# Crowdsourced Measurement of Reaction Times to Audiovisual Stimuli With Various Degrees of Asynchrony

**Pavlo Bazilinskyy** and **Joost de Winter**, Delft University of Technology, the Netherlands

**Objective:** This study was designed to replicate past research concerning reaction times to audiovisual stimuli with different stimulus onset asynchrony (SOA) using a large sample of crowdsourcing respondents.

**Background:** Research has shown that reaction times are fastest when an auditory and a visual stimulus are presented simultaneously and that SOA causes an increase in reaction time, this increase being dependent on stimulus intensity. Research on audiovisual SOA has been conducted with small numbers of participants.

**Method:** Participants ($N$ = 1,823) each performed 176 reaction time trials consisting of 29 SOA levels and three visual intensity levels, using CrowdFlower, with a compensation of US$0.20 per participant. Results were verified with a local Web-in-lab study ($N$ = 34).

**Results:** The results replicated past research, with a $V$ shape of mean reaction time as a function of SOA, the $V$ shape being stronger for lower-intensity visual stimuli. The level of SOA affected mainly the right side of the reaction time distribution, whereas the fastest 5% was hardly affected. The variability of reaction times was higher for the crowdsourcing study than for the Web-in-lab study.

**Conclusion:** Crowdsourcing is a promising medium for reaction time research that involves small temporal differences in stimulus presentation. The observed effects of SOA can be explained by an independent-channels mechanism and also by some participants not perceiving the auditory or visual stimulus, hardware variability, misinterpretation of the task instructions, or lapses in attention.

**Application:** The obtained knowledge on the distribution of reaction times may benefit the design of warning systems.

**Keywords:** crowdsourcing, reaction times, mental chronometry, psychophysics

Address correspondence to Pavlo Bazilinskyy, Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, the Netherlands; e-mail: p.bazilinskyy@tudelft.nl.

*HUMAN FACTORS*
Vol. XX, No. X, Month XXXX, pp. 1–15
DOI: 10.1177/0018720818787126


## INTRODUCTION

Reaction times are widely used to examine human information-processing mechanisms, such as in studies of cognitive ability (Der & Deary, 2006; Jensen, 2006), visual search (Wolfe, 1998), and memory (Baddeley & Ecob, 1973). In human factors science, reaction times are typically measured for applied purposes, for example, to quantify stimulus-response compatibility of human–machine interfaces (Chapanis & Lindenbaum, 1959; Fitts & Seeger, 1953) and the effectiveness of warning systems (Abe & Richardson, 2006). In the design of any warning system, it should be decided whether the warning signal is auditory, visual, vibrotactile, or multimodal. For example, in automated driving, a takeover warning can be a vibrotactile stimulus in the seat (Petermeijer, De Winter, & Bengler, 2016), an auditory signal (Merat & Jamson, 2009), a visual notification on the dashboard (Larsson, Johansson, Söderman, & Thompson, 2015), or a multimodal signal, such as an audiovisual alarm (e.g., Gold, Damböck, Lorenz, & Bengler, 2013) or a vibrotactile-auditory alarm (e.g., Petermeijer, Bazilinskyy, Bengler, & De Winter, 2017). The present study is concerned with a new method for large-scale research on reaction times to multimodal stimuli.

### Previous Research on the Effect of Stimulus Onset Asynchrony (SOA) on Reaction Times

It is well established that in simple reaction time tasks, multimodal feedback yields faster reaction times than unimodal feedback (Diederich & Colonius, 2004; Todd, 1912). However, the timing and intensity of the stimuli have an important effect on reaction times. Literature shows that average reaction times to bimodal stimuli are fastest when the onsets of the stimuli occur at the same moment, with the mean reaction time as a function of SOA exhibiting a

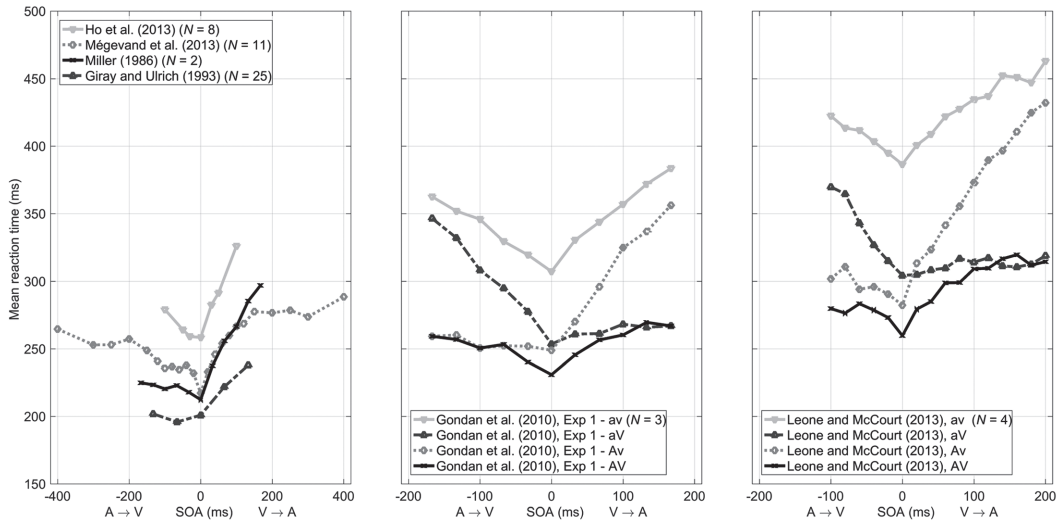


*Figure 1.* Mean reaction times from a selection of literature on stimulus onset asynchrony in audiovisual reaction time tasks. Left = four independent studies (Giray & Ulrich, 1993; Ho, Gray, & Spence, 2013; Mégevand, Molholm, Nayak, & Foxe, 2013; Miller, 1986); middle = study that manipulated stimulus intensity (Gondan, Götze, & Greenlee, 2010); right = study that also manipulated stimulus intensity (Leone & McCourt, 2013); a (A) = low- (high-) intensity auditory stimulus; v (V) = low- (high-) intensity visual stimulus.

*V* shape (e.g., Miller, 1986). This *V* shape is illustrated in Figure 1, showing results from our literature survey on reaction times to audiovisual stimuli as a function of SOA. Only studies that used equivalent task conditions were included in this figure (for additional relevant research on SOA, see Harrar, Harris, & Spence, 2017; Leone & McCourt, 2015; Van der Stoep, Spence, Nijboer, & Van der Stigchel, 2015). Each subfigure shows mean reaction times as a function of SOA, where a negative SOA value means that the onset of the visual stimulus occurred after the onset of the auditory stimulus. The middle and right subfigures concerned studies that focused on manipulating the intensity of the visual and auditory stimuli, as indicated with lowercase (v, a) and uppercase letters (V, A).

