditions, though the name condition was subjectively reported to be favorable. The author reported two possible reasons for this non-difference result. The first concern was also about the long length of speech messages: the largish time duration that preceded signal words possibly invited many nuances in performance. Secondly, some videos probably led to faster recognition of the upcoming actions and therefore should be changed (Giessen, 2021).

As evident from the above studies, there are mixed results on using personal names in speech alerts in automated settings. Maidhof (2020) found one's own names reduced the response time compared to random names, but it is unclear yet whether one's own names can outperform other auditory contents, as studies (Tobias et al., 2013; Giessen, 2021) found no significant difference between the name and a warning sound. This inconsistency of results on whether using a personal name has benefits (Maidhof, 2020) or not (Tobias et al., 2013; Giessen, 2021) motivates this study.

1.4 Goals and hypotheses

This thesis aims to further test whether a driver's personal name is an effective addition for alerting the transition of control in (semi-) automated driving, compared to a traditional warning sound. Two research questions are of this study's interest, which is also similar to Maidhof (2020)'s and Giessen (2021)'s studies:

RQ1: Do people respond faster and more accurately to transitions of control when a speech alert contains a personal name compared to a control condition where an alert contains a warning sound?

RQ2: How do participants experience using their personal name in alerts compared to using a warning sound?

To investigate the questions, I conducted an experiment in a simulated semi-automated car where participants were doing a non-driving task (i.e., playing a mobile game) and were required to make an immediate response (i.e., pressing a button on the steering wheel or pressing the brake) after they heard an alert that either contains their name or a warning sound. For RQ1, I measured the participants' response time and response accuracy to the alerts. For RQ2, I compared participants' preference, openness, perceived helpfulness, and annoyance, analyzed their subjective performances with different alerts, and also thematized the answers to open-ended questions.

Following the idea of Maidhof (2020) and Giessen (2021), I used participants' names in combination with the assertive speech commands originated by Wong et al. (2019), which have been reported to be more urgent and disruptive (in a good way) from a non-driving task. Given that the previous speech commands were relatively long and varied (Giessen, 2021; Maidhof, 2020), I shortened and uniformized the speech command length to a noticeable degree (i.e., from M=4.225s, SD= 0.709s to M= 2.588s, SD= 0.157s, see Appendix A). To address the limitations in the video stimuli (i.e., left-hand traffic and the possibly too long response window), I provided a new driving video set filmed in the Netherlands (right-hand traffic), and modified the moment when the alert was presented such that the interval between the signal word and the situation in need was 3s instead of 5s. Finally, to be more rigorous, I counterbalanced the driving speed and traffic complexity in the videos to reduce the potential interference of irrelevant variables on participants' takeover performance.

Based on these modifications, I hypothesized that the current study could replicate Maidhof (2020)'s results, demonstrating that names can significantly reduce response time when a critical control transition is needed (H1); participants' subjective experience of the alert with their names can be positive (Giessen, 2021; Maidhof, 2020; Tobias et al., 2013) and potentially even better than a warning sound (H2). More generally, as an outcome, I expect this thesis to accumulate more empirical evidence for using personal names to stimulate attention in multitasking scenarios and will also provide valuable practical contributions to informing the speech-based alert design for the rapidly-growing automated systems.

2 Methods

2.1 Materials

2.1.1 Driving videos

To redesign the driving video stimuli, researchers (Janssen and Xu) drove a car (automatic transmission, the driver sits on the left side) for a total of 43-minutes of continuous filming. The route was a loop between the Utrecht "Veemarkt" neighborhood and Utrecht Science Park. This route was chosen as it contained sections with different speed limits, including a relatively low speed- 30km/h (cf. Abe & Richardson, 2004, 2005) and a relatively high speed- 70km/h, and also contained varying traffic density (i.e., the density of the surrounding cars), including a relatively low density and relatively high density. This operationalization was issued to counterbalance the impact of latent variables in the driving situation on participants' takeover behaviors (Gold et al., 2016; Gold et al., 2014; Radlmayr et al., 2014; Vogelpohl et al., 2017).

Video and audio recordings of the driver's forward views were filmed using a GoPro9 camera mounted using the GoPro magnetic swivel clip. The camera was positioned above the front console inside the car to ensure it was not obstructing the driver's view. The camera angle was verified from the driver's position inside the car to confirm it accurately reflected the forward-facing perspective of the road ahead. The camera was controlled by the co-driver (Xu) by a smartphone, and the recording started as soon as the driver (Janssen) released the handbrake.

The 43-minute recording was edited using Final Cut Pro software. To counterbalance driving speed and traffic density, I selected four separate video segments corresponding to four different driving scenarios, including a 30km/h-limit and low-traffic-density driving situation (video A), a 30km/h-limit and a high-traffic-density driving situation (video B), a 70km/h-limit and low-traffic-density driving situation (video C), and a 70km/h- high-traffic-density driving situation (video D) (see Table 1 for the screenshots of four videos). The background noise of the videos was kept to enhance the realism of the driving scenarios. The duration of the videos varied between three and four minutes. Each video had six required take-over actions that needed to be alerted with voice commands (described in more detail below). Video B and video D contained 5 instances where pressing the brake were required and 1 instance where direction had to be indicated, while video A and video C contained 3 instances of pressing the brake and 3 instances of indicating direction (see Table 1).

The videos were inputted into the JavaScript software taken from Wong et al. (2019), which could play the videos, log user actions, and keep track of their response time.

Table 1. *Description of the re-designed four driving videos.*

Videos (screenshot of typical scenario)		Characteristic	Response actions required		
			Action type	n	total
A		30km/h-limit & low-traffic-density driving situation	Pressing brake	5	6
			Signaling direction	1	
B		30km/h-limit & high-traffic-density driving situation	Pressing brake	3	6
			Signaling direction	3	
C		70km/h-limit & low-traffic-density driving situation	Pressing brake	5	6
			Signaling direction	1	
D		70km/h-limit & high- traffic-density driving situation	Pressing brake	3	6
			Signaling direction	3	

2.1.2 Speech alerts

The overall components of the speech alerts used in the current study were in line with the assertive speech alerts of previous relevant studies (Giessen, 2021; Maidhof, 2020; Wong et al., 2019). All speech alerts consisted of a sequence of (1) an attention attractor - each participant's name or a warning sound, and (2) a scenarios command describing the upcoming road situation (e.g., "Beware of red light"), and (3) an executive command containing a signal word which gave explicit instruction on the required action (e.g., "Brake right now", with the signal word "*Brake*").

Compared to the previous study, which used a voice actor (Wong et al., 2019) or Apple's text-to-speech program on a Mac (Giessen, 2021; Maidhof, 2020), all speech components in the current study were developed by a professional AI speech program named MURF Studio (available from https://murf.ai), which could reproduce high-quality natural speech based on the voices of human speakers and offers high-quality speech, adjustability of speech rate, and a variety of national languages. This study used artificial voice instead of a voice actor because the research scope is framed in the context of in-car alerting systems driven by currently available technology (Maidhof, 2020).

2.1.2.1 Personal names

For all the participants, the proper usual name was used (participants were required to provide their usual names using the question "What would you prefer to be called?" before the experiment). The names were pronounced using the "Sophie" voice (female, US English accent) in MURF Studio. In cases where a participant's name could not be pronounced appropriately in English, another female voice from the native language of the participant was used (P09, P12, P14, P18, P19, P23, P28 with the Dutch voice "Mila", P25 with the German voice "Adele", P10 with the Indonesian voice "Indah", P06, P11, P22, P26 with the Chinese (Simplified) voice "Liu", P20 with the Turkish voice "Azra", P24 with Finnish voice "Heta", and P15 with the Italian voice "Adriana"). All names were generated under default speed (1×), pitch (1×), and emphasis.

2.1.2.2 The warning sound

The warning sound selected for the control condition was finalized by two researchers after two rounds of screening from multiple options. To ensure ecological validity and be non-startling for participants, we finalized a sound that people are familiar with in real driving situations instead of a pure tone (cf. Giessen., 2021). The selected warning sound comprised multiple harmonic components and was typically active around 750 Hz, with a timbre characteristic of a crackling "*ding*", close to a seatbelt unfastened alarm (available for download at https://sounds-mp3.com). Based on the suggestion from previous studies, an acoustic sound of around 740Hz is efficacious in conveying urgency and stimulating drivers' responses to risky situations (Cabrera & Ferguson, 2006; Di Stasi et al., 2010). I further modified the loudness of the sound using Audacity software to make it more comparable to the voice of names. As seen in Figure 1, the modified warning sound had similar waveform peaks versus the voice of names (e.g., "Jiaxin"). The sound was also edited to keep a constant period of 0.550 sec, which was approximately consistent with the average time of all the personal names. The comparison of duration between names and sound can be found in Appendix F.

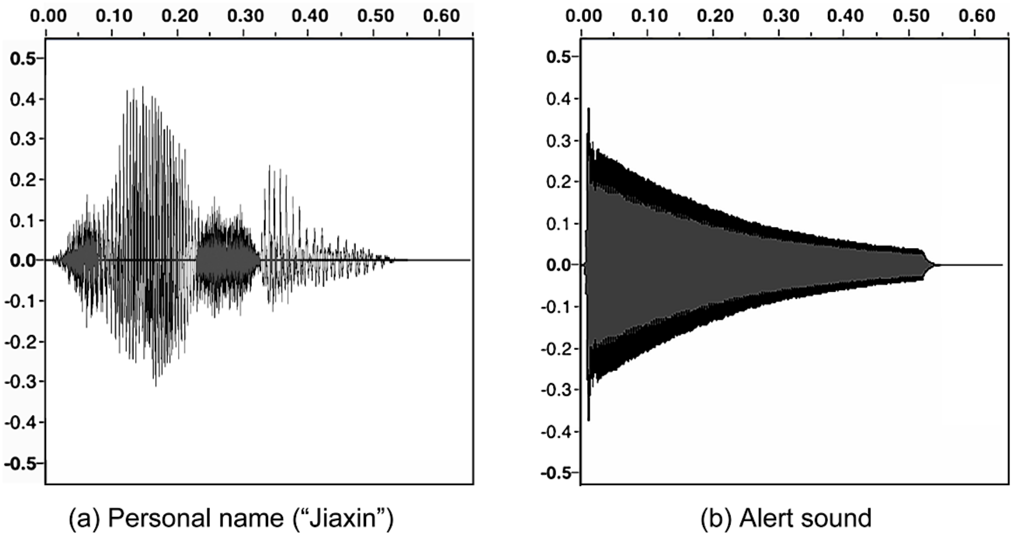

Figure 1. Waveform records of (a) name (e.g., “Jiaxin”) and (b) the warning sound. The x-axis is time, and the y-axis is power (dB).

2.1.2.3 Speech commands

To have a consistent speech experience, all commands were generated using the “Sophie” voice at the same parameters as the personal names and without added noise. In line with (Giessen, 2021; Maidhof, 2020; Wong et al., 2019), the speech commands consisted of the requirements of everyday driving actions, including signaling left or right on the steering wheel and slowing down or braking on the brake pedal. To shorten and uniformize the speech messages, I cut out some redundant parts of the previous sentence without changing their original meaning. For example, I replaced the wording of “Red traffic light coming up!” (scenarios command) with “Beware of red light!”, and replaced “Look up! Action to indicate right is needed” (executive command), with “Indicate right!”. In this manner, I reduced both the length (lowered the mean from 4.225s to 2.588s) and the variations (lowered the standard deviation from 0.709s to 0.157s, see Appendix A), which allowed for more reliable comparability across trials.

Table 2. *Scenario commands and executive commands used in the driving videos.*

Scenario Commands		Executive Commands	
Content	Duration (s)	Content	Duration (s)
Beware of red light	1.400	*Brake* right now	1.117
Beware of T-junction ahead	1.750	*Slow* down	1.017
Car coming from the right	1.500	Indicate *right*	1.050
Car coming from the left	1.533	Indicate *left*	1.083
Moving to the right lane	1.517		
Moving to the left lane	1.550		
Beware of the car ahead	1.400		
Exit roundabout now	1.483		
Narrow road ahead	1.317		

Note. Signal words indicating actions are written in *italics*.

