



# Analyzing crowdsourced ratings of speech-based take-over requests for automated driving



P. Bazilinsky<sup>\*</sup>,<sup>1</sup>, J.C.F. de Winter<sup>1</sup>

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands

## ARTICLE INFO

### Article history:

Received 13 February 2017  
Received in revised form  
18 April 2017  
Accepted 2 May 2017

### Keywords:

Auditory displays  
Autonomous driving  
Speech perception  
Human-automation interaction

## ABSTRACT

Take-over requests in automated driving should fit the urgency of the traffic situation. The robustness of various published research findings on the valuations of speech-based warning messages is unclear. This research aimed to establish how people value speech-based take-over requests as a function of speech rate, background noise, spoken phrase, and speaker's gender and emotional tone. By means of crowdsourcing, 2669 participants from 95 countries listened to a random 10 out of 140 take-over requests, and rated each take-over request on urgency, commandingness, pleasantness, and ease of understanding. Our results replicate several published findings, in particular that an increase in speech rate results in a monotonic increase of perceived urgency. The female voice was easier to understand than a male voice when there was a high level of background noise, a finding that contradicts the literature. Moreover, a take-over request spoken with Indian accent was found to be easier to understand by participants from India than by participants from other countries. Our results replicate effects in the literature regarding speech-based warnings, and shed new light on effects of background noise, gender, and nationality. The results may have implications for the selection of appropriate take-over requests in automated driving. Additionally, our study demonstrates the promise of crowdsourcing for testing human factors and ergonomics theories with large sample sizes.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

### 1.1. Take-over requests

Until cars can drive autonomously, there will be situations where the driver has to resume manual control. Prior to such control transition, the automation may issue a take-over request to the driver (SAE International, 2016; Zeeb et al., 2015). How to provide a take-over request is a widely studied topic in human factors and ergonomics (Hergeth et al., 2015; Naujoks et al., 2014; Petermeijer et al., 2016; Pfromm et al., 2015).

A take-over request can be provided through pre-recorded voice (Gold et al., 2015; Mok et al., 2015; Politis et al., 2015), which may be an effective approach because humans are able to perceive

sounds irrespective of head or eye orientation (Bazilinsky and De Winter 2015). In aviation, a similar approach is used: traffic alert and collision avoidance systems (TCAS), which are mandatory in today's aircraft, apply voice commands (Kuchar and Yang, 2000).

Take-over situations may be of different urgency. Several studies have measured driver behavior in highly urgent situations, such as Mok et al. (2015), who found that 50% of the drivers veered off the road when a critical lane-closure event followed only 2 s after a take-over request ("Emergency, Automation off"). Other studies have been concerned with larger lead times of 5 or 7 s (Gold et al., 2013; see Eriksson and Stanton, 2017; for an overview) or with discretionary transitions having a low urgency (Damböck et al., 2013; Merat and Jamson, 2009; Nilsson et al., 2013). Politis et al. (2015) found that participants reacted 1.3 s faster to urgent take-over requests ("Danger! Collision imminent; You have control!") than to non-urgent ones (e.g., "Warning! GPS signal weak; Want to take over?"). In sum, how to convey the right sense of urgency is regarded as an important topic in automated driving research.

<sup>\*</sup> Corresponding author. Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

E-mail address: [p.bazilinsky@tudelft.nl](mailto:p.bazilinsky@tudelft.nl) (P. Bazilinsky).

<sup>1</sup> The authors contributed equally.

## 1.2. Speech warnings

Previous research has shown that semantics have an effect on urgency, in that a word such as ‘danger’ is perceived as more urgent than ‘attention’ (Arrabito, 2009; Baldwin, 2011; Wogalter and Silver, 1995; Wogalter et al., 2002). Second, emotional tone has important effects: phrases are considered more urgent if spoken in an urgently intoned style (Edworthy et al., 2003a; Ljungberg et al., 2012). Third, it has been found that the greater the speech rate, the higher the perceived urgency (Hollander and Wogalter, 2000; Jang, 2007; Park and Jang, 1999). No clear gender effects seem to exist: words spoken by a female typically yield similar urgency ratings as the same words spoken by a male (e.g., Hellier et al., 2002; Wogalter et al., 2002). However, Jang (2007) and Park and Jang (1999) found that a male voice yielded higher urgency ratings than a female voice. Furthermore, interaction effects have been observed, where the word “Note” received a higher urgency rating when spoken by a male instead of a female (Hellier et al., 2002). Differences in the degree of smoothness, pitch, and timbre may explain these gender differences (Edworthy et al., 2003a,b; Jang, 2007).

In addition to urgency, it is important to consider whether the message is comprehensible and pleasant. If people become displeased with a warning, they may ignore or disable the warning system, potentially causing unsafe situations (Eichelberger and McCart, 2014; Parasuraman and Riley, 1997). A female voice has been regarded as more pleasant (Bazilinskyy and De Winter 2015; Machado et al., 2012) and is more often used in route navigation devices (Large and Burnett, 2013) than a male voice. The female and male voice are supposedly equal in terms of intelligibility, but it has been reported that the male voice is easier to understand in a noisy environment such as an aircraft cockpit (Nixon et al., 1998; Noyes et al., 2006). However, it is unknown whether this effect is replicable. Arrabito (2009) stated that “further research is required to study the effects of speech parameters and word semantics across multiple talkers of each sex for variations of urgency under different background noise sources” (p. 18).