It can also be seen in Figure 1 that the degree with which reaction times increase as a function of SOA depends on the intensity of the stimuli (see also Miller & Ulrich, 2003). More specifically, if the visual stimulus is difficult to see, then participants are likely to respond to the auditory stimulus, and so the onset of the auditory stimulus will have a dominant effect on the mean reaction time. Conversely, if the stimulus is poorly audible, then the onset of the visual stimulus will determine the reaction time. These interactions between SOA and stimulus intensity were illustrated by Gondan, Götze, and Greenlee (2010; see Figure 1 middle) and Leone and McCourt (2013; see Figure 1 right). Thus, the relationship between mean reaction time and SOA is asymmetric (i.e., one side of the *V* shape is flatter than the other) when the auditory stimulus is weak and the visual stimulus is intense (i.e., aV in Figure 1) or when the visual stimulus is weak and the auditory stimulus is intense (Av in Figure 1).

Differences in the overall mean reaction time between the experiments shown in Figure 1 are of lesser interest, as these depend on factors such as the participants' level of experience, outlier removal, overall stimulus intensity, and hardware used during the experiment (e.g., Dodonova & Dodonov, 2013; Gondan & Minakata, 2016). For example, in Diederich and Colonius (2004) and Hershenson (1962), the mean reaction times to audiovisual stimuli were in the range of 135 to 155 ms (SOA 0–50 ms) and 98

to 144 ms (SOA 0–85 ms), respectively. These phenomenally fast reaction times may be explained by the fact that participants were highly trained, the use of intensive stimuli, and specialized hardware that records reaction times with little delay.

The research on the effect of audiovisual SOA has been conducted with small sample sizes (see the legends in Figure 1) but typically with dozens of trials per stimulus condition. Accordingly, investigations of the distributions of reaction times have been performed within subjects rather than between subjects. For example, in Miller (1986), there were two participants who each completed 40 test blocks over a period of about 1 month, each block consisting of 130 test trials. It would be relevant to examine whether there exist individual differences in susceptibility to SOA effects. Within the human factors community, it has been emphasized that the design of warning systems should not be based on the mean reaction time but that slowly responding participants should be considered, too (Eriksson & Stanton, 2017; Wickens, 2001).

## The Potential of Crowdsourcing for Performing Reaction Time Research

The Internet is a now well-established medium for experimental psychological research with large sample sizes (Fortenbaugh et al., 2015). Various studies have replicated classical psychological effects using online crowdsourcing methods (Crump, McDonnell, & Gureckis, 2013; Hilbig, 2016). For example, Barnhoorn, Haasnoot, Bocanegra, and Van Steenbergen (2015), using a JavaScript engine, replicated three reaction time paradigms (Stroop task, attentional blink task, masked priming task) via crowdsourcing.

A number of studies suggest that online software and hardware can cause small delays compared to regular psychophysics methods. For example, De Leeuw and Motz (2016) found an additive reaction time delay of 25 ms, and no difference in variance, when using jsPsych (a library for creating behavioral experiments using JavaScript) running in Google Chrome as compared with MATLAB's Psychophysics Toolbox on the same laptop hardware. Reimers and Stewart (2016) described a limitation of JavaScript, in that audio and visual stimuli scheduled to appear on a Web page at the same time are presented with a small temporal offset that can vary up to 40 ms, depending on the type of browser. Schubert, Murteira, Collins, and Lopes (2013) replicated the Stroop effect online and noted that the online software contributed to additional reaction time variance compared with a controlled lab study. According to simulations by Brand and Bradley (2012), the effect of technical variance (due to e.g., keyboards, CPU load, operating systems) is negligible compared with individual differences in reaction time, and they argued that "researchers' preconceptions concerning the unsuitability of web experiments for conducting research using response time as a dependent measure are misguided" (p. 350).

However, concerns have also been raised about the validity of online research, especially when small stimulus durations are involved. Semmelmann and Weigelt (2017) replicated well-known paradigms (e.g., Stroop test, flanker test) in three settings (classical lab, Web-in-lab, Web), with a total of 147 participants. Although the replication was successful, the mean reaction times in a simple reaction time task were 253 ms, 280 ms, and 318 ms, respectively, for the three settings. That is, the Web-in-lab method caused an additive delay, presumably due to the browser engine and JavaScript, whereas the Web method might be further affected by differences in participants' hardware and testing environments. Woods, Velasco, Levitan, Wan, and Spence (2015) provided a review of 10 online research platforms that can be used for measuring reaction times and concluded that the quality of online data is usually high. However, these authors also discussed sources of technical variability in online reaction time research, such as variability in screen brightness, screen color, and volume of auditory stimuli, and they argued that studies that require short stimulus presentation are not well suitable to online research. Similarly, Schubert et al. (2013) argued that "the smaller the effect, the more problematic the noise introduced by . . . online experimentation" (p. 10).

In summary, although the Internet can be used to replicate psychological phenomena concerning reaction times, online research is associated with additive bias and extra sources of

variance compared to lab-based research, and it is unknown whether reaction times to small temporal manipulations can be replicated online.

### Aim of This Research

Given the knowledge gap, this study was designed to replicate previous research on the effect of SOA and stimulus intensity on audiovisual reaction times using a large sample of participants via crowdsourcing. A replication study of well-established previous findings may contribute to the understanding of the validity of crowdsourcing and yield new knowledge on the relationship between SOA and reaction times.

Our analysis was concerned with investigating whether a *V* shape of mean reaction times (Figure 1) replicates and whether a lower intensity of the visual stimulus causes the slope of the *V* shape to be steeper. As pointed out earlier, crowdsourcing research can yield a high variance in reaction times. Therefore, in addition to investigating mean reaction times, we examined individual differences in reaction times (percentiles and trial-to-trial correlations). Furthermore, we assessed the sources of variability in reaction times by examining learning curves, by comparing the results with a Web-in-lab study using the same software, and by studying the effects of experimental conditions, such as whether participants were using a keyboard or mobile phone or whether they were indoors or outdoors.

## METHOD

This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Committee (HREC) at the Delft University of Technology. Informed consent was obtained from each participant.

### Stimuli

Participants were presented with audiovisual stimuli. The auditory stimuli were single 210-ms-long beeps of 1,840 Hz. The visual stimuli were red circles on a white background. A total of 29 SOA values were used: −1,000, −500, −300, −200, −100, −90, −80, −70, −60, −50, −40, −30, −20, −10, 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 500, and 1,000 ms. These 29 SOA values have a range that is higher than the ranges of SOA values used in previous research (Figure 1) while offering a higher temporal resolution (10 ms for SOA values between −100 and 100 ms). A negative SOA value means that the onset of the auditory stimulus occurred before the onset of the visual stimulus, and a positive SOA value means that the onset of the auditory stimulus occurred after the visual stimulus (as in Figure 1). Figure 2 shows example timelines of reaction time trials with negative and positive SOA.