Moreover, I shortened the time interval from when action keywords (e.g., "left", "brake") were said to the moment where the incident visually took place in the video (i.e., made the response window shorter). In the previous studies, each signal word in a speech voice command was positioned standardly 5 seconds before the incidents happened on the road (Giessen, 2021; Maidhof, 2020). There were indeed suggestions supporting that a take-over request should be issued at least 5–8 s before the critical incident on the road (Gold et al., 2013; Petermann-Stock et al., 2013; Zhang et al., 2019). However, such a timeframe might not be an ideal choice here as our work was meant to focus on the safety-critical situation when an in-car warning is truly a last resort. As reported by Maidhof (2020), a 5s response window was overly long which potentially slower participants' response. To this end, I reduced the time interval of "keyword to incident" from 5s to 3s (i.e., placed the alert closer to the visual appearance of a critical scenario). This operationalization was (assumed to be) more consistent with the "last-minute alerts" in real-world imminent situations that would stimulate more urge to respond faster. Figure 2 shows an example timeline of a speech alert and the subsequent participant take-over procedure. As is shown, the onset of the critical action word (e.g., "*right*") occurred 3s before the incident visually took place on the road (e.g., the car started to move lanes).

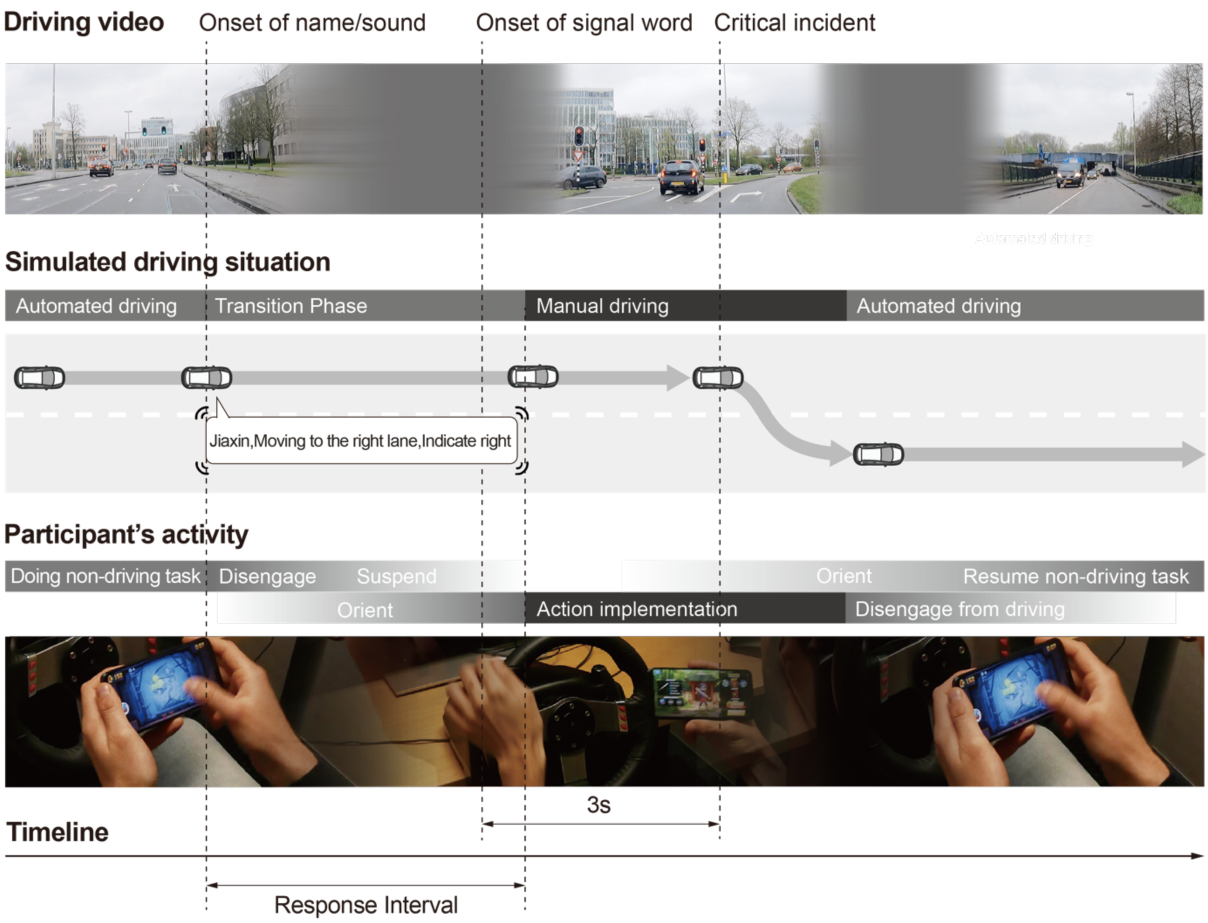

Figure 2. Illustration of the take-over procedure in this experiment.

Note. The response interval starts at the initiation of a name (or the warning sound) and lasts until the user begins to physically take action (e.g., tapping the directional buttons or pressing the brake pedal). This figure is a modification of Figure 2 in Maidhof (2020) and with reference to the framework of transition of control by Janssen et al. (2019).

2.1.3 Apparatus

The experiment was conducted in a 200cm*200cm lab space. The lab space was completely enclosed and equipped with high-quality acoustic barriers, which could effectively prevent other disturbances outside the experiment. The Participants were positioned in a Logitech G27 racing system including a steering wheel and foot pedals (i.e., a clutch, a brake, and an accelerator), which were also used by (Giessen, 2021; Maidhof, 2020; Wong et al., 2019). The steering wheel was mounted on a desk and the foot pedals were placed comfortably on the floor in front of the participants. An HP E27d G4 monitor showing driving videos was positioned 50 cm in front of the participants. Two HP 2.0 speaker systems S5000 were used to play the speech alerts. A black curtain was placed behind the monitor better to focus the participants' attention on the screen.

Figure 3. The driving simulator set-up.

2.1.4 Non-driving task

Playing a mobile game was selected as the non-driving task in this study, which was also in line with (Giessen, 2021; Maidhof, 2020; Wong et al., 2019). We did not use standardized non-driving task[1] because we meant to observe the takeover behaviors of the driver when they are engaging in a realistic distraction task. Similar to Wong et al. (2019), I selected the mobile game Fruit Ninja 2. The game was played on a Samsung Galaxy A40 smartphone. In the game, the player must slice fruit that is randomly thrown into the air by swiping the touch screen with their fingers and they must not slice bombs. The player mode "Arcade+" was selected in which slicing bombs will deduct 10 points and some "special bananas" can appear, which have unique bonuses such as doubling points or freezing the time for a few seconds. Participants were asked to restart the game each time it was over. The game performance was not recorded during the experiment.

[1] Standardized non-driving tasks are those can easily manipulate the task demands or workloads, thereby precisely inducing the target state of the participants, such as peripheral detection tasks and surrogate reference tasks. This type of task were not selected for this study because they are not natural tasks being performed in real driving situations and have low ecological validity (Lee et al., 2021).

2.2 Participants

Twenty-eight participants were recruited through opportunity sampling (10 males and 18 females, M = 25 years of age, SD = 2.22 years of age). All participants had a valid driving license. Seven people usually drove in the Netherlands, eleven people usually drove in China, two participants usually drove in Indonesia, and the other eight usually drove in their home countries (India, Italy, Spain, Turkey, Brazil, Finland, Germany, and Canada). Furthermore, all participants had a self-reported normal or corrected-to-normal vision and no hearing difficulties. This study complied with the tenets of the Declaration of Helsinki and was approved by the Ethical Review Board of the Faculty of Social and Behavioral Sciences of Utrecht University (approval number 22-1435). Participants were rewarded with 0.5 credit points (PPU) or 5 euros for their time. All participants were provided with informed consent before taking part.

2.3 Study design

A one-way within-subject design was used: the speech alert was either preceded by the driver's name (the name condition), or a warning sound (the sound condition). Participants watched four driving videos (i.e., video A, video B, video C, video D), each containing 6 required actions. Each condition was used in two videos, resulting in 12 measurements for each condition per participant.

The video orders that were presented to participants were based on the rule that the name condition and the sound condition had to be alternated. Moreover, given that two pairs of videos (video A & video C; video B & video D) had the same frequency distribution of action type (i.e., signaling direction and pressing the brake, see Table 1), they were banned from coexisting in the same condition (i.e., the name or the sound) to ensure each participant experienced the same amount of braking action and signaling action per condition. These rules finally led to: 2 (start with name or tone) × 4 (the 1st video) × 3 (the 2nd video) × 2 (the 3rd video) × 1 (the last video) – 16 (orders when the two pair coexist in the same condition, 2×4×2×1×1) = 32 unique orders. Of these, 14 orders were randomly selected. To counterbalance the order effects, the other 14 participants were given the inversed orders (see Appendix B for all the finalized orders).

2.4 Procedure

Participants took place in front of the driving set-up upon arrival and received a brief instruction on the context. Specifically, participants were instructed to imagine that they were playing a mobile game during semi-automated driving, and occasionally speech alerts would appear to request them to input in the driving. It was explained that there would be two types of required actions including (1) pressing corresponding buttons on the steering wheel in response to the signal words "indicate left/right" and (2) pressing the brake pedal in response to the signal words "slow down" or "brake". Participants were told that the speech alerts would either start with their own name or with a warning sound, and they were emphatically required to respond as fast and accurately as possible. After the instructions, participants completed informed consent.

After signing, participants were given a practice trial to get acquainted with the set-up and the operation of the steering wheel and foot pedals. They were given a video sample of 60 seconds (similar in style to the experimental videos, but not the same) while also playing the game "Fruit Ninja 2". In the 60-second practice, they executed 1-time indicating left and 1-time pressing the brake. The experimenter stayed nearby to answer possible questions (e.g., not knowing how to play the game) and clarify their improper operations (e.g., pressing wrong buttons on the steering wheel, forgetting to put their foot back in place after pressing the pedal, putting the phone on the table while playing).

In the main experiment, the participants watched the four driving videos while playing the mobile game. The experimenter waited outside and only returned to the lab until each video had finished. Participants' responses and corresponding times were recorded automatically by the JavaScript software of Wong et al., (2019). After the experiment, participants were asked to complete a questionnaire (described in more detail below). In total, the experiment lasted around 30 minutes per participant.

2.5 Measures

2.5.1 Response time

The response time (RT) in the current study was defined as the time that participants started to take physical action after they heard the name (or for the control condition: the warning sound) (see Figure 2 for the schematic representation of RT). This measure was the same as Tobias et al. (2013) and Almén (2002), and also in line with many previous studies which measured response to audio signals during autonomous driving (e.g., van der Heiden et al., 2018; van der Heiden et al., 2019; van der Heiden et al., 2021). However, this measure differed from Maidhof(2020), which used a signal word (i.e., *right*/left/brake/*slow down*) to start the response interval timing (see Figure 2). As in Maidhof (2020)'s study, the speech messages were relatively long and varied; the signal word thus could serve as a clear benchmark that allowed for a fair comparison between trails. In contrast, the current study used speech that was significantly shorter and less varied in length, thus the name (or the sound) was selected as the onset of measurements as it was the most direct reflection of the alerting effect of the stimuli itself. To determine the time interval, I always documented the timecode of every first gamepad response (logged by Javasoft ware) and performed subtraction between it and the timecode of the name (or the sound) onset (logged by the video editing software). I analyzed the mean of median RT to reduce the effect of extreme values (i.e., extremely fast or extremely slow responses).

2.5.2 Response accuracy

The accuracy of the response consisted of a measure of whether a response matched the required action given in the in-car alert. For each trial, only the first response made by a participant after the speech alerts were used in the analysis. An incorrect response included a false response (e.g., pressing right when the correct action was to press left) and a missed response (in line with Maidhof, 2020 and Giessen, 2021,). These data are reported but not part of the statistical analysis.

