Running head: CROSS-MODAL INFLUENCES ON AUTOMOTIVE ICONS

Does a "stoplight!" improve processing a stoplight?
Cross-modal influences of time-compressed spoken denotations on automotive icon classification

Angela Mahr[1] & Dirk Wentura[2]

[1] German Research Center for Artificial Intelligence (DFKI)

[2] Saarland University

---- Journal of Experimental Psychology: Applied, in press ----

Author Note

Angela Mahr, German Research Center for Artificial Intelligence (DFKI), Saarbrücken; Dirk Wentura, Department of Psychology, Saarland University, Saarbrücken, Germany.

Correspondence concerning this article should be addressed to Dirk Wentura, Department of Psychology, Saarland University, Campus A2 4, 66123 Saarbrücken, Germany, e-mail: wentura@mx.uni-saarland.de.

Abstract

Findings from three experiments support the conclusion that semantic auditory primes facilitate processing of complex warning icons in the automotive context. In Experiment 1, we used a cross-modal icon identification task with auditory primes and visual icons as targets, presented in a high perceptual load context. Responses were faster for congruent priming in comparison to neutral or incongruent priming. This effect also emerged for different levels of time-compression of auditory primes. In Experiment 2, participants took part in a driving simulation with target icons on a gantry road sign. Participants had to categorize the color of the icons. Again, compressed auditory primes facilitated responses in cases of congruency (compared to incongruent and neutral primes). This result was replicated in Experiment 3 with more complex responses (i.e., braking, switching lanes). Our results suggest semantic object-based auditory-visual interactions, which rapidly increase the denoted target object's salience.

Keywords: cross-modal, semantic priming, response priming, Stroop, time compression

Public significance statement

This study suggests that spoken words, even if they are time-compressed to 50% of their normal length, facilitate visual processing of corresponding pictorial stimuli in a simulated driving context. Based on these findings, the authors recommend speech warnings in automotive contexts due to their general potential to automatically affect visual processing.

Cross-modal influences of time-compressed spoken denotations on automotive icon classification

Nowadays, advanced driver assistance systems offer a multitude of information about the road environment which can be used to support drivers in critical road situations. However, the majority of these projects have focused on technical and standardization issues, whereas issues related to human factors have received little attention (for a review, see Strandén, Uhlemann, & Ström, 2008). Accordingly, the design of appropriate human-machine interfaces for specific information transfer from technical systems to the driver in time-critical situations is still an open issue. For example, using a simple tonal warning (i.e., a beep) in combination with a display showing textual information is a common practice (see, e.g., Campbell, Richard, Brown, & McCallum, 2007; Cao et al., 2010; Lee, McGehee, Brown, & Reyes, 2002; Spence, 2012). Even though this concept is easy to implement and can be applied to many different types of use cases, it would obviously not be beneficial to adopt this strategy in complex road scenarios in which a driver needs to reliably recognize which out of numerous time-critical warnings is being presented. The driver would be forced to check the display to disambiguate the alert tone, drawing his or her gaze off the road at just the moment when it is necessary to assess the external situation.

In contrast, transferring auditory semantic information about the critical incident would probably be beneficial for instantaneous situation assessment and response selection. Thus, with our present research we want to contribute to the question whether auditory semantically meaningful stimuli (i.e., spoken words) facilitate processing of visual signals in an automotive context. To elucidate with an everyday example: If a co-driver suddenly yells: “stoplight!” (because he thinks that the driver is inattentive), will the driver better recognize the stoplight?

We will focus on spoken words that have an intrinsic relationship to the visual signals (instead of, e.g., abstract sounds with a to-be-learned meaning) because signals that are well mapped to their referents can be easily distinguished in their interpretation (see, e.g., McKeown & Isherwood, 2007). In the past, three basic experimental paradigms were utilized for cross-modal research using semantically meaningful auditory signals. We will give a brief synopsis of this research with a focus on those characteristics that are relevant for our studies.

Cross-modal cueing. A lot of research focused on spatial attention (for reviews, see Hillyard, Störmer, Feng, Martinez, & McDonald, 2016; Spence, 2010), utilizing variants of the cueing paradigm (Jonides, 1981; Posner, Snyder, & Davidson, 1980). Most important in the present context are the studies by Ho and Spence (2005, 2006; but see also Ho, Tan, & Spence, 2006). Ho and Spence (2005) presented the spoken words “front” and “back” (presented by a loudspeaker located beneath the participant) in a simulated driving context which were predictive for the location of a critical scene (i.e., either a fast approach by the car behind, seeable through the rearview mirror, or a fast approach toward the car in front) and the appropriate response (i.e., acceleration versus braking). There was indeed a cueing effect, that is, responses were faster in the valid condition (i.e., location word and location of the critical scene matched; 80% of all trials) compared to the invalid condition (i.e., a non-match of word and location). However, as said, the cue was not only predictive for the location but for the appropriate response as well (i.e., the effect can be understood as if the words had directly triggered the corresponding responses). To relate this to our everyday example: The “stoplight!” of the co-driver should not result in an unconditional braking response, but in a facilitated processing of the visual scene.

To overcome this critical issue, Ho and Spence (2006) conducted a follow-up study, now in a basic research setting. They used again meaningful auditory cues (“left” vs. “right”) in a spatial cueing task with visual targets, which were presented either left or right from fixation. The task was a discrimination task (i.e., whether the target was a *6* or *9*) which was unrelated to the cueing variation. Moreover, the cues were now non-predictive, that is, the number of valid and invalid trials was balanced. This is an important feature to see whether cueing works by mere semantic content. For briefly presented, but clearly seeable targets, there was no cueing effect. However, the effect was found for the accuracy of detecting *masked* targets. Thus, there is evidence that attention is redirected in line with the content of a spoken word. What is, however, not answered by this research is whether auditory messages facilitate processing of meaningful visual stimuli.