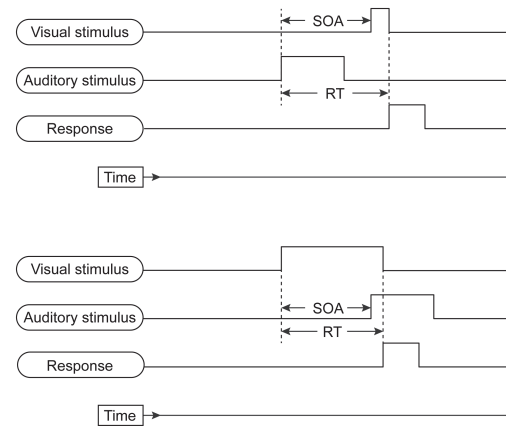


*Figure 2.* Timelines for events of a reaction time trial. The top figure concerns stimulus onset asynchrony (SOA) < 0 ms; the bottom figure concerns SOA > 0 ms. The auditory stimulus always had a duration of 210 ms, whereas the visual stimulus disappeared when the participant pressed the response key.

If the auditory stimulus was presented at the same moment or after the visual stimulus (SOA ≥ 0), then a .png file was presented together with a .wav file, with the time delay (SOA) encoded in the .wav file. If the visual stimulus was presented after the auditory stimulus (SOA < 0), then an animated .gif file was presented together with a .wav file. The animated .gif (via its graphics control extension) was a practical solution to encode a time delay of the onset of the visual stimulus. The rendering of the stimuli was powered by the jsPsych JavaScript library for running behavioral experiments online (De Leeuw, 2015).

The red circles were uniform, had a diameter of 195 pixels, and had three levels of intensity:

*Figure 3.* Visual stimulus in the browser's full-screen mode (RGB 246-166-174, screen resolution: 1,920 × 1,080 pixels).

low, medium, or high (see Figure 3 for an example). These three intensity levels (i.e., shades of red) were selected to be notably different but in such a way that the low-intensity stimulus was still clearly distinguishable from the white background, as we did not want that participants would fail to detect the visual stimuli. High-intensity stimuli were rendered on the screen as RGB 233-33-53. Low- and medium-intensity .png files were created using 40% and 70% transparency setting, respectively, which translates into rendered stimuli of RGB 246-166-174 and RGB 240-99-113, respectively. Low- and medium-intensity .gif files were RGB 251-211-215 and RGB 242-122-134, respectively. Because of the different RGB rendering of .png and .gif files, the reaction times to low- and medium-intensity stimuli between SOA < 0 and SOA ≥ 0 should not be directly compared. The auditory stimuli were always 210-ms beeps; they were not varied in intensity to keep the total number of conditions manageable.

## Crowdsourcing Experiment

Participants in the online experiment participated via the crowdsourcing platform CrowdFlower (https://www.crowdflower.com). Participants became aware of this research by logging into one of many channel websites (e.g., https://www.clixsense.com), where they would see our study in the list of other projects available for completion. We allowed contributors from all countries to participate. It was not permitted to complete the study more than once from the same worker ID. A payment of US$0.20 was offered for the completion of the experiment. A total of 2,000 participants completed the experiment, at a total cost of US$480. Our payment was assumed to be high enough to incentivize participants. Litman, Robinson, and Rosenzweig (2015) investigated the effect of payment for a 6-min task among crowdworkers from India and found that a payment of US$0.10 ("above-minimum-wage condition") yielded higher data quality than a payment of US$0.02 ("below-minimum-wage condition"), whereas a payment of US$1.00 ("far above the minimum wage") did not improve data quality compared with a payment of US$0.10.

Participants first answered a number of questionnaire items. At the beginning of the questionnaire, contact information of the researchers was provided, and the purpose of the upcoming study was described as "to determine reaction times for different types of visual and auditory signals." Participants were informed that the study would take approximately 8 min. The participants were also informed that they could contact the investigators to ask questions about the study and that they had to be at least 18 years old. Information about anonymity and voluntary participation was provided as well. The questionnaire started with the following questions:

- "Have you read and understood the above instructions?" ("Yes," "No")
- "What is your gender?" ("Male," "Female," "I prefer not to respond")
- "What is your age?" (positive integer)
- "In which type of place are you located now?" ("Indoor, dark"; "Indoor, dim light"; "Indoor, bright light"; "Outdoor, dark"; "Outdoor, dim light"; "Outdoor, bright light"; "Other"; "I prefer not to respond")
- "Which input device are you using now?" ("Laptop keyboard," "Desktop keyboard," "Tablet on-screen keyboard," "Mobile phone on-screen keyboard," "Other," "I prefer not to respond")

Several additional questions were asked about driving habits, which were not used in this study. The participants were then asked to leave

the questionnaire by clicking on a link that opened a Web page with the reaction time task. Participants were presented with instructions on how to complete the given task:

> In this experiment, you will hear sounds and see red circles. Please make sure that your audio is on and set your screen to bright. You need to press "F" after hearing a sound OR seeing a red circle (whichever comes first) as fast as possible. Your reaction times will be recorded. After each group of 25 stimuli you will be able to take a small break. Please press any key to start with the first stimulus.

The participants had to respond to 88 different stimuli in random order. Each stimulus was repeated twice, yielding 176 stimuli for each participant (i.e., 29 SOA values × 3 visual intensity levels × 2 repetitions + 2 repetitions of an audio-only stimulus). There was no upper limit to the reaction times; the next stimulus trial was loaded after the participants pressed the *F* button. The stimuli were presented in six batches of 25 and one last batch of 26. After a batch, participants were shown the following text: "You have now completed 25 [50, 75, 100, 125, 150] stimuli out of 176. When ready press 'C' to proceed to the next batch." An analysis of the elapsed times showed that participants took a median time of 9 s to press *C* after the first batch and a median time of 4 s to press *C* after the sixth batch.

After pressing the *F* key, a new stimulus was presented after a uniform random delay between 1,000 and 3,299 ms, in agreement with Diederich and Colonius (2004). The images and sounds were preloaded to eliminate unwanted delays between the stimuli. Data for each participant were saved in a database after the 176th stimulus. Analyses of the distribution of reaction times per participant showed that the temporal resolution of the reaction time measurements (i.e., the minimum difference that could be detected) differed between participants: For the majority of participants (88%), the temporal resolution was between 2.6 and 3.0 ms. For 6% of the participants, the temporal resolution was between 3 ms and 12 ms, whereas 4% of participants had a temporal resolution of 42.7 ms. These differences in temporal resolution may be due to different platforms and browsers used by the participants.

At the end of the experiment, participants were shown a unique code. Participants were asked to note down this code and return to the Web page of the questionnaire. They were required to enter the code on the questionnaire as proof that they completed the experiment and to receive their remuneration.

### Web-in-Lab Experiment

To verify the results of the crowdsourcing experiment in controlled experimental conditions, we launched the same task in a laboratory setting. We collected responses from 42 participants from the university community. All participants completed the task on the same MacBook Air (13-in. screen, 8 GB memory, Intel Core I7 processor, two cores) laptop behind a table in a standard office room of about 3 × 3 m. The blinds were closed to control the lighting conditions; the ceiling lights (fluorescent lamps) were always on. The volume of the laptop was set to 60% (corresponding to a measured sound intensity of 60–65 dBA), and the brightness of the display was 100%. The experimenter started up the task and left the room so that the participant completed the task while being alone in the room. The temporal resolution of the reaction times of the Web-in-lab experiment was 5.8 ms. Participants of the Web-in-lab experiment did not receive remuneration because it is common practice at our institution to not pay participants for a short-lasting experiment.