2.5.3 Subjective experience

A questionnaire containing 29 questions was taken from Maidhof, 2020 and Giessen, 2021. Apart from the demographic questions (Q1-Q15, e.g., gender, age, driving experience), participants were asked to rate their subjective experiences on four dimensions – *preference* (Q16, Q21), *openness* (Q17, Q22), *helpfulness* (Q18, Q23), and *annoyance* (Q19, Q24). These questions originated from the literature (Bazilinskyy & de Winter, 2017; Tobias et al., 2013)(see Table 3). Each dimension was typically made on a 5-point Likert scale ranging from "completely disagree" (1) to "completely agree" (5), with the center point labeled "neither disagree nor agree". Note that for *preference*, the question was asked as a negative statement such that the coding was reversed (see Table 3). There were five open-ended questions in the end regarding participants' subjective general performance (Q25), subjective performance per condition (Q26, Q27), their opinions about using personal names in an in-car alert (Q28), and any further remarks about the experiment(Q29). I used a thematic analysis (Braun & Clarke, 2006) to identify common emerging themes presented within participants' answers. There was also an additional question about people's concerns about using their names. The analysis of this question is reported in Appendix G.

Table 3. *Four dimensions of subjective experience and the corresponding coding and references.*

Dimension	Question	Coding	Reference
Preferences	"I don't like systems using my own name/ the sound in a warning"	1=Completely agree, 2=agree, 3= neither disagree nor agree, 4=disagress, 5= completely disagree	Tobias et al. (2013)
Openness	"I like it when my car uses my own name/ the sound within a warning"	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree	Tobias et al. (2013)
Helpfulness	"Hearing my own name/the sound in a car warning is helpful"	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree	Bazilinskyy & de Winter (2017)
Annoyance	"Hearing my own name/the sound in a car warning is annoying"	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree	Bazilinskyy & de Winter (2017)

Note. The complete questionnaire can be found in Appendix C.

2.6 Statistical analysis

All quantitative data were analyzed using Bayesian hypothesis testing (Wagenmakers et al., 2016), which generates the evidence supporting the alternative hypothesis and is expressed as the odds ratio in favor of the alternative hypothesis compared to the null hypothesis given the data (i.e., Bayes Factor, BF)[2]. BF hypothesis decisions are expressed

[2] In the current study, BF was used instead of the *p*-value to report results. The conventional null hypothesis significance tests offer little information when the difference is not statistically significant and only the alternative hypothesis (H_1) is tested. In contrast, BF allows evidence for the extent to which the data support both two hypotheses(H_1 and H_0). Moreover, Bayesian statistics is more suitable for studies with a relatively small sample size in HCI research (Held & Ott, 2018; Kay et al., 2016).

in a range from “anecdotal” to “extreme” evidence for a hypothesis (Isaacs et al., 2001). For labeling of effect size, I adopted the convention that a BF less than 1/10 implies strong evidence for the lack of a difference, a BF between 1/10 and 1/3 provides moderate evidence for the lack of a difference, and a BF between 1/3 and 3 suggests anecdotal evidence for the lack or presence of a difference (for $1/3 < BF < 1$ or $1 < BF < 3$, respectively), a BF between 3 and 10 denotes moderate evidence for the presence of a difference, a BF between 10 and 100 implies strong evidence (Lee & Wagenmakers, 2014). Table 4 shows the BF classification and the adapted interpretation. Depending on the context, the one-side BF (BF_{+0} or BF_{-0}) or the two-side BF (BF_{10}) was reported. All Bayesian analysis was processed using JASP (Jeffreys' Amazing Statistics Program; version 0.9.2) with a default Cauchy prior width of 0.707.

Table 4. *Bayes factor classification and interpretation.*

H_0	Strength of evidence	H_1
BF = 1	No evidence	BF = 1
$1/3 < BF < 1$	Anecdotal evidence	$1 < BF < 3$
$1/10 < BF < 1/3$	Moderate evidence	$3 < BF < 10$
$1/100 < BF < 1/10$	Strong evidence	$10 < BF < 100$

Note. BF: Bayes factor, H_0: Null hypothesis, H_1: Alternative hypothesis.

3 Results

3.1 RQ1: Assessment of takeover performance

3.1.1 Response time

The experiment produced 28*24 = 672 trials in total, with 336 trials per condition. Due to equipment failure, data of 4 trials for Participant 16 were missed. To avoid statistical problems associated with missing data, I imputed values for the outlier trials using a regression procedure in which missing value estimates were based on the relationships among responses across all trials (Hultsch et al., 2002). There was also 1 trial when participants did not make a response, for which the RT was adjusted to the maximum observed RT, 7.181s. The means of median RT (M_{med}) in both conditions were analyzed by a Bayesian paired-sample t-test. The alternative hypothesis was that the sound condition results in a longer RT than the name condition (H_{+}: $RT_{sound} > RT_{name}$; cf. Maidhof, 2020).

The result showed that the alerts preceded by the warning sound (M_{med} = 3.461s, SD_{med} = 0.434s, 95% CI = [3.293s, 3.630s]) was longer than the alerts preceded by participants' name (M_{med} = 3.280s, SD_{med} = 0.513s, 95% CI = [3.081s, 3.479s]) (see Figure 4). The Bayesian paired t-test found moderate evidence for this difference, BF_{+0}= 4.495. The Bayesian sequential analysis shows that the BF for H_{+} increases and gradually rises above 3 as the number of data points increases (see Figure 5). The additional analysis of RT under different action types can be found in Appendix D.

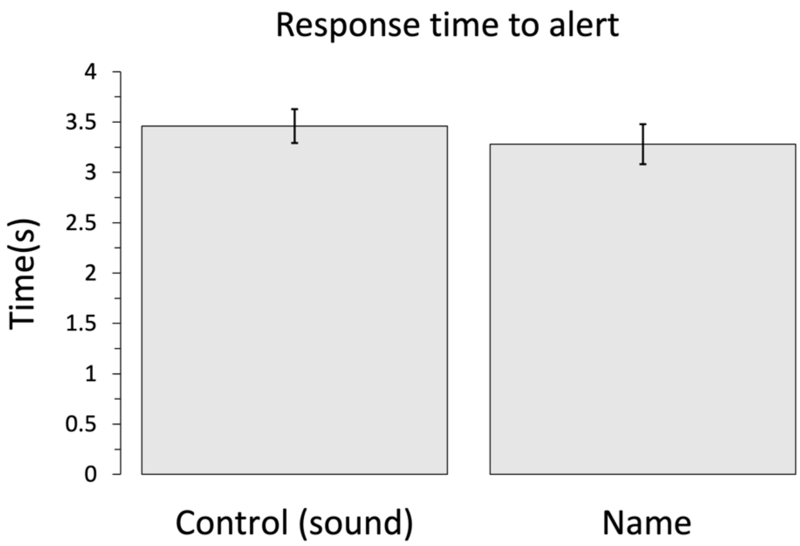

Figure 4. Response time to the alert. Error bars show 95% confidence intervals.

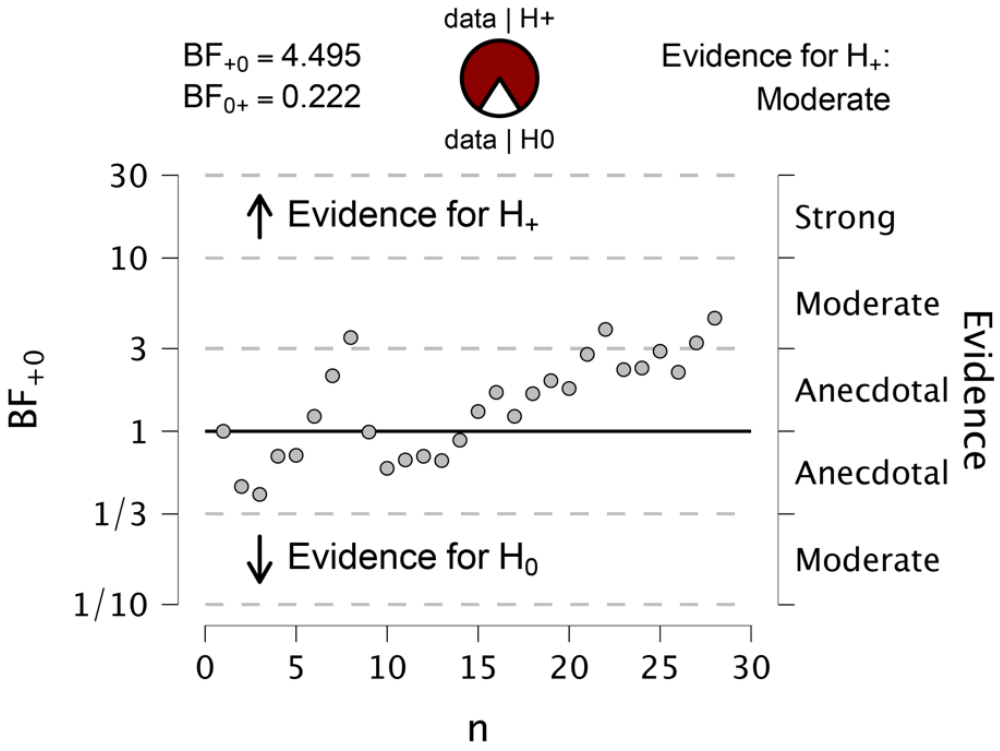

Figure 5. Bayesian sequential analysis showing the BF_{+0} and the progression of the BF_{+0} with the accumulation of data points (H_+: $RT_{sound} > RT_{name}$).

3.1.2 Response accuracy

The accuracy of the name condition was 99.4%, as in 330 out of 332 trials (4 trials missed because of equipment failure) correct responses were given. In the sound condition, the accuracy was 97.6%, with 328 out of 336 trials correct. Only a few incorrect responses were observed, with no clear pattern in the occurrence of those misses and incorrect responses (see Table 5).

Table 5. Response accuracy.

	Name condition	Sound condition
Total number of trials	332 (100%)	336 (100%)
Correct response trials	330 (99.4%)	328 (97.6%)
Incorrect response trials	2 (0.6%)	8 (2.4%)
Trials with false response	2 (0.6%)	7 (2.1%)
Trials with missed response	0 (0.0%)	1 (0.3%)

Note. Incorrect trials included the trials with misses and false responses.

3.2 RQ2: Assessment of subjective experience

3.2.1 Preference, openness, helpfulness, and annoyance

Bayesian paired t-tests found moderate evidence that there is no effect of condition on all four dimensions of subjective experience (for *preference*, BF_{10} =0.201; for *openness*, BF_{10} =0.206; for *helpfulness*, BF_{10} =0.201; for *annoyance*, BF_{10} =0.218). I also conducted one-sample Bayesian t-tests against the midpoint value of three, to test if subjective experience in any dimension was above neutral (H_+: subjective rating ＞3). The results showed strong evidence for the hypothesis on all four dimensions (for preference, openness, helpfulness, all BF_{+0} ＞10, for *annoyance,* both BF_{+0} ＜ 1/10) (see Table 6).

Table 6. *Bayesian statistics for preference, openness, helpfulness, and annoyance.*

Dimension	Condition	M	SD	Bayesian Paired T-Test		Bayesian One-sample Test	
				BF_{10}	error %	BF_{+0}	error %
Preference	Name	3.857	1.239	0.201	0.031	63.231	~ 3.318e-4
	Sound	3.857	1.008			457.903	~ 9.396e-6
Openness	Name	3.750	1.076	0.206	0.031	67.325	~ 2.673e-4
	Sound	3.821	0.819			3277.536	NaN[a]
Helpfulness	Name	4.179	1.020	0.201	0.031	23330.926	NaN[a]
	Sound	4.179	0.723			7.913e+6	NaN[a]
Annoyance	Name	2.179	1.090	0.218	0.031	0.049	~ 0.013
	Sound	2.071	0.979			0.023	NaN[a]

Note. For all tests, the alternative hypothesis specifies that the mean is greater than 3.
[a] t value is large. A Savage-Dickey approximation was used to compute the Bayes factor but no error estimate can be given.