There is currently an irony in automated driving, because the technologies are deployed in the highest-income countries, which already have commendable road safety statistics, while low-income countries account for the vast majority of fatal road traffic accidents (Gururaj, 2008; World Health Organization, 2015). At present, car manufacturers are exploring cross-national perceptions of warnings (Langlois et al., 2008), but it is unknown whether speech-based take-over requests should be differentially developed per country. Research has shown that there are national differences in how people perform at basic visual perception tasks (Henrich et al., 2010). Regarding the appraisal of sounds, similar differences may exist. For example, it has been found that the sound of a bell was rated as pleasant among German listeners (possibly because it yielded connotations to a church bell), whereas this sound was rated as dangerous and unpleasant among Japanese listeners (Fastl, 2006). One specific question is whether a speech-based warning should be tailored to the language and accent of the host country. For example, it is possible that drivers from the UK prefer a British accent, and drivers from the US prefer an American accent. It has been found that a foreign English accent does not reduce the intelligibility and comprehensibility of speech (Munro and Derwing, 1995; Munro, 2008; Smith and Rafiqzad, 1979), but these findings deserve further investigation.

## 1.3. Aim of the study

This paper assesses how different speech-based take-over requests are perceived. Specifically, in line with the above research

gaps, we assessed (1) the effects of speech rate on perceived urgency, commandingness, pleasantness, and ease of understanding, for speakers that differ in gender and emotional tone. Additionally, we investigated (2) the effects of spoken phrase (semantic content) on perceived urgency for a male and female speaker, (3) the effects of noise on the ease of understanding, for a male and female speaker, and (4) the effect of participants’ (i.e., listeners’) gender on pleasantness. Finally, we explored (5) the relationship between the participants’ country and the ease of understanding of the messages. To acquire a large sample, we used crowdsourcing, an approach that is gaining popularity (Bazilinskyy and De Winter 2015; Behrend et al., 2011; Buhrmester et al., 2011; Crump et al., 2013; Kyriakidis et al. 2015; Rand, 2012).

## 2. Methods

This research was approved by the Human Research Ethics Committee at the TU Delft under the ethics approval application titled “Rating audio messages by means of crowdsourcing” on May 24, 2016. Informed consent was obtained from each participant via a dedicated survey item.

### 2.1. Speech-based messages

Speech-based messages “Take over, please” were created using the online tool Acapela-Box (<https://acapela-box.com>). Acapela-Box reproduces the natural sound of language based on voice of human speakers, and was selected because it offers high-quality speech and adjustability of speech rate. Two male voices (Will: US English accent; Graham: UK English accent) and two female voices (Karen: US English accent; Deepa: Indian English accent) were used. These three English accents represent highly populated countries with a strong automotive industry where English is either the first language (US and UK) or one of the official languages (India). The tool offered the option for speech to be generated with an emotional tone. We created recordings for two emotional tones by selecting speakers Will Happy and Will FromAfar. We expected that Will FromAfar, in which the speaker shouts the words from a distance, would be interpreted as urgent. Will Happy was expected to sound pleasant among listeners. Note that Acapela-Box offered a limited number of speakers and emotional tones: there was no male voice with Indian English accent, and among the US English speakers, the Happy and Afar emotional tones were only available for Will. Furthermore, different voices exhibited different speech rates (e.g., Deepa spoke relatively fast).

Using Acapela-Box, each of the six speakers was recorded at eight additional settings of speech rate: –60, –45, –30, –15, +15, +30, +45, and +60, which altered the duration of the sample to approximately 151%, 131%, 119%, 109%, 90%, 85%, 79%, and 76% of its nominal value, respectively. In addition, for each speaker and speech rate, background noise was added, extracted from a YouTube video showing a Tesla Model S in Autopilot mode (Oedegaarde, 2015). For Will and Karen, noise with three extra levels of volume was added (Table 1).

Moreover, 13 phrases were recorded using Will and Karen at a nominal speech rate and without added noise: (1) “Take over please?”, (2) “Take over”, (3) “Please take over”, (4) “Could you please take over”, (5) “Could you please take over?”, (6) “Take over now”, (7) “Take over immediately”, (8) “Hazard: take over”, (9) “Danger: take over”, (10) “Warning: take over”, (11) “Caution: take over”, (12) “Attention: take over”, and (13) “Note: take over”.

In summary, the number of recordings was 140, consisting of 108 recordings where speech rate and noise were varied for each of the six speakers (6 speakers x 9 speech rate levels x 2 noise levels) plus 32 recordings (3 noise levels and 13 additional phrases, for

Download English Version:

<https://daneshyari.com/en/article/4971922>

Download Persian Version:

<https://daneshyari.com/article/4971922>

[Daneshyari.com](https://daneshyari.com)