### Handling of Reaction Time Outliers and Statistical Testing

Reaction times less than 0 ms were removed from the analysis, whereas reaction times greater than 1,500 ms were set equal to 1,500 ms. Using this so-called winsorization method, extremely slow reaction times (>1,500 ms) were retained in the analysis (as recommended by Gondan & Minakata, 2016), while limiting the skewness and kurtosis of the reaction time distribution (Ratcliff, 1993). Differences between participants' conditions (e.g., input device) were compared using an unequal-variance *t* test (Welch, 1947) after performing an inverse transformation of the reaction

times (Ratcliff, 1993). Effect sizes were assessed using Cohen's *d* of the inverse reaction times.

## RESULTS

The responses were collected between March 3, 2017, 12:42 and March 4, 2017, 17:30 (GMT). Two hundred twenty-four participants completed an optional user satisfaction survey offered by CrowdFlower. The study received an overall satisfaction score of 4.4 out of 5.0 on a scale from 1 (*very dissatisfied*) to 5 (*very satisfied*). The mean response to the question "How clear were the task instructions and interface?" was 4.6 on a scale from 1 (*very unclear*) to 5 (*very clear*), and the mean response to "How would you rate the pay for this task relative to other tasks you've completed?" was 4.2 on a scale from 1 (*much worse*) to 5 (*much better*).

### Participant Filtering and Participant Characteristics

Out of 2,000 participants, 177 were removed during data filtering. These were participants for whom no reaction time data were available due to a server/recording error ($n = 119$), participants with more than 20% negative reaction times (due to pressing the response key before the stimulus was presented; $n = 55$), or participants who answered "no" to the question whether they had read and understood the task instructions ($n = 9$). The 20% threshold was assumed to discriminate between participants with genuine anticipatory reaction times (i.e., accidentally pressing the *F* key too early) and participants cheating the system by repeatedly pressing *F*.

In the group of the remaining 1,823 participants, 1,283 were male, 533 were female, and 7 did not specify their gender. Three participants reported an unrealistic age or an age that was not in agreement with the task instructions (3, 5, and 17 years). Because these ages could be the result of a basic typographical error, and because these three participants did complete the task, they were retained in the analysis. The participants' mean age for the 1,820 participants of 18 years and older was 33.9 years ($SD = 10.1$, min = 18, max = 71).

The participants were from 83 different countries, with 22 countries having 25 or more respondents and four countries (Spain, Russia, Serbia, Venezuela) having more than 100 respondents.

### Learning Curve

Figure 4 shows that the mean reaction times decreased with trial number, that is, the participants showed faster reaction times as the experiment progressed. The spikes in the graph represent the trials that directly followed the breaks after each 25th stimulus. We removed Trials 1 through 5, 26, 51, 76, 101, 126, and 151 from the remaining analysis (except the correlations among trials), because these trials may be invalid as it is likely that some participants were still learning the basics of the task or pressed an incorrect key during these trials.

### Effects of SOA and Stimulus Intensity on Reaction Time

Figure 5 shows the mean reaction times as a function of SOA. Note that the results for SOA = −10 ms are not shown in the figures because the animated .gif files showed a delay of 100 ms when programmed with a delay of 10 ms. It can be seen that the lowest reaction times were obtained when the SOA was 0 ms. Furthermore, the visually delayed stimuli (i.e., SOA < 0 ms) yielded a mean reaction time that was about 43 ms higher than the auditorily delayed stimuli (i.e., SOA > 0 ms). It can also be seen that the low-intensity visual stimuli were associated with a stronger increase of the mean reaction time for increasing SOA than the high-intensity visual stimuli, which is consistent with the literature presented in Figure 1.

Individual differences were assessed using percentiles of the observed reaction times, from low (i.e., fast reactions) to high (i.e., slow reactions). Figure 6 shows that the lowest reaction times were hardly affected by SOA, whereas the 95th percentile is strongly sensitive to SOA. In other words, the changes in mean reaction time observed in Figure 5 can be largely attributed to differences in the right tail of the reaction time distribution.

### Effects of Experimental Conditions on Reaction Time

Figure 7 shows that indoor lighting condition did not have a large impact on the mean reaction times. A Welch's test showed no significant difference between dark and bright indoor light, $t(189.3) = 0.51$, $p = .608$, $d = 0.05$. However,

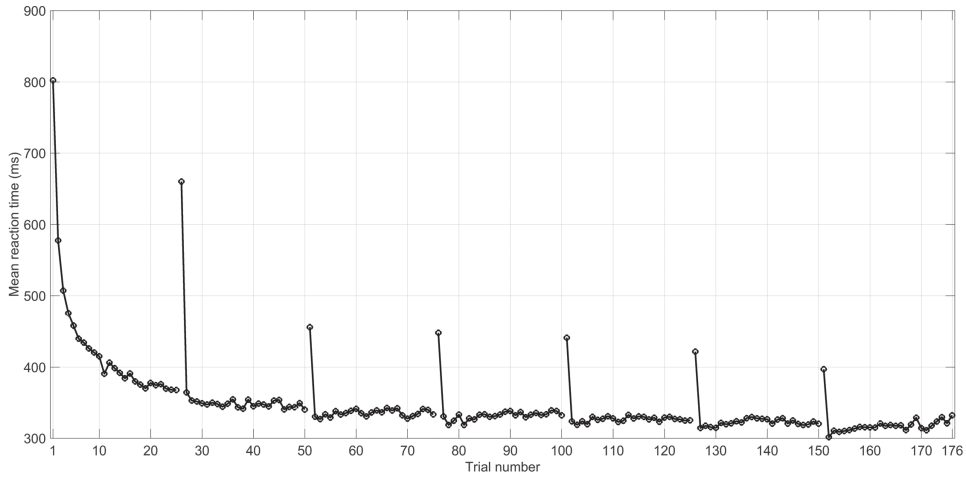


*Figure 4.* Mean reaction times versus trial number in the crowdsourcing study ($N = 1{,}823$). Each data point represents the mean across approximately 1,815 trials (i.e., 1,823 participants minus excluded responses).