Further, the frequency distribution of each dimension per condition was analyzed (see Figures 6 and 7). Answers towards the left in the histogram imply a more negative view, while answers towards the right suggest a more positive experience (Maidhof, 2020). In the figures, right-hand-sided frequency distributions are noticeable in all four dimensions in both the name and the sound conditions, indicating that most participants reported a positive subjective experience. Specifically, for the name condition, 20 out of 28 participants reported that they (completely) disagreed with the statement stating not to like the alert contained their names; 21 out of 28 would be open to a car that uses their names; 22 out of 28 indicated that using names in a warning was helpful, and 19 out of 28 did not find the use of their name annoying. However, 5 participants agreed that hearing their name induced annoyance. For the sound condition, 21 out of 28 liked the sound; 20 out of 28 felt open to a car using the sound; 23 out of 28 perceived that the sound was helpful, with no participants reporting disagreement; and 22 out of 28 did not find the sound annoying, with also 5 finding it annoying.

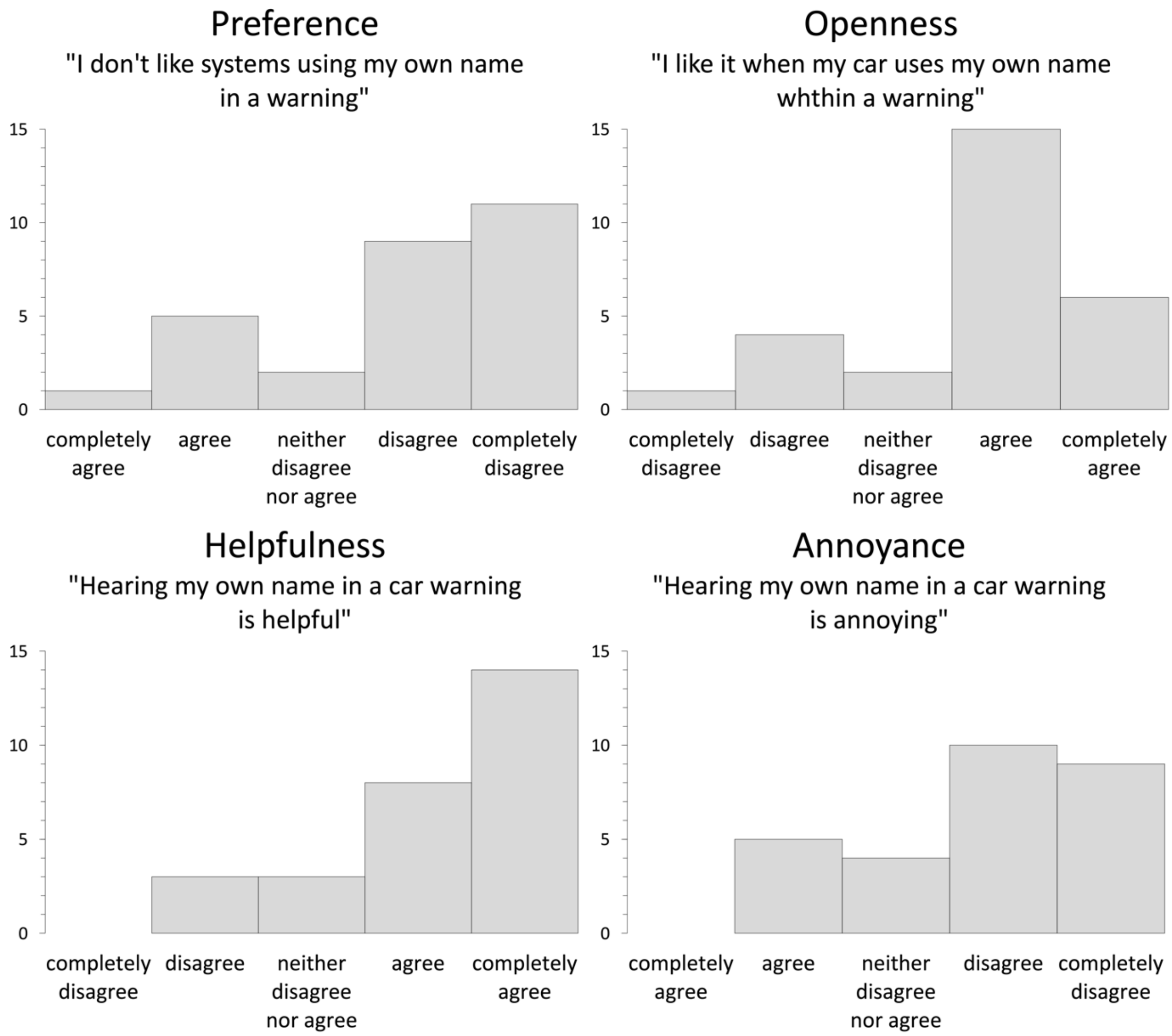

Figure 6. Frequency distributions of the subjective ratings of the name condition.

Note. For the “like” and “annoying,” the labels are presented in inverse order from how they were delivered to participants to have a consistent interpretation of all figures: values towards the right reflect a more positive evaluation.

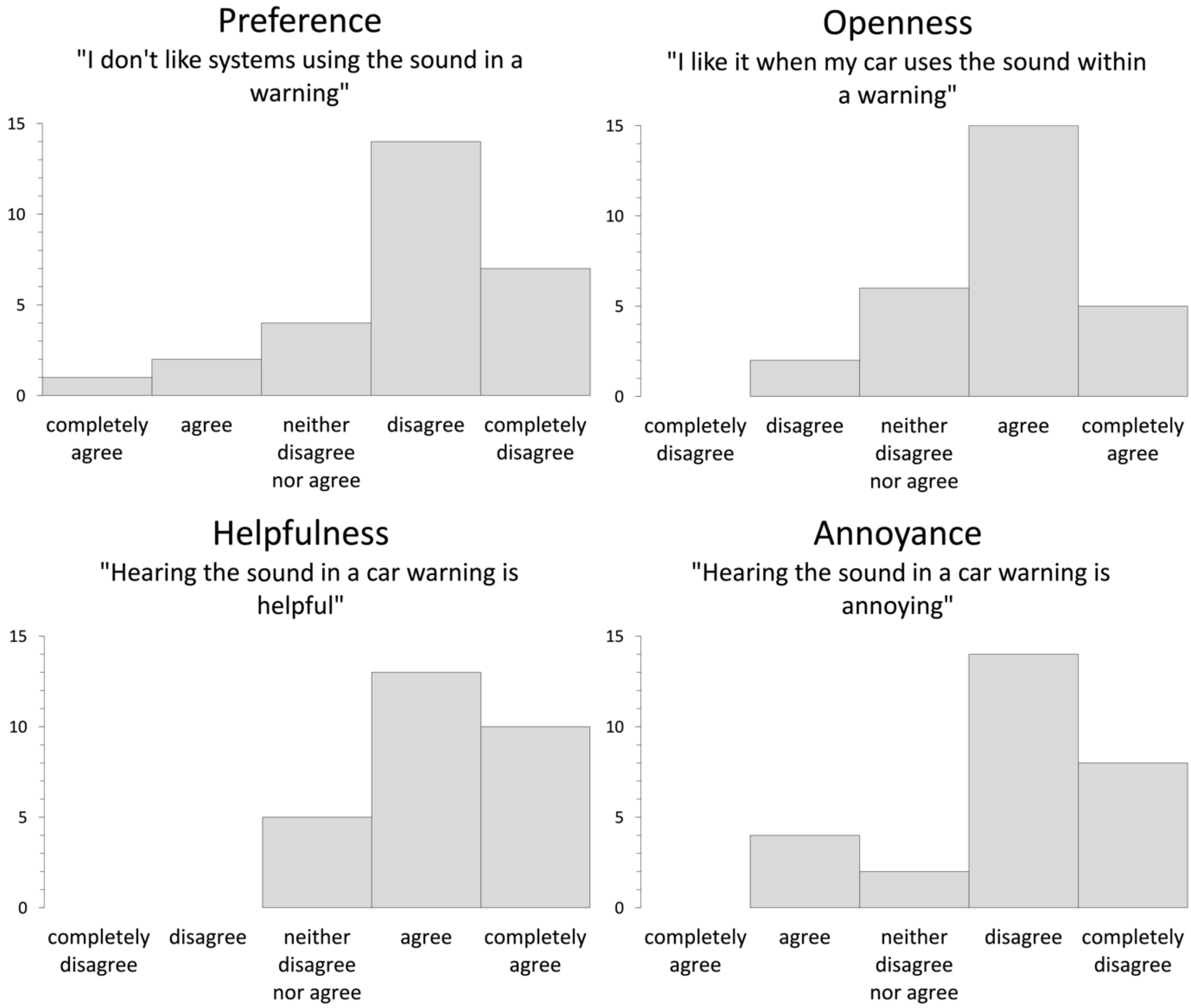

Figure 7. Frequency distributions of the subjective ratings of the sound condition.

Note. Same as above figure.

3.2.2 Subjective performance

3.2.2.1 General subjective performance

In the open question Q25, 17 out of 28 participants rated their performance positively (e.g., "*very nice*", "*quite well*", "*very well*", "*very nice*", "*good*", "*fine*", "*alright*", "*okay*", "*not bad*", "*above average*", "*managed to response all*"). 2 out of 28 gave a more neutral assessment (e.g., "*moderate", "average*"). 5 out of 28 rated their performance negatively (e.g., "*not so good*", "*not really well*", "*not so well*", "*not good*"). One participant (P21) thought they performed differently regarding different takeover requirements: "*good with braking, not so good with the turning left/right because sometimes I was trying to anticipate what to do before hearing the full instructions."* The other 3 gave no or irrelevant answers.

3.2.2.2 Subjective performance in different conditions

When asked about subjective performance in the different conditions (Q26 for the name condition, Q27 for the sound condition), 15 out of 28 gave explicit answers that they thought they performed better in the name condition than in the sound condition. For example, P08 wrote in Q26 that "*It is better than the warning tone because I suddenly put my attention to the sound*" whereas wrote in Q27 that "*Good, but I prefer the system to mention my name*"; P23 wrote for the name condition: "*a lot better, when I heard my own name I reacted faster I think*" while wrote for the sound condition: "*not as good as with hearing my own name, so not so well*". In contrast, 3 participants valued their performance worse in the name condition than in the sound condition, for example, P06 wrote in Q26 that "*not good, I only concerned my name rather than the other cars on the road"* whereas wrote in Q27: *"good, it can help me take action in time*". Besides, 4 out of 28 gave statements that they thought they performed equally in both conditions; for example, P2 wrote in both Q26 and Q27: "*I really do not think there is any difference."* The remaining 6 participants did not value their performances as better or worse in either condition but only gave answers such as "*well*" and "*fine."* The frequency distribution of participants' subjective comparison can be found in Figure 8.