Cross-modal Stroop color-naming. The second line of research comes from visual-auditive Stroop studies (Stroop, 1935). In these experiments, the color of a visually presented stimulus has to be named or categorized by keypress. The stimulus is accompanied by an auditory distractor that either named the target color (congruent condition), a different color (incongruent condition), or is neutral with regard to color; usually, responses are faster in the congruent compared to the incongruent condition (see, e.g., Cowan & Barron, 1987; Elliott, Cowan, & Valle-Inclan, 1998; Elliott et al., 2014; Roelofs, 2005). For two reasons, with regard to the goal of the present experiments, we want to highlight a recent study by ourselves (Mahr & Wentura, 2014). First, we used time-compressed spoken words. Thinking about in-car warning systems, in time-critical driving situations the system could present time-compressed speech to quickly inform the driver about the current safety issue or suggest a specific action. It goes without saying that savings of even a fraction of a second can sharply reduce accident rates (see, e.g., Spence, 2012, for a discussion of that point). Second, we combined the Stroop task with a variation of perceptual load (Lavie, 1995, 2005). Driving contexts are likely to be characterized as a situation with high perceptual load. In detail, participants were presented with a colored target circle in a ring-like arrangement along with several other circles, either colored in non-target colors (i.e., high perceptual load context) or grey-shaded (i.e. low perceptual load), and instructed participants to quickly categorize the target color via a key-press. Shortly before the target display (i.e., with a stimulus onset asynchrony, SOA, of 100 ms), an auditory prime started which was either uncompressed (i.e., 400 ms long), compressed to 30% (i.e., 120 ms), or compressed to 10% (i.e., 40 ms). In the high perceptual load condition, large congruency effects (i.e., response times being faster in congruent conditions compared to incongruent ones) were found even for 30% compression; for 10% compression, the effect was still significant and had a medium size. Thus, a critical feature of cross-modal studies using auditory stimuli, that is, their typically long duration with regard to time-critical paradigms, was mitigated.

Of note, congruency effects were smaller in the low perceptual load context, thereby replicating similar results found by Tellinghuisen and Nowak (2003; see also recently Tellinghuisen, Cohen, & Cooper, 2016). This was remarkable because uni-modal (i.e., visual) effects in comparable designs (Lavie, 1995, 2005) as well as cross-modal studies exploring the phenomenon of inattentional deafness (i.e., a single in principle noticeable, but unexpexted and task irrelevant auditory stimulus was not noticed by participants; see, e.g., Macdonald & Lavie, 2011; Raveh & Lavie, 2015; see Murphy & Greene, 2015, 2017, for applications in a driving context) show exactly the opposite pattern.

With regard to applied contexts, this feature indicates that auditory primes are especially of help in a noisy environment (which comes closer to real driving contexts than the low load condition). However, Stroop effects might be driven by response competition processes (i.e., prime and target compete for access to the response system) or by semantic facilitation processes (i.e., the prime facilitates the encoding of congruent targets; see De Houwer, 2003; Wentura & Degner, 2010). If the former explanation applies, these processes are of limited usefulness in an applied context because the auditory stimulus does not really help to process the visual scene. In contrast, if the latter comes true, this would be an especially attractive feature for applied contexts because it would show that auditory primes help people to detect an important event in the outside visual world.

Semantic priming. An unequivocal test for semantic facilitation processes is the semantic priming paradigm (Meyer & Schvaneveldt, 1971; for a review, see, e.g., McNamara, 2005). Chen and Spence (2011, 2017) tested crossmodal semantic priming via naturalistic sounds (e.g., a barking dog) and spoken words. The task of participants was to detect the presence of briefly presented and masked visual targets (line drawings; e.g., of a dog) in target-present trials (and to avoid false alarms in target-absent trials). Target-congruent sounds and words enhanced object detection as compared with incongruent sounds/words. However, effects with words were found with rather long SOAs (1000 ms) and only given some specified conditions at a medium SOA condition (350 ms).[1] Results by Lupyan and Ward (2013) corroborated the target detection effect with an SOA of app. 1100 ms. We mention this detail because from uni-modal visual priming studies it is known that long-SOA effects might be driven by non-automatic expectancy-based processes (Neely, 1977). Of course, direct transfer of this finding to cross-modal research is problematic and even if the transfer holds, this does not mean that the long-SOA effects found in previous cross-modal semantic priming experiments must be non-automatic. However, we are interested to see whether short-SOA effects can be found as more unequivocal evidence for automatic cross-modal semantic facilitation effects.

[1] These conditions were: adding a picture identification task after the object detection task (Chen & Spence, 2011, Exp. 4B); the 350ms-SOA-block subsequent to a 1000-ms-SOA-block (Chen & Spence, 2017). A further study (Chen & Spence, 2013) suggests *prima facie* effects with shorter SOA. But note that participants had the task to categorize the target as depicting either an animal or a non-animal. Since incongruent priming was always realized by a stimulus of the other response category (e.g., *guitar* as prime for *dog*), the design had the basic characteristic of a response priming paradigm (i.e., the prime is either compatible or incompatible to the correct target response) and cannot be directly compared to the other experiments.

Both the studies by Chen and Spence (2011, 2017) as well as Lupyan and Ward (2013) tested the hypothesis of a cross-modal semantic priming with a variety of different stimuli (as is usual in semantic priming research). In our earlier study (Mahr & Wentura, 2014), we took a somewhat different route. Since our Stroop effects (see above) were dominant facilitation effects, we wondered whether they are mainly driven by semantic facilitation processes. Therefore, we changed the basic paradigm (see above) to a semantic priming paradigm in our third experiment: Now, participants only had to categorize whether or not one of the four target colors was present in the (high load) display. Thus, response-related processes could no longer explain congruency effects. Nevertheless, a clear congruency effect was observed, given a SOA of 100 ms with time-compressed prime stimuli (30%; see above) and a high perceptual load context (see also Tellinghuisen et al., 2016, Exp. 3, for a related result). Thus, at least with a restricted and very simple set of stimuli, visual-auditive semantic priming effects can be found with a very short SOA.

Taken together, there is evidence for facilitatory effects of spoken words on perceiving visual events and responding to them. Moreover, there is evidence that these facilitatory effects are more pronounced in a complex visual environment and that these effects can be found with time-compressed auditory stimuli that allow for a use of a short SOA.

With the present experiments we want to transfer what we have found with our former experiments into an automotive context. Do findings of cross-modal Stroop/priming effects achieved with simple color stimuli apply equally to more complex objects? And do they apply in a more realistic environment, that is, a driving context? For the automotive context, it is important to show that traffic-related objects can be primed by their denotations. For this purpose, we switched to icons from the automotive context and their spoken two-syllable denotations in the present experiments.