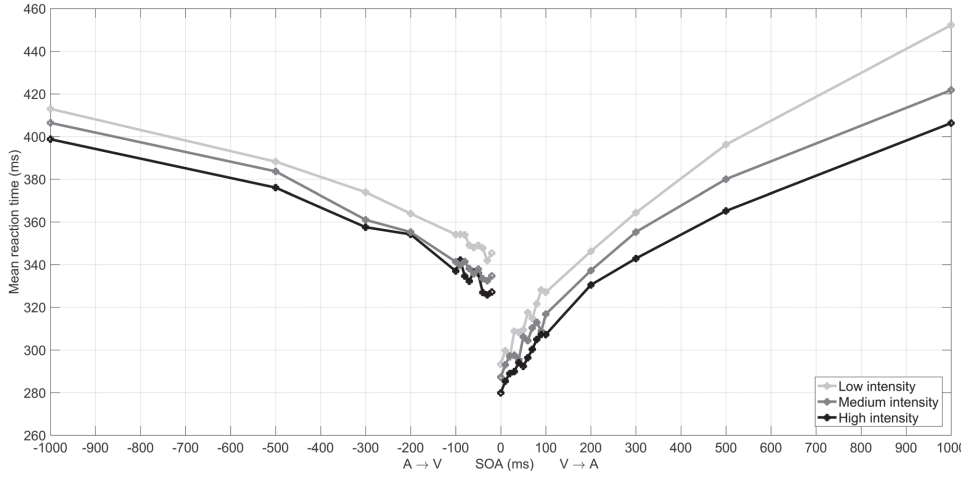


*Figure 5.* Mean reaction times for 28 levels of stimulus onset asynchrony (SOA) and three levels of visual intensity in the crowdsourcing study. Each data point represents the mean across approximately 3,400 trials (i.e., 1,823 participants × 2 trials per participant minus excluded responses).

completing the task outdoors was associated with significantly higher reaction time than completing the task indoors, $t(40.7) = 3.32$, $p = .002$, $d = 0.55$.

Figure 7 also shows that participants who completed the task with a laptop or desktop keyboard had faster reaction times than participants who used other input devices (e.g., tablets or mobile phones), $t(20.4) = 2.73$, $p = .013$, $d = 0.62$. There were no statistically significant gender differences, $t(1020.9) = 0.39$, $p = .697$, $d = 0.02$, nor age differences in reaction time (Spearman's correlation between age and mean reaction time: $\rho = 0.03$, $N = 1{,}820$).

## Trial-to-Trial Correlations (Stability) of Reaction Times

Finally, we calculated trial-to-trial correlations to obtain an indication of the stability of

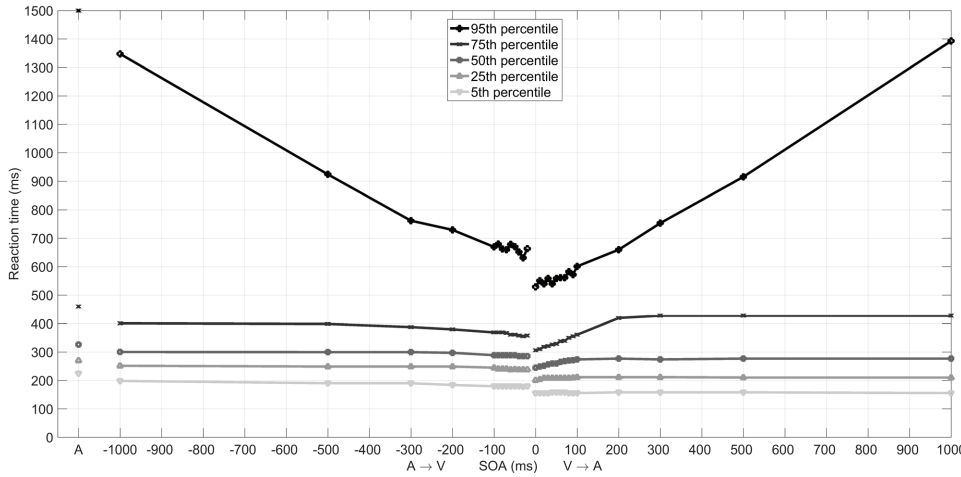


*Figure 6.* Percentiles of the reaction times for 28 levels of stimulus onset asynchrony (SOA) in the crowdsourcing study. Each data point is based on approximately 10,200 trials (i.e., 1,823 participants × 6 trials per participant minus excluded responses). The auditory-only trial (A) is based on 3,397 trials (i.e., 1,823 participants × 2 trials per participant minus 249 excluded responses).

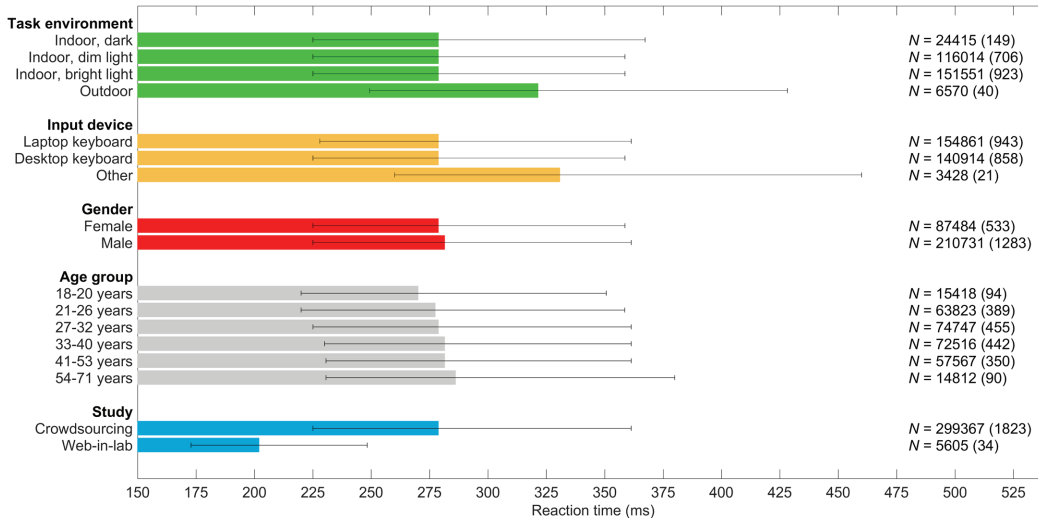


*Figure 7.* Median reaction time at the level of trials, per task environment, input device, gender, and age group for the crowdsourcing study and for the crowdsourcing versus the Web-in-lab study. The error bars denote the 25th and 75th percentiles. Also listed is the number of trials with the number of participants in parentheses. *Outdoor* refers to "Outdoor, dark"; "Outdoor, dim light"; and "Outdoor, bright light" combined. *Other* refers to "Tablet on-screen keyboard," "Mobile phone on-screen keyboard," and "Other" combined. Ages of 20, 26, 32, 40, 53, and 71 years are the 5th, 25th, 50th, 75th, 95th, and 100th percentiles of participants' ages, respectively.

participants' reaction times. Figure 8 shows a Spearman correlation matrix among the reaction times per trial number for the crowdsourced participants. A high correlation between a pair of trials means that participants' reaction times had a similar rank ordering in these two trials, whereas a correlation of zero would be expected if participants were not consistent at all. A clear