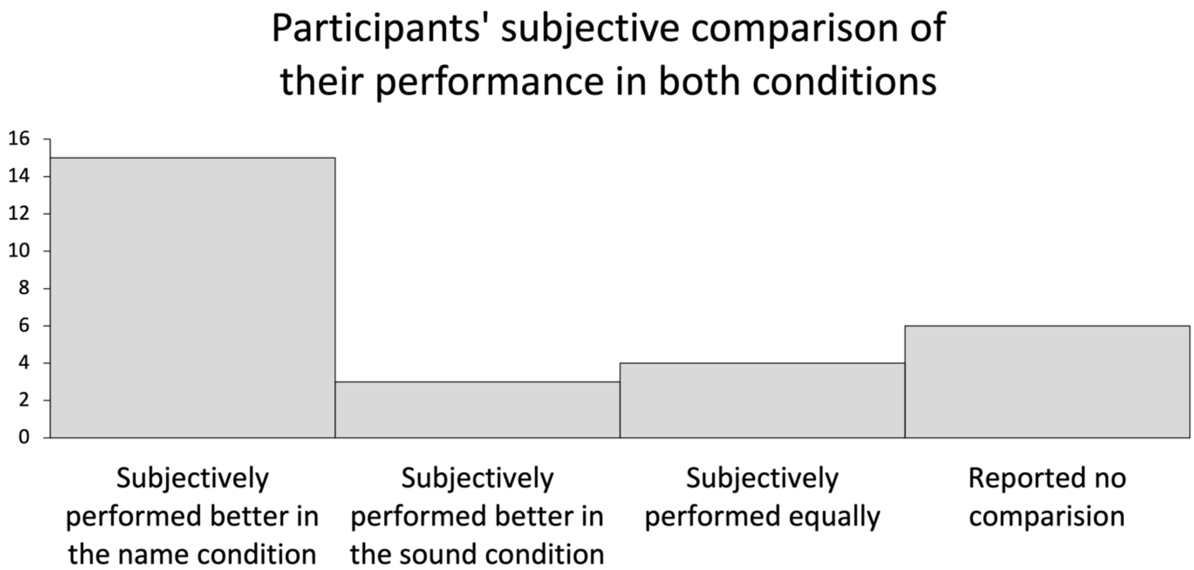

Figure 8. Frequency distribution of participants' subjective comparison of their performance in both conditions.

3.2.2.3 Thematic analysis

Through the thematic analysis of the open question feedback, the first emerged theme was a subjective faster attention-getting effect. 13 participants mentioned that hearing their names can effectively divert their attention back to driving. For example, P20 wrote about the name condition: "*I think I could pay more attention, so the performance might be better.*" P27 wrote about the name condition: "*Even better. It really alerted me when my name was used. Got my attention quicker from playing the game*". Likewise, P08 mentioned: "*It is better than the warning tone because I suddenly put my attention to the sound*".

Another emerging theme was the importance of pronunciation (suggested by 6 participants). 4 out of 6 expressed their appreciation of the well-pronounced names. For example, P15 wrote: "*my name was pronounced correctly so that was nice*"; P25 wrote: "*The pronunciation was good. I think that is extremely important!*". The remaining 2 participants indicated that their names were not pronounced accurately. For example, P12 wrote: "*My name was pronounced not fully corrected by the system, so sometimes I felt like that was funny which may affect how fast I respond to the messages*".

Another emerging theme was the negative emotions associated with hearing personal names (mentioned by 5 participants). 3 out of 5 indicated that names might be annoying. For example, P10 suggested: "*It increases my awareness yet annoying, maybe better not for all warnings.*" One participant (P23) mentioned that hearing names might cause stress: "*It helped to grab my attention faster, but it also stressed me more.*" The remaining 2 shared that hearing their names triggered other memories. For example, P06 noted: "*I don't think it's a good idea... It like the teacher calls my name*".

Table 7. *Description of emerging themes.*

Themes	Significant statement examples	n	%
Attention-getting effect	P11: "*I think it is helpful because I can immediately pay attention to the instruction after hearing my name.*"	13	46.4%
Importance of pronunciations	P25: "*The pronunciation was good. I think that is extremely important!*"	6	21.4%
Negative emotions (i.e., annoyance, stress, strangeness)	P10: "*It increases my awareness yet annoying, maybe better not for all warnings.*"	5	17.9%

4 Discussion

4.1 Discussion of results

In this study, I conducted a driving simulation experiment to assess the takeover performance and subjective experience when drivers heard their name or a warning sound in a speech alert. The results demonstrated that participants' responses were faster when the alert was preceded by their name compared to the warning sound (RQ1) (conform to Maidhof, 2020, but in contrast to Giessen, 2021). There was no difference in response accuracy (RQ1) (conform to Maidhof, 2020; Giessen, 2021). These results support the more general notion that the own name is a powerful attentional attractor (Cherry, 1953; Moray, 1959). The subjective experience was positive in the name condition (conform to Tobias, 2013; Maidhof, 2020; Giessen, 2021), and similarly positive in the sound condition (RQ2).

For the response time (RQ1), this study demonstrated that a personal name could trigger a faster response to required actions, which to some extent, replicated Maidhof (2020)'s results. However, it should be noticed that our results were derived from different measures of RT. In Maidhof (2020)'s study, RT was calculated from the onset of an action keyword (e.g., *brake*) to an action made, which was the prerequisite of how the difference was found. However, when the onset of RT was chosen as the name itself (as the current study), Maidhof did not find a difference. Such inconsistency might be partly due to the different lengths of the speech commands we used. The longer and more varied messages used in Maidhof (2020)'s study might have resulted in a larger standard deviation, making it harder to detect the differences statistically. In contrast, the speech commands in the current study were significantly shortened and uniformized, which was the most likely reason why the significance surfaced here. Further study would need to balance the speech length to draw a more robust conclusion. Moreover, the current result contradicts Giessen (2021), which found no difference between name and sound. The opposition between us is also likely due to shortened speech commands. As reported by Giessen, the two-sentence-long speech messages might have given participants too much time to examine their surroundings and develop situational awareness, which might partly shrink the effect of the name or the sound itself. In addition, the current results also partially conform to Tobias et al. (2013)'s results which showed a slightly faster response time to a personal name compared to a standard warning tone.

For the response accuracy (RQ1), similar to Maidhof (2020) and Giessen (2021), this study did not find a difference between the own name and control conditions: accuracy was high in both, which revealed that participants had no or at least minimal problem following the speech commands. Considering that the commands we used were relatively straightforward, it isn't surprising that participants performed highly accurately. Each alert had a scenario command (explaining the scenario) and an executive command (instructing what to do), making it very clear and easily understood. Another possible explanation for the high accuracy might lie in the non-driving task (i.e., playing the mobile game Fruit Ninja 2). This mobile game might be too easy to overload participants' cognitive processes. Participants might therefore interleave their attention between the non-driving and driving tasks. Moreover, The turn-based playing mode (one turn per minute) might also result in natural breakpoints, which gave the participants opportune moments to reorient attention to the driving circumstance. Those factors might to some extent have improved participants' performance. It could be hypothesized that, when the non-driving tasks become more demanding and continuous, people's auditory susceptibility might become lower (van der Heiden et al., 2021), then the benefits of having a name in an alert might show even stronger, as the attention-grabbing effect (Moray, 1959) might help to return attention faster back before a transition is needed. Future work can explore how varying task loads of non-driving activities affect participants' take-over accuracy.

In terms of subjective experience (RQ2), the current study replicated the previous results that participants were positive about using their own names for alerting purposes (Giessen, 2021; Maidhof, 2020; Tobias et al., 2013). However, such positive experiences did not differ between conditions (i.e., name versus sound) in any of the tested dimensions (i.e., preference, openness, helpfulness, annoyance)(Maidhof and Giessen did not report those comparisons). Further research is needed to identify the cause of these null differences. I suspect this is partly due to the careful selection of the sound used for the control condition. Compared to Giessen (2021)'s study which used a pure tone, I selected a nearly realistic car warning sound that people are relatively familiar with in their everyday driving. The frequency and loudness have also been well adjusted to make it more psychologically pleasant. All these operationalizations might have contributed to higher subjective ratings. However, though the Likert scale ratings seemed similar in both conditions, the thematic analysis reveals that most of the participants considered their performance to be better when hearing their names compared to hearing the sound, and subjectively thought the personal names strongly grabbed their attention. Putting it all together, the results suggest that, though the name and the warning sound were both subjectively accepted, participants performed both subjectively and objectively better in responding to transitions when hearing their names.

4.2 Implications

Given the above results, the primary theoretical implication of this study is extending the body of empirical evidence for a fundamental insight from cognitive science, that the personal name is a salient attractor of human attention (Cherry, 1953; Moray, 1959), and possesses the unique ability to evoke more rapid responses compared to other auditory signals (e.g., a warning sound) (Lucas, 1994). The second theoretical implication is developing a new research direction for interruption and multitasking research in dynamic environments (e.g., semi-automated driving). Previous research in this domain has frequently outlined ways of initiating proactive interruptions from the perspective of presentation modality (e.g., Petermeijer et al. 2017), the timing of the alert (e.g., van der Heiden et al., 2017), and the reliability of the alert (e.g., Wickens et al., 2009). However, little is known about how proactive interruptions should be designed as spoken interactions, in terms of content and delivery (Edwards et al., 2021). The current work retested the novel idea pioneered by Maidhof (2020) of alerting transitions by "calling a name", which might imply a piece of thread that an immediate interruption can also be effectively initiated by social norms and rituals from human-human interaction, and therefore illuminate more possible human-automation interaction strategies.

Regarding the implications for practice, this study provides a possible solution to customize and enhances the user experience for future (semi-) automated cars. Participants in the study were observed to be highly open and favorable when they heard their names in the alerts, which implied that personal names could be registered and used by the built-in speech agent system. However, the pronunciation should be treated with special caution. Indeed, many participants in this study emphasized the importance of the accuracy of the name pronunciation, and some of them reported unpleasant feelings when their names were not pronounced correctly. Moreover, some people might even object to hearing their names in cars even if their names are well pronounced, as a name might trigger annoyance, stress, and some episodic memories that are irrelevant to the context (e.g., the experience of being called by a teacher, as the P06 reported). Although these negative effects were only observed in a small number of participants, they might still deserve attention. Overall, all these findings might instruct future automation designers and developers to exercise more caution when using high self-relevant information of their users; for example, they probably should make sure that the system allows for an acoustic try-out of the audio file, the adjustment of voice parameters, language origins (Maidhof, 2020), and more personalized options.

4.3 Limitations and future work

The current study only analyzed a more traditional response time, the time consumed to reengage a physical control. However, it is noteworthy that the physical response might not be entirely enough to measure the success of a control transition (Frison et al., 2019). According to the transition framework of Janssen et al. (2019), there are other nuanced responses between an alert and actual contribution to driving, including disengaging from the original non-driving task, orienting to the driving task, and physical transfer of control. Then the next possible research question might be: in which stages can the personal name play its maximum alerting effect? To answer the question, further study should compensate more time-based measures, such as (1) the time when participants disengage from (or suspend) the non-driving task; (2) the time that is needed to orient the traffic situation, and possibly (3) the time that is needed to accomplish the take-over actions (i.e., from the onset of physical contribution until it is stopped). For unpacking these more granular levels of user behaviors, more complex non-driving tasks, and more sophisticated measures (e.g., gaze pattern data from eye tracker) might be needed.

Second, the driving task used in this study might be oversimplified. This study was conducted in a low-fidelity driving simulator, where participants only had to press a button on the steering wheel or press the brake pedal (which was similar to previous studies by Wong et al., 2019; Maidhof, 2020; Giessen, 2021). However, under real (semi-) automated driving, drivers usually need to take over more complex tasks and need more time to make an appropriate decision to avoid safety threats. Therefore, the too-simple driving task here (or, in some sense, the lack of natural driving tasks) might have resulted in the participants feeling less urgently, which probably limited the alerting effect of the stimuli. Moreover, this study displayed driving videos, which could just be played as preset and not provide just-in-time feedback corresponding to participants' actions on the gamepad. For example, if participants responded incorrectly, the video could not give any error indication. This might have also influenced the ecological validity of this study. Future studies may reexamine the effect of the stimuli by involving higher-complexity transition tasks (e.g., steering wheel movements, multiple decisions on route alternatives, task combinations) and involved with more perceptible feedback for indicating the consequences of the responses. This might call for higher fidelity driving simulator set-ups that enable a more realistic driving experience.