Overview

In Experiment 1, we adapted Mahr and Wentura's (2014) cross-modal Stroop task to automotive icons to show that the basic results are replicated with these stimuli. Experiments 2 and 3, however, lie at the heart of the present article for two reasons: First, we transferred the target detection task to gantry road signs that repeatedly occurred while driving in a driving simulation environment. Second, we used a semantic priming design. That is, in each trial, one target icon appeared along with three non-target icons on the gantry in either red or green color; participants had to categorize the color as quickly as possible. Thus, a priming effect (induced by a congruent auditory prime) can only be explained by facilitated detection and encoding of the target icon and not by response interference and/or facilitation as in the Stroop-like task of Experiment 1. In Experiment 2, we recorded responses to targets simply via steering wheel buttons. In Experiment 3, we introduced more natural response modes: depending on the target color, participants had to either brake or merge into a designated lane. Moreover, in Experiment 3 we additionally varied the contingency of the prime-target relation. Whereas in one condition, the prime was not informative with regard to the upcoming target, it predicted the target in the majority of trials in a second condition.

Experiment 1

Experiment 1 was designed to find cross-modal Stroop effects with more complex time-compressed auditory primes and visual target objects (compared to color words and colored circles). We applied three different levels of spoken word duration: In addition to the benchmark of an uncompressed version of each word, we used a 50% and a 30% compression rate. Thus, we discarded the 10% condition of Mahr and Wentura (2014) because after compressing the two-syllable words down to 10% of their original duration, it was actually impossible to recognize which denotation they comprised. Therefore, we decided to replace the most extreme level of time compression with an intermediate one (50% of the original duration) to explore what might be the best choice for the subsequent experiments. We expected Stroop effects (i.e., faster responses for congruent compared to incongruent conditions) for all levels of compression.

We confined the design of Experiment 1 to what was the high perceptual-load condition in the earlier experiments (i.e., the target icons were presented in the context of other non-target icons which share characteristics of the target set). Our focus on a rather difficult visual search task not only aligned with previous findings, but also coincided with practical considerations: Given our applied research question, it was of special interest whether speech warnings particularly support drivers in highly complex and rather ambiguous situations. But note, that due to not varying perceptual load in the present experiments, we cannot make any claim about the causal role of perceptual load.

Method

Participants. A total of 19 students (16 females, 3 males; age range 18 – 28 years, median = 22 years) from Saarland University took part for course credit. All participants had normal or corrected-to-normal vision, and none reported any color blindness or hearing problems.

The Stroop effects found in Mahr and Wentura (2014) for “no compression” and “30% compression” in the high perceptual load condition were $d_Z = 0.89$ and $d_Z = 1.27$, respectively. To detect an effect of $d_Z = 0.89$ with a probability of $1 - \beta = .95$, an α-value of .05 (two-tailed), a sample size of 19 participants was required (calculations were done using G.Power 3.1.3; Faul, Erdfelder, Lang, & Buchner, 2007).

Design. A 3 (semantic congruency of auditory prime and target icon: neutral, congruent, incongruent) × 3 (duration of the auditory prime: 100% [i.e., 700 ms], 50% [i.e., 350 ms], 30% [i.e., 210 ms]) design was used with all factors manipulated within participants. Technically, the congruency factor was realized using a 5 (auditory prime type: traffic light, tractor, ambulance, children, and neutral) × 4 (visual target type: traffic light, tractor, ambulance, children) design, resulting in non-contingent auditory priming (i.e., the probability that a congruent target follows the auditory prime is 25%). That is, participants did not benefit from listening to the words, since they did not contain any information about the subsequent target (see Logan & Zbrodoff, 1979; Mordkoff, 2012, with regard to this issue). Congruency was manipulated on a trial-by-trial basis, while compression level was manipulated blockwise, resulting in a total of three blocks, each of which comprised 100 trials: 20 congruent, 60 incongruent, and 20 neutral trials. The order of compression level presentation was counterbalanced among the participants.

Materials. Each target display contained eight visual stimuli presented in a ring-like arrangement on a white background (see Figure 1). The ring was presented with a visual angle of 18.9° (diameter 400 pixels ≙ 20 cm), with each icon (40-72 pixels wide × 50-68 pixels high) centered on a square white patch (size 73 × 73 pixels ≙ 3.6 × 3.6 cm) spanning 3.4°. One of these visual stimuli was one of the four target objects (traffic light, tractor, ambulance, or children; see *Appendix B*). Seven filler icons (motorcycle, stop sign, truck, bicycle, cow, fallen tree, and roadwork; see *Appendix B*) were clearly distinguishable from the target icons, and randomly located at the seven remaining non-target positions in each trial.

We generated auditory word stimuli by having a male speak the words several times and recording them with a Sennheiser condenser microphone (ME104; 16-bit mono, 44100-Hz digitization). Audacity (for Windows) served as the recording software and a MOTU (8pre) sound card as the hardware. We selected sound files lasting exactly 700 ms,[2] which was to be the average length of the two-syllable words traffic light (“Ampel”, [ˈampl̩]), tractor (“Traktor”, [ˈtʁaktoːɐ̯]), ambulance (“Notarzt”, [ˈnoːtʔaʁt͡st]), and children (“Kinder”, [ˈkɪndɐ]). The German word for ‘object’ (“Objekt” ([ɔpˈjɛkt]) served as the neutral prime.

[2] Each stimulus word was spoken (and recorded) about 30 times by the same male person. For each auditory stimulus we selected the file that lasted for exactly 700 ms (from word onset until word offset).

We created two different time-compressed versions of each of the five sound files: One with a length of 350 ms (i.e., a compression to 50% of the original duration), and another with a length of 210 ms (i.e., a compression to 30% of the original duration). The 50% compression files were still understandable, whereas the 30% compression files were hardly discriminable outside the context of the given task: That is, arbitrary 30% compressed two-syllable words without any constraining expectations are not identifiable; however, in the present context with only five known words used, categorization is rather easy after hearing the samples.

Time compression of the stimuli was performed with the Praat linguistic freeware program (Boersma & Weenink, 2011) using the PSOLA ("pitch-synchronous-overlap-add") algorithm, which ensured phonologically adequate time compression with unchanged pitch. The sounds were presented over closed-ear headphones (TerraTec headset master 5.1 USB) and ranged in volume from 67 to 72 dB SPL.

Procedure. Participants were individually tested in a sound proof chamber. They were seated in front of a 15-inch monitor (60 Hz refresh rate, resolution 640 × 480 pixels) controlled by a personal computer in an experimental cabin with dimmed light. The participants' viewing distance was about 60 cm. The experiment was conducted using E-prime software (E-prime 1.1). In each trial of the experiment, participants had to categorize the target icon within the ring-like arrangement of eight items by pressing the appropriate key (keys d, f, j, and k on a standard keyboard). Participants were informed that the auditory words preceding the target display were non-informative.