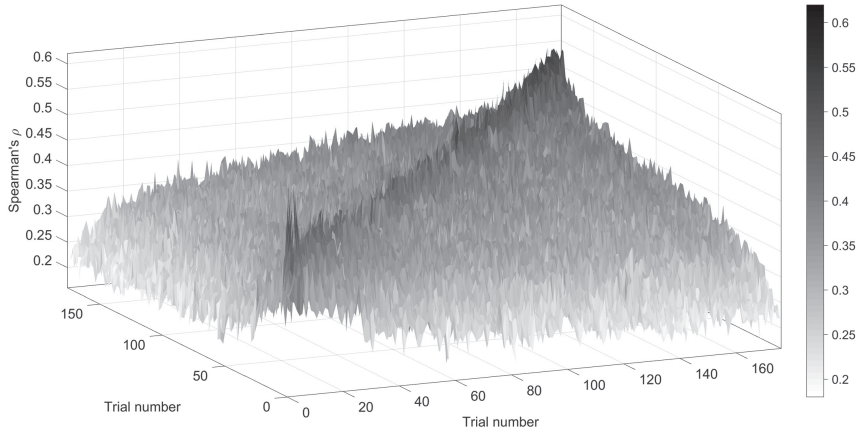


*Figure 8.* Heat map of Spearman rank-order correlations of crowdsourced participants' reaction times between Trials 1 and 176.

simplex pattern can be seen, with temporally adjacent trials showing higher correlations than temporally disparate trials (see also Ackerman, 1987). It can also be seen that reaction times stabilized (i.e., higher correlations) in later trials.

### Reaction Times From the Web-in-Lab Experiment Compared With the Crowdsourcing Experiment

In the laboratory setting, we retained responses from 34 participants, obtained between March 7, 2017, 10:57, and March 10, 2017, 13:13 (GMT). We removed four participants with incomplete data and four participants who were involved in pilot tests conducted during the design of the study. The participants were six females and 28 males, having a mean age of 27.1 years ($SD$ = 6.6 years, min = 18, max = 56). The reaction times were processed identically to the crowdsourcing experiments.

The results in Figure 7 show substantial differences between the international crowdsourcing method and the local lab method, $t(34.8) = 10.68, p < .001, d = 1.57$. The Web-in-lab method featured a lower mean reaction time and lower variability of reaction time (Figure 9).

## DISCUSSION

### Replicated Effects

In this study we aimed to replicate published research regarding the effects of audiovisual SOA and visual stimulus intensity on reaction times with a large sample of crowdsourced participants and to examine sources of variability of mean reaction times (e.g., learning curves, task conditions, comparison with Web-in-lab study).

Our findings replicated the $V$ shape as observed in past research, with the mean reaction time being fastest when SOA = 0 ms and increasing monotonically both with increasing and decreasing SOA. The effect of stimulus intensity was also replicated, as evidenced by the higher reaction times for visual stimuli of lower intensity, as well as by the relatively steep slope of mean reaction times for low-intensity visual stimuli when the auditory stimulus was presented after the visual one (SOA > 0). This steep slope could also be seen for low-intensity visual stimuli in Figure 1 (Av condition).

Crowdsourcing allows researchers to access a large pool of participants, thereby yielding high statistical power. This can be illustrated with a post hoc power analysis: For a false-positive rate of 1%, a sample size of 1,823, and an effect size for a pair of conditions ($d_z$) of 0.109 (calculated from a mean difference of 20 ms, an observed $SD$ across participants of 179 ms, and an observed correlation between the two groups of 0.47), the achieved statistical power is 98.0% (Faul, Erdfelder, Lang, & Buchner, 2007). The results in the figures allowed for a reliable assessment of experimental effects and individual

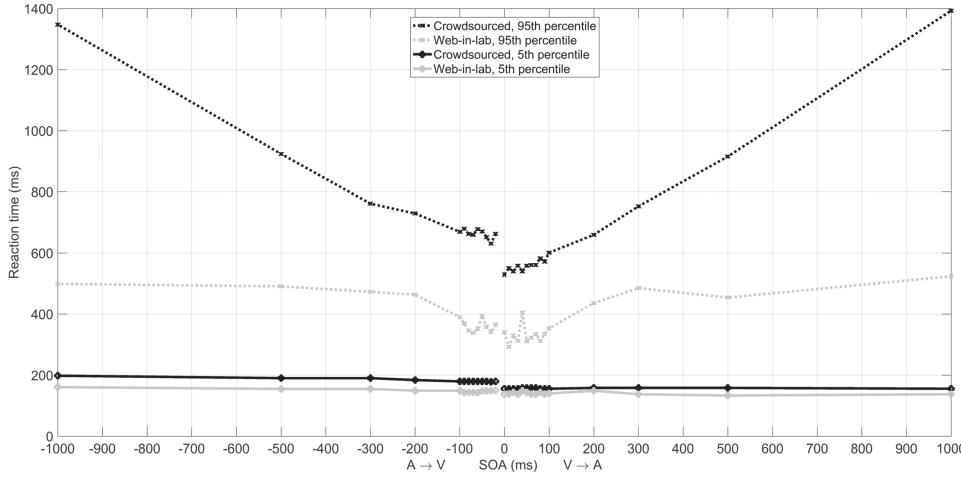


*Figure 9.* The 5th and 95th percentiles for 28 levels of stimulus onset asynchrony (SOA) in the crowdsourcing and Web-in-lab studies. Each data point of the crowdsourcing study is based on approximately 10,200 trials (i.e., 1,823 participants × 6 trials per participant minus excluded responses). Each data point of the Web-in-lab study is based on approximately 191 trials (34 participants × 6 trials per participant minus excluded responses).

differences results (effect of SOA, stimulus intensity, learning curves, percentiles).

## Effects of Experimental Conditions and Comparison Between the Crowdsourcing Experiment and the Web-in-Lab Experiment

Although the expected effects were clearly replicated, there were substantial differences in reaction times between the crowdsourcing study and the Web-in-lab study. The differences between the two methods may be because the Web-in-lab participants used the same high-quality laptop, which displayed the stimuli with the same intensity, whereas it is plausible that at least some crowdsourcing participants had poor or malfunctioning hardware or did not have their audio turned on despite the task instructions. Some of the crowdsourcing participants completed the task outdoors, which was associated with slower reaction times, possibly due to poor lighting conditions or distractive elements in the environment. Also, crowdsourcing participants who used a handheld device had a higher mean reaction time than participants who used a laptop or PC, which may be because the former involves hardware delays or may be hard to use if one's task is to provide input as quickly as possible. Furthermore, it is possible that the lab participants were concentrated and motivated to perform well, whereas the crowdsourced participants may have taken the task less seriously because they were anonymous. Previous research shows that IQ and reaction time share a negative correlation (Jensen, 2006; Madison, Mosing, Verweij, Pedersen, & Ullén, 2016). The lab participants, who were mostly students at a technical university, may have faster reaction times than the typical international crowdsourcing participant.