Another potential limitation is the participants. The participants in this study were from varied countries and almost all (27/28) were not native English speakers. This might not align very well with the speech commands pronounced in an English voice. Though a foreign English accent does not seem to reduce the intelligibility of speech (Munro, 2008; Munro & Derwing, 1995; Smith & Rafiqzad, 1979), according to the similarity-attraction effect (Byrne et al., 1967), people might still tend to trust speakers similar to them, such as similar accent and ethnic background (Nass & Brave, 2005; Reeves & Nass, 1996). For example, drivers from the UK might prefer a British accent, and those from the US might prefer an American accident (Bazilinskyy & de Winter, 2017). Therefore, it is unclear yet to what extent the cultural differences within our cross-national samples influenced our results. One possible question left for future studies is whether a speech alert should be exactly tailored to the language and accent of the participants' own country. Finally, the recruited participants in this study were all relatively young drivers (all under the age of 30). Future studies should consider middle-aged or older drivers to enhance the generalization of the results.

5 Conclusion

This study demonstrates that speech alerts starting with a driver's personal name can reduce the response time to control transitions compared to the alerts starting with a warning sound, and demonstrates that people have overall positive attitudes toward being alerted by their names. These results suggest that a personal name can be considered an effective design element in auditory alerting systems of future (semi-) automated cars that potentially contribute to better driving safety and experience.

References

[1] Abe, G., & Richardson, J. (2004). The effect of alarm timing on driver behavior: an investigation of differences in driver trust and response to alarms according to alarm timing. Transportation Research Part F: Traffic Psychology and Behavior, 7(4-5), 307-322.

[2] Abe, G., & Richardson, J. (2005). The influence of alarm timing on braking response and driver trust in low speed driving. Safety Science, 43(9), 639-654.

[3] Almén, L. (2002). Comparing audio and tactile inputs as driver attention control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[4] Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. Journal of Personality and Social Psychology, 37(5), 715.

[5] Arrabito, G. R. (2009). Effects of talker sex and voice style of verbal cockpit warnings on performance. Hum Factors, 51(1), 3-20.

[6] Atchley, P., & Chan, M. (2011). Potential benefits and costs of concurrent task engagement to maintain vigilance: A driving simulator investigation. Human Factors, 53(1), 3-12.

[7] Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. Theoretical Issues in Ergonomics Science, 8(4), 321-348.

[8] Baldwin, C. L. (2011). Verbal collision avoidance messages during simulated driving: Perceived urgency, alerting effectiveness and annoyance. Ergonomics, 54(4), 328-337.

[9] Baldwin, C. L., & Moore, C. (2002). Perceived urgency, alerting effectiveness and annoyance of verbal collision avoidance system messages. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[10] Barthes, R.(1977). Image, music, text.(S. Heath, Ed.) The Journal of Aesthetics and Art Criticism, 37, 220.

[11] Bazilinskyy, P., & de Winter, J. C. (2015). Auditory interfaces in automated driving: an international survey. PeerJ Computer Science, 1, e13.

[12] Bazilinskyy, P., & de Winter, J. C. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. Applied Ergonomics, 64, 56-64.

[13] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. Human–Computer Interaction, 4(1), 11-44.

[14] Bliss, J. P., & Acton, S. A. (2003). Alarm mistrust in automobiles: How collision alarm reliability affects driving. Applied Ergonomics, 34(6), 499-509.

[15] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

[16] Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. Journal of personality and social psychology, 5(1), 82.

[17] Cabrera, D., & Ferguson, S. (2006). Considerations arising from the development of auditory alerts for air traffic control consoles. Georgia Institute of Technology.

[18] Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? Human Factors, 54(5), 747-761.

[19] Chae, J., & Kang, Y. (2021). Designing an experiment to measure the alert fatigue of different alarm sounds using the physiological signals. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction.

[20] Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. The Journal of the Acoustical Society of America, 25(5), 975-979.

[21] Clark, H., McLaughlin, A. C., Williams, B., & Feng, J. (2017). Performance in takeover and characteristics of non-driving related tasks during highly automated driving in younger and older drivers. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[22] Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. Psychonomic Bulletin & Review, 8(2), 331-335.

[23] Dahlbäck, N., Wang, Q., Nass, C., & Alwin, J. (2007). Similarity is more important than expertise: Accent effects in speech interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[24] de Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. Transportation Research Part F: Traffic Psychology and Behavior, 27, 196-217.

[25] Di Stasi, L. L., Contreras, D., Cañas, J. J., Cándido, A., Maldonado, A., & Catena, A. (2010). The consequences of unexpected emotional sounds on driving behavior in risky situations. Safety Science, 48(10), 1463-1468.

[26] Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proceedings of the National Academy of Sciences, 113(10), 2636-2641.

[27] Edwards, J., Janssen, C., Gould, S., & Cowan, B. R. (2021). Eliciting spoken interruptions to inform proactive speech agent design. CUI 2021-3rd Conference on Conversational User Interfaces.

[28] Edworthy, J., Hellier, E., Walters, K., Clift-Mathews, W., & Crowther, M. (2003). Acoustic, semantic and phonetic influences in spoken warning signal words. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17(8), 915-933.

[29] Eichelberger, A. H., & McCartt, A. T. (2014). Volvo drivers' experiences with advanced crash avoidance and related technologies. Traffic Injury Prevention, 15(2), 187-195.

[30] Farmer, G. D., Janssen, C. P., Nguyen, A. T., & Brumby, D. P. (2018). Dividing attention between tasks: Testing whether explicit payoff functions elicit optimal dual-task performance. Cognitive Science, 42(3), 820-849.

[31] Fischer, C., Dailler, F., & Morlet, D. (2008). Novelty P3 elicited by the subject's own name in comatose patients. Clinical Neurophysiology, 119(10), 2224-2230.

[32] Fischer, C., Luaute, J., & Morlet, D. (2010). Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. Clinical Neurophysiology, 121(7), 1032-1042.

[33] Folmer, R. L., & Yingling, C. D. (1997). Auditory P3 responses to name stimuli. Brain and Language, 56(2), 306-311.

[34] Friedenberg, J., Silverman, G., & Spivey, M. J. (2021). Cognitive science: An introduction to the study of mind. Sage Publications.

[35] Frison, A.-K., Wintersberger, P., Schartmüller, C., & Riener, A. (2019). The real T (h) OR: Evaluation of emergency take-over on a test track. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings.

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

[37] Gerber, M. A., Schroeter, R., Xiaomeng, L., & Elhenawy, M. (2020). Self-Interruptions of non-driving related tasks in automated vehicles: Mobile vs head-up display. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.

[38] Giessen, I. v. d. (2021). Spoken auditory alerts in semi-automated vehicles: Using personal names for attracting attention.

[39] Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime–performing non-driving related tasks while driving highly automated. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[40] 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(1), 1938-1942.

[41] Gold, C., Korber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. Hum Factors, 58(4), 642-652.

[42] Gold, C., Lorenz, L., & Bengler, K. (2014). Influence of automated brake application on take-over situations in highly automated driving scenarios. Proceedings of the FISITA 2014 World Automotive Congress.

[43] Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. Journal of Experimental Social Psychology, 40(2), 216-224.

[44] Hancock, P. A. (2017). Driven to distraction and back again. In Driver Distraction and Inattention (pp. 9-26). CRC Press.

[45] Hasenjäger, M., & Wersing, H. (2017). Personalization in advanced driver assistance systems and autonomous vehicles: A review. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (itsc).

[46] Held, L., & Ott, M. (2018). On p-values and Bayes factors. Annual Review of Statistics and Its Application, 5(1), 593-419.

[47] Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. Human Factors, 44(1), 1-17.

[48] Holeckova, I., Fischer, C., Giard, M.-H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. Brain Research, 1082(1), 142-152.

[49] Holeckova, I., Fischer, C., Morlet, D., Delpuech, C., Costes, N., & Mauguière, F. (2008). Subject's own name as a novel in a MMN design: a combined ERP and PET study. Brain Research, 1189, 152-165.

[50] Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57(2), P101-P115.

[51] Iqbal, S. T., Horvitz, E., Ju, Y.-C., & Mathews, E. (2011). Hang on a sec! Effects of proactive mediation of phone conversations while driving. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[52] Isaacs, E., Edmonds, C., Lucas, A., & Gadian, D. (2001). Calculation difficulties in children of very low birthweight: a neural correlate. Brain, 124(9), 1701-1707.

[53] Janssen, C. P., Iqbal, S. T., Kun, A. L., & Donker, S. F. (2019). Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. International Journal of Human-Computer Studies, 130, 221-233.

[54] Kay, M., Nelson, G. L., & Hekler, E. B. (2016). Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.

[55] Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. New England Journal of Medicine, 370(1), 54-59.

[56] Köhn, T., Gottlieb, M., Schermann, M., & Krcmar, H. (2019). Improving take-over quality in automated driving by interrupting non-driving tasks Proceedings of the 24th International Conference on Intelligent User Interfaces.

[57] Lee, M. D., & Wagenmakers, E.-J. (2014). Bayesian cognitive modeling: A practical course. Cambridge university press.

[58] Lee, S. C., Yoon, S. H., & Ji, Y. G. (2021). Effects of non-driving-related task attributes on takeover quality in automated vehicles. International Journal of Human–Computer Interaction, 37(3), 211-219.

[59] Llaneras, R. E., Salinger, J., & Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

[60] Lucas, P. A. (1994). An evaluation of the communicative ability of auditory icons and earcons. An evaluation of the communicative ability of auditory icons and earcons. Georgis Institute of Technology.

[61] Maidhof, C. (2020). Car, call me by my name: Effectiveness of using the driver´ s own name for in-car voice alerts.

[62] Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. Psychological Science, 6(5), 314-317.

[63] Mandel-Emer, D., Jusczyk, P. W., Houston, D., Seidl, A., Hollich, G., Johnson, E., & Jusczyk, A. (2003). What's in a name?: How infants respond to some familiar sound patterns. Jusczyk Lab Final Report: http://hincapie.

[64] Martens, M. H., & van den Beukel, A. P. (2013). The road to automated driving: Dual mode and human factors considerations. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013).

[65] McKeown, D. (2005). Candidates for within-vehicle auditory displays.

[66] Merat, N., Jamson, A. H., Lai, F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual: Driver behavior when resuming control from a highly automated vehicle. Transportation Research Part F: Traffic Psychology and Behavior, 27, 274-282.

[67] 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. 2015 IEEE 18th International Conference on Intelligent Transportation Systems.

[68] Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. Quarterly Journal of Experimental Psychology, 11(1), 56-60.

[69] Müller, H. M., & Kutas, M. (1996). What's in a name? Electrophysiological differences between spoken nouns, proper names and one's own name. NeuroReport, 8(1), 221-225.

[70] Munro, M. J. (2008). Foreign accent and speech intelligibility. Phonology and Second Language Acquisition, 5, 193-218.

[71] Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Language Learning, 45(1), 73-97.

[72] Nass, C., & Lee, K. M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[73] Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[74] Nass, C. I., & Brave, S. (2005). Wired for speech: How voice activates and advances the human-computer relationship. MIT press Cambridge.

[75] Neubauer, C., Matthews, G., & Saxby, D. (2014). Fatigue in the automated vehicle: do games and conversation distract or energize the driver? Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[76] Newman, R. S. (2005). The cocktail party effect in infants revisited: listening to one's name in noise. Developmental Psychology, 41(2), 352.

[77] Nitz, D. A. (2006). Tracking route progression in the posterior parietal cortex. Neuron, 49(5), 747-756.

[78] Noyes, J., Hellier, E., & Edworthy, J. (2006). Speech warnings: a review. Theoretical Issues in Ergonomics Science, 7(6), 551-571.

[79] Orth, D., Schömig, N., Mark, C., Jagiellowicz-Kaufmann, M., Kolossa, D., & Heckmann, M. (2017, September). Benefits of personalization in the context of a speech-based left-turn assistant. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 193-201.

[80] Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3), 381-410.

[81] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230-253.

[82] Perrin, F., García-Larrea, L., Mauguière, F., & Bastuji, H. (1999). A differential brain response to the subject's own name persists during sleep. Clinical Neurophysiology, 110(12), 2153-2164.

[83] Perrin, F., Schnakers, C., Schabus, M., Degueldre, C., Goldman, S., Brédart, S., Faymonville, M.-E., Lamy, M., Moonen, G., & Luxen, A. (2006). Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome. Archives of Neurology, 63(4), 562-569.