To start each block, participants pressed the space bar. In each trial, following a 1000 ms blank (black) screen, a white fixation cross appeared for 500 ms, followed by a blank screen for another 250 ms. Subsequently, the target screen was presented until a response was given (see Figure 1). The auditory prime presentation started with an SOA of 100 ms.

Four warm-up-trials (not included in the analyses) preceded each block. As an introduction to the experimental task, participants completed three practice phases. In Phase 1, they were simply presented with one of the four target icons in a given trial (four times each) and had to learn the response key assignment. To support this process, stickers with the four target icons were affixed to their left and right index and middle fingers. Written feedback (i.e., "correct", "wrong") was provided on each trial. Phase 2 had 22 trials, and participants had to categorize icons (again presented as stand-alone stimuli) as either one of the four target icons (using the four response keys) or one of the filler items (by pressing the space bar). Feedback was given on each trial. The final practice phase comprised 30 trials identical to the trials of the main experiment. The only difference was that feedback was provided on each trial to ensure participants had understood the task and were responding to visual targets only.

Each of the three experimental blocks consisted of 100 trials composed of 20 neutral, 20 congruent, and 60 incongruent trials, randomly intermixed. The trial list was completely balanced with regard to the four icons (i.e., each of the 16 possible target-prime combinations was presented five times, and each target icon was used five times in the neutral condition). In order to control for response repetition effects, target icons were not repeated in two directly adjacent trials. Participants took self-timed breaks between blocks. The experiment lasted approximately 45 minutes.

Results

Error trials (4.3%) and outliers (3.4%; i.e., RTs in correct trials that were below 250 ms or were 1.5 interquartile ranges above the third quartile with respect to the individual distribution; Tukey, 1977) were discarded. Table 1 shows the mean RTs for the conditions of our design.

RTs were analyzed in a 3 (semantic congruency: neutral vs. congruent vs. incongruent) × 3 (duration: 100% vs. 50% vs. 30%) repeated measures MANOVA (see, e.g., O'Brien & Kaiser, 1985). For tests involving the congruency factor, we additionally report the results for two a priori planned orthogonal contrasts (Helmert). The contrast of dominant interest is the contrast between the congruent and incongruent conditions. It is the decisive contrast with regard to whether there is a congruency effect or not. However, although of secondary importance, the other contrast is of interest as well. It is the contrast of the neutral condition compared to the congruent and incongruent conditions collapsed. Statistically, it indicates whether the neutral condition significantly deviates from the average of congruent and incongruent conditions, or in other words: whether it deviates from the midpoint of the distance between the incongruent and the congruent mean. Thus, if we interpret the distance between incongruent priming and neutral priming as “interference” and the distance between neutral priming and congruent priming as “facilitation”, the contrast indicates whether facilitation and interference should be considered of equal magnitude or whether they diverge in magnitude.

The analysis revealed a significant main effect for the semantic congruency manipulation, $F(2, 17) = 35.87, p < .001, \eta_p^2 = .808$, which is dominantly due to the contrast congruent versus incongruent condition, $F(1, 18) = 75.70, p < .001, \eta_p^2 = .808$. The other orthogonal contrast (i.e., neutral vs. congruent/incongruent collapsed; see above) was non-significant, $F(1, 18) = 2.58, p = .125, \eta_p^2 = .125$, thereby indicating that facilitation and interference are not significantly different (but see also below). The interaction of semantic congruency condition and duration was not significant, $F < 1$, nor was the main effect of duration, $F(2, 17) = 1.82, p = .192, \eta_p^2 = .176$. Post-hoc tests (using Bonferroni-Holm alpha adjustments) showed all three congruency effects (i.e., the mean difference between the incongruent and the congruent condition for the three durations) to be significantly above zero, *t*s(18) > 4.63, *p*s < .001 (see Figure 2).

In addition, difference scores for facilitation (i.e., neutral - congruent) and interference (i.e., incongruent - neutral) were calculated for all three duration conditions (see Figure 2). We analyzed these indicators in two separate repeated measures MANOVAs (i.e., one for facilitation, one for interference) with regard to the duration factor (100% vs. 50% vs. 30%). For facilitation, there was a significant constant effect overall (i.e., the mean facilitation score averaged across duration conditions was significantly above zero), $F(1,18) = 26.12$, $p < .001$, $\eta_p^2 = .592$, but we did not find a significant main effect of duration, $F < 1$. Similarly, overall interference was significantly above zero, $F(1, 18) = 9.81$, $p = .006$, $\eta_p^2 = .353$, without revealing a significant main effect of duration, $F < 1$. Descriptively, facilitation appears to be more pronounced than interference (see also Figure 2); however, as noted above, this difference is not significant.

We conducted the same analysis with the error rates (see Table 1) as with the RTs. The 3 (semantic congruency) × 3 (duration) repeated measures MANOVA did not show any contradictory results to that of the RTs. There were no significant effects: $F(2,17) = 2.47$, $p = .12$, $\eta_p^2 = .225$, for the main effect for duration; $F < 1.07$ for the remaining effects. Despite the non-significance of the MANOVA results for errors, we tested the three simple congruency effects for the three duration levels in order to provide full transparency with regard to potential speed-accuracy trade-offs. None were significant (all |*t*|s(18) < 1.25, *p*s > .227).

Discussion

We investigated whether spoken object denotations influence the categorization of automotive target icons even when they do not reveal any valid information. The significant and large difference in RTs between congruent and incongruent trials in all experimental conditions clearly support this assumption. Another remarkable point is the pattern of cost-benefit partitioning of the Stroop effect. In accordance with our former findings (Mahr & Wentura, 2014), rather large benefits and rather minor interference were revealed. (However, differences were not significant.) Furthermore, analogously to the former experiments, no main effect of duration on RTs was identified. With regard to time compression, it is worth noting that we found clear cross-modal effects of semantic congruency independently of the duration of spoken word primes. Accordingly, for non-contingent presentation of spoken word stimuli, time compression seems to be neither a beneficial nor a detrimental factor.

To sum up, we found cross-modal semantic Stroop/priming effects with complex materials, namely two-syllable spoken words as auditory primes and automotive icons as visual targets. We applied two reasonable time compression levels to the irrelevant primes, but no difference in effects was revealed. Overall, our results are promising since they indicate that cross-modal semantic priming is effective even for more complex objects and their denotations. This constitutes a major step towards critical real-world road scenarios.