We did not find a statistically significant correlation between the mean reaction time and the mean age of the crowdsourced participants. This lack of correlation may be because the oldest participant in our study was 71 years old, whereas simple reaction times increase with age especially for people above 70 years old (Der & Deary, 2006). It is also possible that the relationship between age and reaction time is confounded because CrowdFlower participants from lower-income countries tend to be younger (De Winter & Dodou, 2016).

Another source of difference between the crowdsourcing and Web-in-lab study may concern differences in understanding of the task instructions. In previous CrowdFlower research,

we found that participants from English-speaking countries (De Winter, Kyriakidis, Dodou, & Happee, 2015) and participants from countries with a higher gross domestic product (GDP) per capita (De Winter & Dodou, 2016) took less time to complete a questionnaire, which may have been due to difficulty in processing English-language text. Similarly, in a supplementary analysis of the present study, we found that participants from countries with a higher GDP per capita had a lower median time to complete the experiment, including the questionnaire (Spearman's $\rho = -0.67$, $p < .001$, based on 22 countries with 25 or more respondents), and had a faster mean reaction time as well (Spearman's $\rho = -0.36$, $p = .101$). (Supplemental material is available at https://doi.org/10.4121/uuid:673c9bbc-bf17-42fa-a23a-3d716e141b1f) In summary, national differences may be a source of heterogeneity in the crowdsourcing study.

In our analysis, 55 of 2,000 participants were excluded due to negative reaction times. We aimed to show the variability of reaction times and therefore did not exclude slow-responding participants. However, others who use crowdsourcing and aim for clean data could opt for applying stricter screening criteria.

### Learning Curve and Trial-to-Trial Correlations

The first trials were associated with slower reaction times as the participants needed time to get used to the system. Also, the participants showed increased reaction times after the breaks, which is presumably because some participants did not have their finger on the keyboard yet or initially pressed an incorrect key. That is, participants had to press *C* to proceed to the next batch of trials but had to press *F* after each trial, which may result in initial confusion. We also found that performance became more stable (i.e., higher between-trial correlations) as the experiment progressed. This increase of stability may be caused by the fact that participants learned the nature of the task and entered the autonomous phase of skill learning, in which performance is less susceptible to task-irrelevant distractions (Fitts & Posner, 1967).

### Individual Differences and Reinterpretation of the Effects of SOA on Reaction Times

We found that SOA hardly affected the fastest reaction times, but it did have a substantial effect on the slowest (e.g., 95th percentile) of the reactions. That is, the hypothesized *V* shape of mean reaction time as a function of SOA was evident only in the mean reaction time and the higher percentiles of reaction time but was hardly evident from the 5th, 25th, and 50th percentiles of reaction time. These observations suggest that reductions in mean reaction times caused by simultaneous multimodal feedback are not necessarily due to multisensory neural integration, in which auditory and visual information is summed or combined in the central nervous system or at the level of individual neurons (Stein & Stanford, 2008). Our findings can be explained using an independent-channels mechanism where the visual and auditory channels operate in parallel (see Nickerson, 1973, for a review). That is, our results can be explained by the notion that participants sometimes do not attend to the auditory or visual stimulus. For example, a participant may be temporarily blinded due to an eye blink (typically lasting 150 ms; Wang, Toor, Gautam, & Henson, 2011), as a result of which he or she is more likely to react to the auditory stimulus. Similarly, a participant may have a lapse in hearing (e.g., due to an internal distraction or masking due to external noise), as a result of which he or she is more likely to react to the visual stimulus. Future research could use eye tracking and neurophysiological measures to investigate how reaction times depend on eye blinks and lapses in attention.

Participants were not prescreened based on their hearing or visual disabilities or other criteria that could affect their ability to complete this task. The variability in the right tail of the reaction time distribution may also be caused by individual differences in the understanding of the task instructions (i.e., to respond to the first of the two stimuli), in sensory ability, and in computer hardware. People with a hearing disability or with malfunctioning speakers, for example, by definition have to react to the visual

stimuli. The auditory-only stimulus caused a relatively high proportion of delayed responses (≥1,500 ms), which suggests that a portion of participants were "waiting" for the visual stimulus to arrive or did not have their sound enabled. More generally, our findings suggest that warning signals should be audiovisual rather audio only or visual only and that the visual and auditory warning should be presented simultaneously (SOA = 0 ms), as was done in an automated driving study by Petermeijer et al. (2017), for example.

### Limitations of Crowdsourcing Regarding Temporal Resolution and Timing of Audiovisual Stimuli

Based on their review, Woods et al. (2015) argued that "only subset of studies, specifically those requiring short stimulus presentation, are not so well suited to online research" (p. 15). We indeed did have some technical problems in the presentation of the stimuli. First, we observed a limited temporal resolution of the reaction time measurements in the crowdsourcing experiment, being 2.6 to 3 ms for 88% of the participants but 42.7 ms for 4% of the participants. Second, the animated .gifs do not render properly for delays of 10 ms (i.e., SOA = −10 ms), a known issue in computer graphics (Karonen, 2012). Also, the .gif files were associated with an average additive delay of about 43 ms. This additive delay was not observed in the lab study and was hardly present among the faster responses. Thus, it is possible that the 43-ms delay was caused by certain browsers not displaying the animated .gif files properly. Despite the problems observed with animated .gif files, differences in reaction times could be detected even for auditory and visual delays of 10-ms increments, which is noteworthy when considering that a typical screen refresh rate is only 17 ms (see Woods et al., 2015, for further discussion). In summary, we obtained credible experimental effects despite imperfect control of the SOA and despite a limited temporal resolution of the measurements. The robustness of reaction times to noise and temporal resolution is in agreement with simulations by Ulrich and Giray (1989) and Reimers and Stewart (2015). Authors of future research could extend our approach by varying not only the intensity of visual stimuli but also the intensity of the auditory stimuli.

## CONCLUSIONS

We conclude that crowdsourcing may allow for large-scale reaction time research, at the expense of a lack of control of the test environment. For example, screen brightness and rendering problems may affect the perception of visual stimuli, whereas hardware volume level can affect the perception of auditory stimuli. The expected effects of SOA were replicated despite the variable test environment, which indicates that crowdsourcing is a powerful tool in reaction time research.

## ACKNOWLEDGMENTS

The research presented in this paper is being conducted in the project HFAuto–Human Factors of Automated Driving (PITN-GA-2013-605817).

## KEY POINTS

- A large crowdsourcing study ($N$ = 1,823) on audiovisual reaction times was performed.
- Stimulus onset asynchrony and visual stimulus intensity were varied.
- Results replicated past psychophysics research that used comparatively small sample sizes.
- Crowdsourcing yielded higher variability in reaction times than a local Web-in-lab study.
- Crowdsourcing is a promising medium for reaction time research.