[84] Petermann-Stock, I., Hackenberg, L., Muhr, T., & Mergl, C. (2013). Wie lange braucht der Fahrer-Eine Analyse zu Ubernahmezeiten aus verschiede-nen Nebentatigkeiten wahrend einer hochautomatisierten Staufahrt. 6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren.

[85] Petermeijer, S., Bazilinskyy, P., Bengler, K., & de Winter, J. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. Applied Ergonomics, 62, 204-215.

[86] Pfleging, B., Rang, M., & Broy, N. (2016). Investigating user needs for non-driving-related activities during automated driving. Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia.

[87] Politis, I., Brewster, S., & Pollick, F. (2015). Language-based multimodal displays for the handover of control in autonomous cars. Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

[88] Portas, C. M., Bjorvatn, B., & Ursin, R. (2000). Serotonin and the sleep/wake cycle: special emphasis on microdialysis studies. Progress in Neurobiology, 60(1), 13-35.

[89] Pratt, H., Berlad, I., & Lavie, P. (1999). Oddball'event-related potentials and information processing during REM and non-REM sleep. Clinical Neurophysiology, 110(1), 53-61.

[90] Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[91] Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. Cambridge, UK, 10, 236605.

[92] Richie, E., Offer-Westort, T., Shankar, R., & Jeon, M. (2018). Auditory displays for take-over in semi-automated vehicles. International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management.

[93] SAE International (2016). Definitions for terms related to driving automation systems for on-road motor vehicles (J3016). Soc. Automot. Eng., Warrendale, PA, USA, Tech. Rep. J3016_201806.

[94] Šabić, E., Henning, D., & MacDonald, J. (2019). Adaptive auditory alerts for smart in-vehicle interfaces. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[95] Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. European Journal of Social Psychology, 8(4), 467-487.

[96] Schnakers, C., Perrin, F., Schabus, M., Majerus, S., Ledoux, D., Damas, P., Boly, M., Vanhaudenhuyse, A., Bruno, M.-A., & Moonen, G. (2008). Voluntary brain processing in disorders of consciousness. Neurology, 71(20), 1614-1620.

[97] Smith, G. A., Chounthirath, T., & Splaingard, M. (2019). Effectiveness of a voice smoke alarm using the child's name for sleeping children: A randomized trial. The Journal of Pediatrics, 205, 250-256.e251.

[98] Smith, G. A., Splaingard, M., Hayes, J. R., & Xiang, H. (2006). Comparison of a personalized parent voice smoke alarm with a conventional residential tone smoke alarm for awakening children. Pediatrics, 118(4), 1623-1632.

[99] Smith, L. E., & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. Tesol Quarterly, 371-380.

[100] Tobias, C., Su, C.-Y., Kolburg, L., & Lathrop, B. (2013). Cocktail party effect & attention capture in semi-autonomous driving. 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.

[101] Treisman, A. M., & Riley, J. G. (1969). Is selective attention selective perception or selective response? A further test. Journal of Experimental Psychology, 79(1p1), 27.

[102] van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.

[103] van der Heiden, R. M. A., Janssen, C. P., Donker, S. F., Hardeman, L. E. S., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. PloS One, 13(8), e0201963.

[104] van der Heiden, R. M. A., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond? Transportation Research Part F: Traffic Psychology and Behavior, 65, 748-759.

[105] Van der Heiden, R. M. A., Kenemans, J. L., Donker, S. F., & Janssen, C. P. (2021). The effect of cognitive load on auditory susceptibility during automated driving. Hum Factors, 18720821998850.

[106] Vogelpohl, T., Vollrath, M., Kühn, M., Hummel, T., & Gehlert, T. (2017). Übergabe von hochautomatisiertem Fahren zu manueller Steuerung-Teil 2. Unfallforschung der Versicherer.

[107] Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. Current Directions in Psychological Science, 25(3), 169-176.

[108] Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. Georgia Institute of Technology.

[109] Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a “cry wolf” effect? Human Factors, 51(4), 446-462.

[110] Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Based guidelines for warning design and evaluation. Applied Ergonomics, 33(3), 219-230.

[111] Wong, P. N. Y., Brumby, D. P., Babu, H. V. R., & Kobayashi, K. (2019). Voices in self-driving cars should be assertive to more quickly grab a distracted driver's attention. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

[112] Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(1), 255.

[113] Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science, 3(2), 178-194.

[114] Young, M. S., & Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. Ergonomics, 50(8), 1324-1339.

[115] Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. Transportation Research Part F: Traffic Psychology and Behavior, 64, 285-307.

Appendix A

Comparison of the length of the speech commands between the current study and Maidhof (2020)'s study

In this study, I shortened the length of the speech commands of Maidhof (2020), in order to minimize the potential impact of the overly long and varied messages on participant performance and to allow for more reliable comparability across conditions. To demonstrate that the commands were shortened, I analyzed the mean time duration by a Bayesian independent student t-test. The result showed strong evidence (BF_{10} =39984.540) supporting that the speech commands used in this study (M = 2.588s, SD = 0.157s) were shorter and less varied than Maidhof's (M = 4.225s, SD = 0.709s).

Table A1. *Speech commands used in the current study. Action signal words (i.e., left, right, brake, slow down) are in italics.*

Speech commands	Total Duration(s)
Beware of T-junction ahead. Indicate *right.*	2.800
Narrow road ahead. *Slow down.*	2.334
Car coming from the left. *Slow down.*	2.550
Car coming from the right. *Brake* right now.	2.517
Beware of red light. *Slow down.*	2.417
Exit roundabout now. Indicate *right.*	2.533
Moving to the left lane. Indicate *left.*	2.633
Beware of red light. *Brake* right now.	2.517
Moving to the right lane. Indicate *right.*	2.567
Beware of T-junction ahead. Indicate *left.*	2.833
Beware of T-junction ahead. *Slow down.*	2.767

Table A2. *Speech commands used in Maidhof (2020)'s study.*

Speech commands	Total Duration(s)
Beware of T-junction ahead. Look up! Action to indicate *right* is needed.	4.233
Beware of the car exiting on your left. You need to *slow down* right now.	4.033
Exiting roundabout ahead. Look up! Action to indicate *left* is needed.	4.267
Beware of the oncoming vehicles ahead. Watch out! *Brake* right now.	4.150
Two-way road being blocked on the side. Beware of oncoming vehicles ahead. You need to *slow down* right now.	6.150
Beware of the oncoming vehicles ahead. You need to *slow down* right now.	4.067
Narrow road ahead. Beware of the oncoming vehicles. You need to *slow down* right now.	4.483
Beware of T-junction ahead. Look up! Action to indicate *left* is needed.	4.233
Red traffic light coming up. Watch out! *Brake* right now.	3.383
Moving towards the right lane. Look Up! Action to indicate *right* is needed.	4.217
Moving towards the left lane. Look Up! Action to indicate *left* is needed.	4.250
There is traffic coming up. You need to *slow down* right now.	3.233

Appendix B

Video orders used for participants

Table B. *Video orders used for participants.*

Participant	1st Video	2nd Video	3rd Video	4th Video
P01	N-Video A	S-Video B	N-Video D	S-Video C
P02	N-Video A	S-Video C	N-Video B	S-Video D
P03	N-Video D	S-Video B	N-Video A	S-Video C
P04	N-Video C	S-Video B	N-Video D	S-Video A
P05	S-Video C	N-Video D	S-Video B	N-Video A
P06	S-Video D	N-Video B	S-Video C	N-Video A
P07	S-Video A	N-Video D	S-Video B	N-Video C
P08	N-Video A	S-Video C	N-Video D	S-Video B
P09	N-Video B	S-Video D	N-Video A	S-Video C
P10	S-Video B	N-Video D	S-Video C	N-Video A
P11	N-Video C	S-Video A	N-Video B	S-Video D
P12	N-Video B	S-Video D	N-Video C	S-Video A
P13	S-Video D	N-Video B	S-Video A	N-Video C
P14	S-Video A	N-Video C	S-Video D	N-Video B
P15	S-Video A	N-Video C	S-Video B	N-Video D
P16	N-Video D	S-Video B	N-Video C	S-Video A
P17	S-Video B	N-Video A	S-Video C	N-Video D
P18	N-Video D	S-Video A	N-Video C	S-Video B
P19	S-Video A	N-Video B	S-Video D	N-Video C
P20	N-Video C	S-Video D	N-Video B	S-Video A
P21	S-Video B	N-Video C	S-Video A	N-Video D
P22	N-Video A	S-Video D	N-Video B	S-Video C
P23	S-Video C	N-Video B	S-Video D	N-Video A
P24	N-Video D	S-Video C	N-Video A	S-Video B
P25	N-Video B	S-Video A	N-Video C	S-Video D
P26	S-Video C	N-Video A	S-Video D	N-Video B
P27	S-Video D	N-Video C	S-Video A	N-Video B
P28	S-Video C	N-Video A	S-Video B	N-Video D
	S-Video D	N-Video A	S-Video C	N-Video B
	N-Video B	S-Video C	N-Video A	S-Video D
	N-Video C	S-Video A	N-Video D	S-Video B
	S-Video B	N-Video D	S-Video A	N-Video C

Note. N=Name condition, S= Sound condition.

Appendix C

All Questionnaire items

The questionnaire was administered digitally using a digital tool (Qualitrics, www. qualtricsm.com). The final digital version of the questionnaire can be found at https://survey.uu.nl/jfe/form/SV_6EUXaA66PGrHf8y. Below I only provide the questions, corresponding variables, and the used coding method.

Table C. *All Questionnaire items.*

Section	Variable	Question	Full question displayed	Used coding
1	Gender	Q1	What's your gender?	−1 = Prefer not to respond, 1 = Female, 2 = Male, 3=Non-binary/third gender
	Age	Q2	How old are you?	Positive integer value
	Nationality	Q3	What's your nationaliy?	Textual response
	Language	Q4	Is English your native language?	1=Yes, 2=No
	Vision	Q5	Is your vision nomal or corrected?	1=Normal, 2= Corrected (glasses or prescription lenses)
	Hearing	Q6	Do you suffer from hearing problems	1=Yes, 2=No
		Q7	(If yes is selected) What kind of hearing problem do you have?	Textual response
	Driver qualification	Q8	Do you have a driving licence for a car?	1=Yes, 2=No
	Driver seniority	Q9	For how long do you have the driving license?	1 = 1 year or less, 2= 1-2 years, 3= 2-3 years, 4=3-4 years, 5= 4-5 years, 6= More than 5 years
	Drive frequency	Q10	On average how frequently do you drive a car or other motorised vehicles?	1 = Couple time a week, 2= weekly, 3= couple times a month, 4=monthlu, 5= never
	Drive location	Q11	In which contry do you drive most?	Textual response

	Education	Q12	Are you cuttently a student?	1=Yes, 2=No
		Q13	What is your last diploma?	1=PhD,2=Master's degree, 3=Bachelor's degree, 4=High school degree, 5= secondary education diploma, 6=Other
	Experience to the non-driving task	Q14	Have you ever played the mobile game "Fruit Ninja" ? (or a very similar mobile game)?	1=Yes, 2=No
		Q15	(If yes is selected) What kind of hearing problem do you have?	1=once, 2=2-6 times, 3= 7-11 times, 4=12-16 times, 5=more than 16 times
2	Perefernce (to name)	Q16	I don't like systems using my own name in a warning	1=Completely agree, 2=agree, 3= neither disagree nor agree, 4=disagress, 5= completely disagree
	Openness (to name)	Q17	I like it when my car uses my own name whthin a warning	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree
	Helpfulness (to name)	Q18	Hearing my own name in a car warning is helpful	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree
	Annoyance (to name)	Q19	Hearing my own name in a car warning is annoying	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree
	Concern (to name)	Q20	How concerned are you about systems recording such personal information about you?	1=very concrened, 2=concerned, 3= neutral, 4=not concerned, 5= not concerned at all
	Perefernce (to sound)	Q21	I don't like systems using the tone in a warning	1=Completely agree, 2=agree, 3= neither disagree nor agree,

				4=disagress, 5= completely disagree
	Openness (to sound)	Q22	I like it when my car uses the tone whthin a warning	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree
	Helpfulness (to sound)	Q23	Hearing the tone in a car warning is helpful	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree
	Annoyance (to sound)	Q24	Hearing the tone in a car warning is annoying	1=Completely disagree, 2=disagree, 3= neither disagree nor agree, 4=agree, 5= completely agree
3	General performance	Q25	In general, how well do you think you performed in the driving task?	Textual response
	Subjective Performance in name condition	Q26	How well do you think you performed in the driving task after hearing a warning that starts with your own name?	Textual response
	Subjective Performance in sound condition	Q27	How well do you think you performed in the driving task after hearing a warning that starts with a tone?	Textual response
	Subjective opinion about using personal name	Q28	Do you have any remarks about warnings in cars using your own name?	Textual response
	Subjective opinion about the experiment	Q29	Do you have any further remarks about the experiment?	Textual response

Appendix D

Additional analysis of response time: What if response type (hands or foot) is taken into account?