For the forthcoming experiment, we specifically designed and implemented a simulated driving task that maintained the typical trial-based structure of our former computer experiments. This enabled us to leave many experimental parameters unchanged, while at the same time investigating whether cross-modal effects of spoken words could be found during a dynamic driving task comprising continuous visual and motor demands.

Moreover, in Experiment 2 we switched the paradigm from a Stroop task to a semantic priming paradigm. Now, merely the color of an automotive target icon should be categorized, thus keeping response mode orthogonal to the variation of congruency of primes and targets.

Experiment 2

For the new simulated driving scenario, we transferred the target detection task to gantry road signs that repeatedly occurred while driving (see Figure 3). We recorded responses to targets simply via steering wheel buttons. This allowed us to retain our hitherto well-examined procedure and materials, while at the same time switching to a dual task situation and optical flow conditions. In order to separate response competition processes from semantic facilitation processes, we switched from target identification to target feature identification: participants simply had to indicate the color of the present target icon. With this, we were able to explore the effects of spoken word stimuli on visual perception and attention within a RT paradigm. For Experiment 2, we decided to use a time compression level of 50% of the original duration for the auditory primes. This corresponds to the medium duration of auditory primes in Experiment 1, with primes remaining well intelligible. In addition to a neutral prime word, we added a silence control condition to test for possible interference effects by the mere presence of an auditory prime.

Method

Participants. The participants were a new group of 25 students (13 females, 12 males; age range 20 – 31 years, median = 22 years) from Saarland University who were paid 8 Euro for participation. All had normal or corrected-to-normal vision and none reported color blindness or hearing problems. All participants had possessed a valid driver's license for at least two years.

The congruency effect found in Experiment 3 of Mahr and Wentura (2014) was $d_Z = 0.85$. To detect an effect of $d_Z = 0.85$ with a probability of $1 - \beta = .95$, an α-value of .05 (two-tailed), a minimum sample size of 21 participants was required (according to G.Power 3.1.3; Faul et al., 2007).

Design. The priming factor (neutral silence vs. neutral word vs. congruent vs. incongruent) was technically realized by a 5 (auditory prime word: *traffic light*, *tractor*, *ambulance*, *children*, or neutral [prime word *object* or silence; see Procedure]) × 4 (visual target type: traffic light, tractor, ambulance, children) design. The conditional probability of a target appearing after a given prime was at chance level, resulting in non-contingent auditory priming. That is, participants again did not benefit from listening to the words.

Materials. The experimental track of the simulated driving scenario comprised a straight road (width 18.5 m) with five lanes next to each other (width 3.7 meters [m] each). Gantry road signs (height 7.3 m) were positioned every 210 m (see Figure 3). In line with the number of trials (172), the track was about 37 km long. A start and a goal symbol signalized the beginning and end of data recording. Each gantry road sign (height above street 18.5 m) contained four square displays (2.5 m × 2.5 m) to present the icons.

Each of the four displays showed one luminous automotive icon from the same set of targets and distractors as in Experiment 1 (see *Appendix B*). Each icon could be presented in red or green, and each trial had two red and two green icons in order to prevent participants from using strategies for target color identification. In most of the trials, one of the four target icons was presented on one of the four displays of the gantry road sign; in a minority of trials no target stimulus was presented at all (catch trials). In each trial, the remaining three (target present condition) or four (catch trials) icons were chosen from the set of seven filler items, which were also the same as in Experiment 1 (see also *Appendix B*). Auditory prime stimuli were identical to the medium compression rate (50%) used in Experiment 1.

The experiment was conducted and data collected using the open source driving simulation software OpenDS 1.0 (Math, Mahr, Moniri, & Müller, 2012). In order to generate different tracks for each participant in accordance with our requirements, we used the adjunct tool GenXDS before feeding the respective files into OpenDS.

Procedure. Participants were tested individually in a room with modest daylight. They were seated in front of a 19-inch monitor (60 Hz refresh rate, resolution 1280 × 1024 pixels) controlled by a personal computer. The participants' viewing distance was about 70 cm and they wore closed-ear Sennheiser PC 151 headphones. The auditory stimuli ranged in volume from 70 to 72 dB SPL. A MiMo game steering wheel and pedals were used to control the vehicle within the simulation.

Throughout the experimental track of the simulated driving scenario, participants drove at a constant speed of 100 kilometers per hour (km/h), leading to time intervals of about 7.6 seconds between every two gantry road signs. Their task was to stay in the center of the middle lane and react to target icons appearing on the gantry road signs (see Figure 3b). The icons appeared sixty meters before each gantry road sign and remained visible until participants passed under the sign (about 2.2 s). For each gantry road sign encountered, participants had to decide by key press whether a red target object (“target red” button on the steering wheel with right thumb) or green target object (“target green” button on the steering wheel with left thumb) was presented at one of the four positions on the gantry road sign. In target absent trials, no answer was to be given.

Participants could respond from the onset of the target display up until the next trial started (maximum response time: 7.5 s). Auditory primes were presented just before the visual target onset with an SOA of 100 ms. The presentation duration for the auditory primes was 350 ms (i.e., 50% compression). Participants were informed that the auditory words were non-informative.

As an introduction to the experimental task, participants completed an initial practice phase to familiarize themselves with the simulator. They drove on a straight road without road signs present. Afterwards, participants were given the opportunity to familiarize themselves with the four target icons (each in red and green) presented on a static screen. Subsequently, a short practice track containing 20 sample trials had to be navigated. This short practice track was similar to the subsequent main experimental track; accordingly, all auditory prime types were included.

The main experimental track comprised 160 trials with a target icon present (32 congruent trials, 96 incongruent trials, and 32 neutral trials). Each non-neutral auditory prime type was paired eight times with each target icon, four times with a red target and four times with a green target. The neutral trials consisted of 16 trials (i.e., twice with each icon in red and twice in green) with a control word ("Objekt" [object], as in Exp. 1), and 16 trials (same division as above) with silence instead of a prime word. In twelve catch trials, no target icon was present; each prime type (four target icon denotations, neutral word, silence) was presented twice in these catch trials. Thus, overall the track comprised 172 trials, randomly intermixed, except for the condition that the target icon not be repeated in two directly adjacent trials. The track started with four additional warm-up trials which were not included in the analyses. The experiment lasted for approximately 40 minutes.