## ORCID ID

J. C. F. de Winter https://orcid.org/0000-0002-1281-8200

## SUPPLEMENTAL MATERIAL

Supplemental material is available at https://doi.org/10.4121/uuid:673c9bbc-bf17-42fa-a23a-3d716e141b1f

## REFERENCES

Abe, G., & Richardson, J. (2006). Alarm timing, trust and driver expectation for forward collision warning systems. *Applied Ergonomics*, *37*, 577–586. https://doi.org/10.1016/j.apergo.2005.11.001

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3–27. https://doi.org/10.1037/0033-2909.102.1.3

Baddeley, A. D., & Ecob, R. J. (1973). Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology*, *25*, 229–240. https://doi.org/10.1080/14640747308400342

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & Van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods*, *47*, 918–929. https://doi.org/10.3758/s13428-014-0530-7

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *30*, 350–357. https://doi.org/10.1177/0894439311415604

Chapanis, A., & Lindenbaum, L. E. (1959). A reaction time study of four control-display linkages. *Human Factors*, *1*, 1–7. https://doi.org/10.1177/001872085900100401

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE*, *8*, e57410. https://doi.org/10.1371/journal.pone.0057410

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*, 1–12. https://doi.org/doi:10.3758/s13428-014-0458-y

De Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*, 1–12. https://doi.org/10.3758/s13428-015-0567-2

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging*, *21*, 62–73. https://doi.org/10.1037/0882-7974.21.1.62

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

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

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388–1404. https://doi.org/10.3758/BF03195006

Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, *41*, 674–687. https://doi.org/10.1016/j.intell.2013.09.001

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, *59*, 689–705. https://doi.org/10.1177/0018720816685832

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Fitts, P. M., & Seeger, C. M. (1953). SR compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210. https://doi.org/10.1037/h0062827

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science*, *26*, 1497–1510. https://doi.org/10.1177/0956797615594896

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1278–1291. https://doi.org/10.1037/0096-1523.19.6.1278

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

Gondan, M., Götze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics*, *72*, 1692–1709. https://doi.org/10.3758/APP.72.6.1692

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, *78*, 723–735. https://doi.org/10.3758/s13414-015-1018-y

Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of perceived simultaneity. *Experimental Brain Research*, *235*, 763–775. https://doi.org/10.1007/s00221-016-4822-2

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289–293. https://doi.org/10.1037/h0039516

Hilbig, B. E. (2016). Reaction time effects in lab-versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*, 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Ho, C., Gray, R., & Spence, C. (2013). Role of audiovisual synchrony in driving head orienting responses. *Experimental Brain Research*, *227*, 467–476. https://doi.org/10.1007/s00221-013-3522-4

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam, Netherlands: Elsevier.

Karonen, I. (2012, March 8). Why is this GIF's animation speed different in Firefox vs. IE? [Answer]. Retrieved from https://webmasters.stackexchange.com/questions/26994/why-is-this-gifs-animation-speed-different-in-firefox-vs-ie

Larsson, P., Johansson, E., Söderman, M., & Thompson, D. (2015). Interaction design for communicating system state and capabilities during automated highway driving. *Procedia Manufacturing*, *3*, 2784–2791. https://doi.org/10.1016/j.promfg.2015.07.735

Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *i-Perception*, *4*, 213–228. https://doi.org/10.1068/i0532

Leone, L. M., & McCourt, M. E. (2015). Dissociation of perception and action in audiovisual multisensory integration. *European Journal of Neuroscience*, *42*, 2915–2922. https://doi.org/10.1111/ejn.13087

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical

Turk. *Behavior Research Methods*, *47*, 519–528. https://doi.org/10.3758/s13428-014-0483-x

Madison, G., Mosing, M. A., Verweij, K. J., Pedersen, N. L., & Ullén, F. (2016). Common genetic influences on intelligence and auditory simple reaction time in a large Swedish sample. *Intelligence*, *59*, 157–162. https://doi.org/10.1016/j.intell.2016.10.001

Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLOS ONE*, *8*, e71608. https://doi.org/10.1371/journal.pone.0071608

Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car. In *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 514–521). Iowa City: University of Iowa.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Attention, Perception, & Psychophysics*, *40*, 331–343. https://doi.org/10.3758/BF03203025

Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology*, *46*, 101–151. https://doi.org/10.1016/S0010-0285(02)00517-0

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review*, *80*, 489–509. https://doi.org/10.1037/h0035437

Petermeijer, S. M., Bazilinskyy, P., Bengler, K. J., & De Winter, J. C. F. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics*, *62*, 204–215. https://doi.org/10.1016/j.apergo.2017.02.023

Petermeijer, S., De Winter, J. C. F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, *17*, 897–907. https://doi.org/10.1109/TITS.2015.2494873

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532.

Reimers, S., & Stewart, N. (2015). Presentation and response time accuracy in Adobe Flash and HTML5/Javascript Web experiments. *Behavior Research Methods*, *47*, 309–327. https://doi.org/10.3758/s13428-014-0471-1

Reimers, S., & Stewart, N. (2016). Auditory presentation and synchronization in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, *48*, 897–908. https://doi.org/10.3758/s13428-016-0758-5

Schubert, T. W., Murteira, C., Collins, E. C., & Lopes, D. (2013). ScriptingRT: A software library for collecting response latencies in online studies of cognition. *PLOS ONE*, *8*, e67769. https://doi.org/10.1371/journal.pone.0067769

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*, 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266. https://doi.org/10.1038/nrn2331

Todd, J. W. (1912). Reaction to multiple stimuli. In R. S. Woodworth (Ed.), *Archives of psychology, XXI*. Columbia Contributions to Philosophy and Psychology. New York: Science Press.

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement-Good news for bad clocks. *British Journal of Mathematical and Statistical Psychology*, *42*, 1–12. https://doi.org/10.1111/j.2044-8317.1989.tb01111.x

Van der Stoep, N., Spence, C., Nijboer, T. C. W., & Van der Stigchel, S. (2015). On the relative contributions of multisensory integration and crossmodal exogenous spatial attention to multisensory response enhancement. *Acta Psychologica*, *162*, 20–28. https://doi.org/10.1016/j.actpsy.2015.09.010

Wang, Y., Toor, S. S., Gautam, R., & Henson, D. B. (2011). Blink frequency and duration during perimetry and their relationship to test-retest threshold variability. *Investigative Ophthalmology & Visual Science*, *52*, 4546–4550. https://doi.org/10.1167/iovs.10-6553

Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, *34*, 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Wickens, C. D. (2001, March). *Attention to safety and the psychology of surprise*. Paper presented at the 2001 International Symposium on Aviation Psychology, Columbus, OH.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33–39. https://doi.org/10.1111/1467-9280.00006

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the Internet: A tutorial review. *PeerJ*, *3*, e1058. https://doi.org/10.7717/peerj.1058

Pavlo Bazilinskyy concluded an Erasmus Mundus Double MSc in dependable software systems with double distinction at the University of St Andrews (Scotland) and Maynooth University (Ireland) in 2014. He is currently a Marie Curie research fellow in the BioMechanical Engineering Department of the Delft University of Technology.

Joost de Winter is a human factors scientist with interest in automated driving. He is an associate professor in the Department of BioMechanical Engineering of the Delft University of Technology.

*Date received: August 6, 2017*
*Date accepted: June 10, 2018*