I conducted paired t-tests to further compare the RT between conditions under different response types: the responses with the hands (signaling left or right) or with the foot (slowing down or braking). When designing the video stimuli I counterbalanced the frequency of each action type between conditions (For each participant per condition, 4 times of signaling direction and 8 times of braking were executed in total, see Table 1), which enabled paired t-tests to be performed.

RT when the braking is needed

The RT (the mean of median RT, M_{med}) to braking is faster in the name condition (M_{med} = 3.081s, SD_{med} = 0.505s, 95% CI = [2.885s, 3.277s]) compared to the control condition with the warning sound (M_{med} = 3.273s, SD_{med} = 0.414s, 95% CI = [3.113s, 3.434s]). A Bayesian paired t-test found strong evidence for this difference, BF_{10} = 16.089.

RT when the signaling direction is needed

The RT (the mean of median RT, M_{med}) to signaling direction in the name condition (M_{med} = 3.856s, SD_{med} = 0.541s, 95% CI = [3.647s, 4.066s]) seems to be slightly faster compared to the control condition with the warning sound (M_{med} = 4.041s, SD_{med} = 0.542s, 95% CI = [3.831s, 4.251s]). However, a Bayesian paired t-test found only anecdotal evidence for this difference, BF_{10} = 0.475.

Appendix E

Alternative measure of response time: What if the onset of response time is chosen differently?

In the main thesis, I analyzed participants' response time relative to the onset of the name/sound and the onset of an action by the user (i.e., button press on the steering wheel or pressing the brake). This analysis was in line with many previous studies which involved audio signals during autonomous driving (e.g., van der Heiden et al., 2018; van der Heiden et al., 2019; van der Heiden et al., 2021). In this appendix, I also reported an alternative analysis when RT was measured as the interval between the onset of a signal word (i.e., left / right / brake/ slowdown) and an action. Here I named it RT_2. Note that when choosing to analyze RT_2, there were "logging errors" (i.e., participants sometimes gave their response before the signal word) (cf. Maidhof, 2020; Giessen, 2021). I adjusted the RT to 0s for these trials, considering these situations to represent "extremely fast" responses. To allow a more direct comparison with Maidhof's results, I conducted classical paired t-tests where RT_2 used both mean and median (M_{med}: Mean of median RT; M_{mean}: Mean of mean RT).

E1 Assessment of response time

Results when RT_2 using median. A classical paired t-test found no significant difference on RT_2 (t (27) = 1.312, p = .200) between the name condition (M_{med} = 0.985s, SD_{med} = 0.393s, 95% CI = [0.832s, 1.137s]) and the control condition (M_{med} = 1.055s, SD_{med} = 0.381s, 95% CI = [0.908s, 1.203s]).

Results when RT using mean. A classical paired t-test found no significant difference on RT_2 (t (27) = 1.047, p = .305) between the name condition (M_{mean} = 1.041s, SD_{mean}= 0.408s, 95% CI = [0.883s, 1.199s]) and the control condition (M_{mean} = 1.101s, SD_{mean}= 0.394s, 95% CI = [0.949s, 1.254s]).

E2 Assessment of response accuracy

I did not report "logging errors" (i.e., the response that was given earlier than signal words) in the main text because I used a different measure of RT (i.e., choosing the name/sound as the onset instead of signal words, cf. Maidhof, 2020; Giessen, 2021). All responses in the current study were given after the name/sound, so no “logging errors” occurred. I reported such data in this appendix to allow for a direct comparison with Maidhof (2020)’s and Giessen (2021)’s analysis.

As seen from the Tabel E1, logging errors in the current study (8.73% in the name condition, 6.25% in the sound condition) were generally more frequent than in Maidhof (2020)’s study (2.2% in the own-name condition and 0.7% in the random-name condition) but were similar to Giessen (2021)’s study (9.06% in the name, 8.33% in the tone). Most of the logging errors occurred in video C (19 logging errors) and video D (18 logging errors) (see Table E2). I hypothesized two possible reasons for this observation. First, which is consistent with Giessen (2021)’s hypothesis, there might be some scenarios or the speech commands in video C and video D that enabled participants to predict in advance what actions were required to be taken, which might have caused users to react so fast that prior to the signal words. Second, participants might respond faster in higher-speed driving, because video C and video D in this study presented a relatively higher-speed driving context (70km/h limit) compared to video A and video B (30 km/h limit). Further studies might be needed to validate these hypotheses.

Table E1. *Too early response trials per condition.*

Condition	Total number of trials (100%)	logging error (Too early response trials)
Name	332	29 (8.73%)
Tone	336	21 (6.25%)

Table E2. *Too early response trials per video.*

	Logging errors trial 1	Logging errors trial 2	Logging errors trial 3	Logging errors trial 4	Logging errors trial 5	Logging errors trial 6	Total
Video A	1	0	1	0	1	2	5
Video B	1	1	1	2	1	4	10
Video C	3	3	2	2	6	3	19
Video D	5	2	2	3	3	3	18

Appendix F

Analysis of duration of the names and the sound: Did the duration of the auditory cues differ between conditions?

In this supplementary file, I compared the duration of pronunciation of the names and the sound. A Bayesian one-sample t-test found anecdotal evidence for a difference (BF_{10} = 1.749) between the duration of names (N=28; M = 0.511s, SD = 0.09s) and the sound duration (0.550s), which means the duration of stimuli were comparable between conditions.

Table F. *Duration of the personal names of the participants.*

Number	name	Duration(s)	Number	Name	Duration(s)
P01	*	0.517	P15	*	0.500
P02	*	0.517	P16	*	0.583
P03	*	0.633	P17	*	0.500
P04	*	0.517	P18	*	0.450
P05	*	0.633	P19	*	0.583
P06	*	0.367	P20	*	0.467
P07	*	0.517	P21	*	0.717
P08	*	0.517	P22	*	0.333
P09	*	0.467	P23	*	0.583
P10	*	0.633	P24	*	0.500
P11	*	0.467	P25	*	0.367
P12	*	0.517	P26	*	0.350
P13	*	0.583	P27	*	0.533
P14	*	0.517	P28	*	0.450

Note. Personal names are replaced by asterisks for privacy consideration.

Appendix G

Concerns about using the name

In the final questionnaire, a separate dimension of subjective experience - *concern* (Q20) was designed specifically for the name condition, measuring how concerned participants were about systems calling their names. This was asked using the question: "How concerned are you about systems recording such personal information about you?" The scale had five anchors ("very concerned", "concerned", "neutral", "not concerned", and "not concerned at all").

Results are shown in Figure G below. Most participants (17/28) were not concerned, 5/28 were neutral, and 6/28 had concerns. Future studies would need to investigate where these concerns came from.

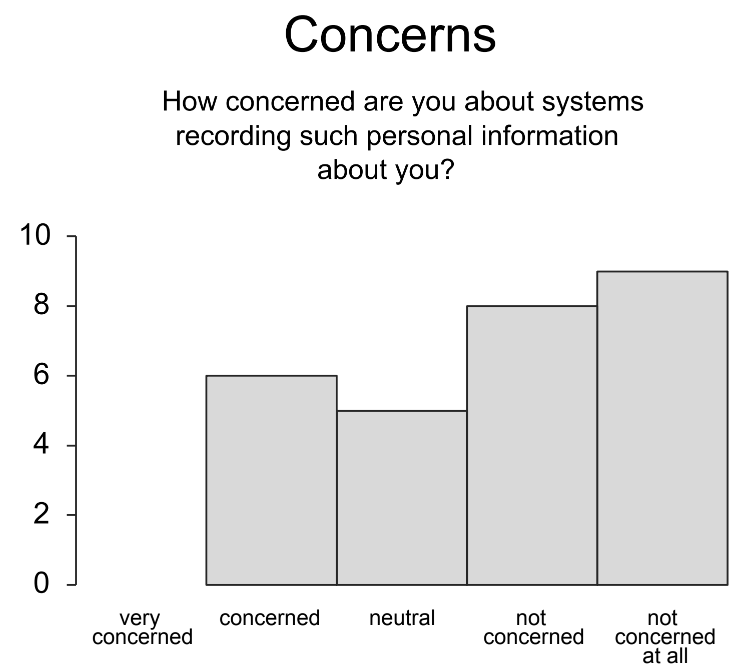

Figure G. Participants' concerns about the car using personal names.

Acknowledgments

It is with both excitement and frustration that I am finishing this thesis as the final delivery of my master trajectory. When I looked back at this document, I have to admit many shortcomings: the theoretical territory still needs further exploration, the experimentation and statistics might be too simple to be convincing, and some of the writing might be wordy and ambiguous. These deficiencies caused by my limited ability were often a source of frustration for me. But on the positive side, I believe this thesis at least presents a completed research process and gives a meaningful answer to the defined question, which might somewhat inform subsequent readers who are interested in the relevant field. Personally, I can also feel a significant self-improvement throughout this 5-month thesis period, in terms of not only the knowledge and skills but also the more determined motivation to do scientific research in the future, which I think, should be something worthy of a celebration.

I would like to thank my supervisor, Chris, for offering me the opportunity to run this exciting study, for his sufficient guidance, and for his always showing great patience and warmness to me - an English-limited international student and a complete newbie in psychology. Honestly, I felt very unconfident when I jumped into ACP from my previous major of industrial design, and even overwhelmed when I started this thesis. But fortunately, thanks to Chris's effective and kind supervision, the process was not as that tough as I expected. I enjoyed and felt highly enriched by every conversation with Chris. He provided his constructive comments, taught me the necessary skills at each stage of the research process, and more importantly, he inspired me to be a qualified scientific researcher with a critical, rigorous, and original spirit, which meant a lot to me.

I would also like want to thank my course coordinator, Stella, for helping me during my courses and internship; thank her for always being available for my all kinds of questions. Without Stella, I wouldn't have run so smoothly in accomplishing this master. Thanks to Caterina and Ilse for helping me with the technical puzzles. Thanks to the classmates I met at UU, Lan, Ziyu, and Wenkai, for working together with me to solve the daily challenges. Thanks to my wonderful friends, Jiahui, Jiaxing, and Huahua for maintaining a close emotional connection with me though we are far across the ocean. Thanks to my parents for funding my study and their constant support. And a final and special thanks to Zhuochao for giving me the best care and love, and always being by my side.

Jiaxin Xu

July 2022, in Utrecht

Curriculum Vitae

Jiaxin Xu was born on August 14th, 1996, in Hulunbuir, China. He studied industrial design for his Bachelor degree at Nanjing Agricultural University and graduated top first of his class in 2018. He was then recommended for admission to study in the Design Science Master program at Hunan University, with a first-class graduate student scholarship granted by Hunan University. During this master period, Jiaxin further developed his skills in human-centered design and design-oriented human factors research. In 2020, he was granted a China national graduate student scholarship. After receiving his Master degree in 2021, Jiaxin came to Utrecht University in the Netherlands to study applied cognitive psychology for his second Master degree.