Results

Prior to aggregation, we discarded error trials (3.5% incorrect reaction, 0.4% no reaction) and individual outliers (2.6%; see Experiment 1). Table 2 shows mean RTs and error rates for all conditions.

The RTs for the four semantic congruency conditions (neutral silence vs. neutral word vs. congruent vs. incongruent) were analyzed in a repeated measures MANOVA, which revealed a significant main effect, $F(3, 22) = 19.30, p < .001, \eta_p^2 = .725$. A priori Helmert contrasts further showed, first, that trials without prime word presentation (i.e., neutral silence) did not significantly differ from trials with prime words (i.e., the mean of all other trial types), $F(1,24) = 2.98, p = .097, \eta_p^2 = .110$. Looking at Table 2 indicates that neutral silence is at the same level as the incongruent and the neutral word condition. Thus, there is no indication for interference by the mere presence of an auditory signal. Second, the neutral prime word trials did significantly differ from trials with relevant prime words (i.e., congruent and incongruent pooled), $F(1,24) = 6.40, p = .018, \eta_p^2 = .211$. Note, that this result indicates the imbalance of facilitation and interference (see below). Third, the contrast of main interest (congruent vs. incongruent trials) revealed a significant difference, $F(1,24) = 59.33$, $p < .001$, $\eta_p^2 = .712$ (see Figure 4).

The second Helmert contrast (i.e., neutral versus congruent/incongruent collapsed) indicates that facilitation and interference significantly diverge. Difference scores for facilitation (i.e., RT neutral control - RT congruent) and interference (i.e., RT incongruent – RT neutral control) were calculated (see Figure 4). Facilitation significantly differed from zero, $t(24) = 4.69$, $p < .001$, whereas interference did not, $t(24) < 1$.

For the error rates (i.e., incorrect reaction or missing reaction, see Table 2), a repeated measures MANOVA (semantic congruency: neutral silence vs. neutral word vs. congruent vs. incongruent) revealed no significant overall effect of semantic congruency, $F(3, 22) = 1.99$, $p = .15$, $\eta_p^2 = .21$. In addition, the positive difference between error rates in incongruent and congruent trials speaks against a potential speed-accuracy trade-off.

Discussion

When presenting time-compressed spoken denotations of objects just before their appearance, we could find faster responses in trials with a congruent cross-modal semantic presentation compared to incongruent or neutral trials. Moreover, there was no general impairment of performance by presenting auditory stimuli, as can be seen by the comparison of silent trials with neutral and incongruent trials (see Table 2).

These findings are remarkable for several reasons. First, the effects seem to be robust and pronounced enough to occur even under simulated dynamic driving conditions. Second, since we controlled for stimulus-response compatibility effects in this experiment, our results imply that semantic priming rather than response priming seems to be the major cause of our cross-modal effects. We could confirm clear cross-modal semantic facilitation of visual processing in a RT paradigm.

In Experiment 3, we changed the experimental setting in two ways. First, we altered the response type from a simple key press to demanding driving maneuvers: Depending on the target color, participants had to either brake or merge into a designated lane. The introduction of such complex motor responses constitutes another controlled step towards in-car warning scenarios, in which drivers might have to respond appropriately to imminent road hazards. Our claim that previous findings are relevant for critical road incidents would only be able to be maintained if driving performance enhancement via congruent speech warnings were to be confirmed in this new setting.

Second, we varied the contingency between prime and target category in a between-participants design. Whereas the "no contingency" condition was identical to Experiment 2 (in all aspects except the new response types), in the "high contingency" condition, we increased the ratio of congruent to incongruent trials. With regard to the applied context, it is clear that real in-car warning systems will (should) provide valid warnings which mostly match the upcoming road incident. For example, Ngo, Pierce, and Spence (2012) showed in an air traffic simulation task (i.e., a kind of radar screen had to be observed) that auditory signals (i.e., a beep) which were completely redundant to a visual signal (i.e., an aircraft's symbol turned into red if it was near collision) helped to respond faster to this event. For the same reason, Ho and Spence (2005) compared exogenous spatial cueing (by a car horn) that was non-predictive (50% valid trials in a two-location task) with predictive cueing (80% valid trials) and found larger effects for the latter. Accordingly, it is important for us to show that cross-modal semantic priming effects are at least not reduced when auditory warnings are highly redundant with visual icons.

Experiment 3

Method

Participants. Forty-six students (34 females, 12 males; age range 19 – 35 years, median = 24 years) from Saarland University participated in Experiment 3; they were paid 8

Euro for their participation. All had normal or corrected-to-normal vision and none reported color blindness or hearing problems. All participants had possessed a valid driver's license for at least two years. Data from two further participants had to be excluded due to remarkably lower accuracy rates (< 56%) than the remaining participants.

Design. We used a 4 (semantic congruency: neutral silence vs. neutral word vs. congruent vs. incongruent) × 2 (response type: braking vs. steering) × 2 (contingency: no vs. high) mixed design with congruency and response type varied within participants and contingency varied between participants. We decided in favor of a fixed assignment of color to response type (i.e., red targets were assigned "braking" and green targets were assigned "steering") because the meaning of red in driving contexts is so ingrained that the opposite assignment would have certainly created (non-informative) interference.[3] Moreover, although our intention was to explore priming effects for complex responses, we did not want to emphasize potential differences between the two types of complex responses.

The "no contingency" condition was realized exactly as in Experiment 2 (see also *Procedure*). For the "high contingency" condition, we increased the conditional probability of a congruent target appearing after a given prime word considerably above chance level (80%), thus resulting in highly contingent auditory priming (see *Procedure* for details).

Power planning was oriented on, first, replicating the overall priming effect in both contingency samples, second, having enough power to detect potential differences between the two samples. The first constraint is easily satisfied, since the congruency effect found in Experiment 2 was extremely large ($d_Z = 1.54$). Even a more conservative planning indicates that with n=23 (i.e., our subsample n), an effect of $d_Z = 0.80$ (i.e., a large effect according to Cohen, 1988) can be detected with power $1\text{-}\beta = .96$ (assuming $\alpha = .05$, two-tailed). With

[3] In Experiment 2, the priming effects for red and green targets did not significantly differ.

regard to the second constraint: With power 1-β = .80 (assuming α = .05, two-tailed), we were able to detect large between-participant effects (i.e., *d* = .84; Faul et al., 2007).

Materials and Procedure. Design, Materials, and Procedure were the same as in Experiment 2, except for the following details: Target color classification (red or green) indicated that drivers were to conduct the corresponding maneuver in the simulated driving scenario: braking or changing to the target icon lane, respectively. RTs based on driving performance in lane changes (until a steering wheel turning angle of 3 degrees from the straight position was exceeded) or braking (time until gas pedal was released) maneuvers constituted the dependent variable in this experiment. The only slight adaptation to the tracks was the distance between gantry road signs, which was slightly reduced to 200 m. A Logitech Driving Force GT steering wheel and pedals for throttle and brake were used. The general driving speed was reduced to a maximum of 60 km/h, resulting in a time window of about twelve seconds between gantry road signs. The icons appeared about 48.3 m before each gantry road sign remained visible until participants passed under the sign (minimum 2.9 s). In cases where the target icon was red, participants were to immediately brake and decelerate sharply to a target speed of 20 km/h or lower. In cases with a green target icon, they were to keep their speed and merge into the lane (over which the icon was presented). After the designated speed or lane was reached for 2,000 ms, a signal tone sounded and drivers were to speed up to 60 km/h again or change back to the middle lane. For a correct response, participants needed to accomplish the required maneuver within 5,000 ms after target display onset. If they reacted incorrectly, too late, or missed a reaction, the signal tone was still presented 7,000 ms after target onset.[4] This ensured that participants returned to the middle lane or accelerated back up to the standard driving speed before the next trial started.

[4] For the braking task, an incorrect reaction was recorded (a) if participants slowed down, but the speed remained above 20 km/h, (b) if they turned the steering wheel beyond three degrees, or (c) if they moved outside the lane markings of the middle lane. For the steering task, an

As an introduction to the experimental task, participants completed an initial practice phase to familiarize themselves with the simulator. They drove on a straight road for 4.3 km with a single icon (i.e., either a red circle with a black cross or a green circle with an arrow; see *Appendix*) presented on each of the 20 gantry road signs. Rules for responding (red → brake; green → change lane) corresponded to the main trials. Afterwards, participants were given the opportunity to familiarize themselves with the four automotive target icons, as in Experiment 2. Subsequently, they navigated another practice track containing 20 sample trials (4.3 km), which were equivalent to the subsequent main experimental track.

In the "no contingency" condition, the main experimental track comprised 172 trials (160 main trials plus 12 catch trials), as in Experiment 2. In the "high contingency" condition, the main experimental track comprised 164 trials (152 trials with a target icon present: 96 congruent trials, 24 incongruent trials, 32 neutral trials, and 12 catch trials). Each non-neutral auditory prime type was paired twenty-four times with the corresponding target icon (twelve times with a red target and twelve times with a green target) and twice with each non-corresponding target icon (once with each color). The neutral trials and catch trials were identical to those in the "no contingency" condition. In the "high contingency" condition, participants were informed that the auditory words predicted the target object in most cases. The experiment lasted for approximately 45 minutes.

Results

Prior to aggregation, we discarded error trials (i.e., false reactions or no reactions; 0.7%/0.5% for braking, 2.9%/1.9% for steering, for no/high contingency, respectively) and individual outliers (braking: 3.1%/4.0%; steering: 5.0%/4.4%, for no/high contingency,

incorrect reaction was recorded if (a) participants changed to an incorrect lane or (b) took their foot off the gas pedal. "No response" was recorded in target present trials if participants did not respond either correctly or incorrectly within the given time window.

respectively; see Experiment 1). Table 2 shows mean RTs and error rates for all conditions after averaging braking and steering performance.

We analyzed RTs in a 4 (semantic congruency: neutral silence vs. neutral word vs. congruent vs. incongruent) × 2 (response type: braking vs. steering) × 2 (contingency: no vs. high) mixed factors MANOVA for repeated measures. The analysis revealed a significant main effect of semantic congruency, $F(3,42) = 49.37$, $p < .001$, $\eta_p^2 = .779$, which was significantly moderated by contingency, $F(3,42) = 7.11$, $p = .001$, $\eta_p^2 = .337$. The Response type × Congruency interaction, $F(3,42) = 2.78$, $p = .053$, $\eta_p^2 = .165$, as well as the three-way interaction just missed the conventional level of significance, $F(3,42) = 2.76$, $p = .054$, $\eta_p^2 = .165$; $F < 1.13$ for the remaining effects.

Helmert contrasts further showed, first, that trials without prime word presentation (neutral silence) did not differ from trials with prime words (all other trial types), $F < 1$; this contrast was not moderated by response type, contingency, or response type × contingency, all $F < 1$.

Second, the neutral prime word trials did significantly differ from trials with relevant prime words (congruent and incongruent [pooled]), $F(1,44) = 23.00$, $p < .001$, $\eta_p^2 = .343$. Note, that this contrast tests for the (im-)balance of facilitation and interference (see below). Again, this contrast was not moderated by the other factors, all $F < 1.36$. Third, the contrast of main interest (congruent vs. incongruent trials) revealed a significant difference, $F(1,44) = 138.61$, $p < .001$, $\eta_p^2 = .759$. It was moderated by response type, $F(1,44) = 8.47$, $p = .006$, $\eta_p^2 = .161$, contingency, $F(1,44) = 20.25$, $p < .001$, $\eta_p^2 = .315$, and response type × contingency, $F(1,44) = 6.17$, $p = .017$, $\eta_p^2 = .123$.

For the “no contingency” condition, there were significant priming effects (see Figure 4) for braking, $F(1,22) = 72.73$, $p < .001$, $\eta_p^2 = .768$ ($d_Z = 1.78$), and steering, $F(1,22) = 9.22$, $p = .006$, $\eta_p^2 = .295$ ($d_Z = 0.63$); however, the former was significantly larger than the latter,

F(1,22) = 24.26, p < .001, η_p^2 = .524 (d_Z = 1.03). For the "high contingency" condition, there were significant priming effects (see Figure 4) for braking, F(1,22) = 102.03, p < .001, η_p^2 = .823 (d_Z = 2.11), and steering, F(1,22) = 55.02, p < .001, η_p^2 = .714 (d_Z = 1.55); these effects were not statistically different, F < 1. Both the braking as well as the steering priming effect were significantly larger in the high contingency condition compared to the no contingency condition, F(1,44) = 8.99, p = .004, η_p^2 = .170, and F(1,44) = 21.63, p < .001, η_p^2 = .330, respectively.

The second Helmert contrast (i.e., neutral word versus congruent/incongruent collapsed) indicates that facilitation and interference significantly diverge. Difference scores for facilitation (neutral control - congruent) and interference (incongruent - neutral control) were calculated. Figure 4 shows facilitation and interference scores for the conditions of interest. A 2 (response type: braking vs. steering) × 2 (contingency: no vs. high) mixed factors MANOVA with facilitation scores as the dependent variable yielded a significant constant effect (i.e., on average there was significant facilitation), F(1,44) = 88.06, p < .001, η_p^2 = .667. It was moderated by contingency, F(1,44) = 4.27, p = .045, η_p^2 = .089; that is, facilitation was slightly larger for the high contingency condition, with all other Fs < 1.07.

A 2 (response type: braking vs. steering) × 2 (contingency: no vs. high) mixed factors MANOVA with interference scores as the dependent variable yielded a non-significant constant effect, F(1,44) = 2.65, p = .111, η_p^2 = .057, but a main effect of contingency, F(1,44) = 5.43, p = .024, η_p^2 = .110. The interaction missed the criterion of significance, F(1,44) = 3.82, p = .057, η_p^2 = .080; F < 1 for the main effect of response type. For the no contingency condition, there was no interference overall,[5] F < 1; for the high contingency condition, there was a significant interference effect overall, F(1,22) = 5.60, p = .044, η_p^2 = .203.

[5] Despite the non-significant interaction Response type × Contingency, the reader might want to know that for the no contingency condition there was a significant effect of response type,

For the error rates (see Table 2), a 4 (semantic congruency: neutral silence vs. neutral word vs. congruent vs. incongruent) × 2 (response type: braking vs. steering) × 2 (contingency: no vs. high) mixed factors MANOVA yielded a significant main effect of response type (more errors while steering), $F(1,44) = 12.06$, $p = .001$, $\eta_p^2 = .215$, and a main effect of congruency that just missed the criterion of significance, $F(3,42) = 2.70$, $p = .058$, $\eta_p^2 = .162$; all other *F*s < 1.39. The Helmert contrast of main interest (congruent vs. incongruent trials) revealed a significant difference, $F(1,44) = 7.41$, $p = .009$, $\eta_p^2 = .144$. There were fewer errors in the congruent condition compared to the incongruent one.

Discussion

In the no contingency condition of Experiment 3, we were able to replicate the results pattern from Experiment 2. Even though the driving maneuvers were much more complex than simply pressing a key, we again found a clear priming effect. Once more, the effect occurred due to considerable facilitation, with no significant interference involved. We found significant priming for both response types, although braking was associated with larger effects than steering. However, we do not want to emphasize this latter result (which was not replicated in the high contingency condition) since a variety of differences between the two response modes might be responsible for it.

Comparing the no contingency versus high contingency conditions yielded a clear result as well: High contingency priming generated a larger priming effect than the no contingency condition. Given the applied perspective of this research, this result comes as a relief, since auditory warnings in the real world will of course only be applied if they have

$F(1,22) = 6.17$, $p = .021$, $\eta_p^2 = .219$ ($F < 1$ for the corresponding test in the high contingency condition). While there was no interference effect for braking, $F(1,22) = 1.59$, $p = .221$, $\eta_p^2 = .067$, steering was associated with a negative interference score, $F(1,22) = 4.39$, $p = .048$, $\eta_p^2 = .166$; that is, the neutral condition was slower than the incongruent one. Because of the singularity of this result in all of our experiments, we refrain from putting much emphasis on it.

sufficient validity. Therefore, it was important to show that our priming effects are not reduced in cases of high validity.

However, we have to concede that there was a small but significant interference effect in the high contingency condition. Thus, there is indeed a potential negative effect of an incorrect warning when warnings tend to be quite accurate overall. This issue should be explored more in future research. But note, we are not talking about costs of false positive warnings (i.e., an auditory warning appears although the situation is not critical). The interference effect refers to a mismatch of auditory warning and visual sign. In real driving systems, auditory warnings with a mismatch rate of twenty percent will certainly not be allowed.

General Discussion

Our experiments show that short, time-compressed but meaningful auditory warnings can clearly facilitate responses to visual events that semantically correspond to them. This priming effect could be observed in a rather complex visual search context (Experiment 1) as well as in a dynamic driving simulation context (Experiment 2 and 3) requiring complex response maneuvers (Experiment 3). With these results, we have taken a major step forward from our former research (Mahr & Wentura, 2014).

In our former experiments (Mahr & Wentura, 2014), we exclusively used auditory color words and visual color symbols as materials. In the present experiments, we used icons that symbolized complex semantic concepts. Thus, the facilitating effect of a congruent auditory prime has to be explained at the level of semantic representations. By way of contrast, color patches do not *symbolize* specific colors; they *are* specific colors, and color is a single, easy-to-process feature. Thus, although the link between the phonological code and the visual event is given by a semantic link even in the case of colors, it was not self-evident that the priming effects found for colors would be replicated with more complex materials.

A further difference is given by the task used in the present Experiments 2 and 3 in comparison to Mahr and Wentura (2014) as well as Chen and Spence (2011). All these experiments indicate that auditory primes do not simply prepare a corresponding response (which is a priori the dominant explanation of the effects found in the Stroop-like paradigm used in the present Experiment 1 and Experiments 1 and 2 of Mahr & Wentura). However, Mahr and Wentura (2014) as well as Chen and Spence (2011) found a larger detection sensitivity for targets that were preceded by a congruent auditory prime. Since the identity of the target did not need to be categorized, a devil's advocate might claim that the effect can in principle be explained by reversing the roles of prime and target: Whenever a task-relevant auditory prime (e.g., a color name) is identified, participants are prepared to quickly press the "target present" button if they additionally have a feeling that there is a match between the auditory stimulus and the visual display; this feeling of match can of course only arise in the congruent case. This feeling might result already from early states of processing the visual display, which yield feature maps that indicate the presence of simple features like specific colors (Treisman & Gelade, 1980). Thus, we cannot be sure whether the auditory prime really facilitates encoding of the visual target with its specific feature combinations (e.g., location, form). In the present Experiments 2 and 3, we had participants categorize the targets (i.e., whether they were red or green). Thus, a priming effect can only be explained by enhanced full processing of the target object.

In Mahr and Wentura (2014), we discussed the cross-modal semantic priming effect in terms of working memory processes. We will only briefly recapitulate the argument. To solve the task, the participants have to keep the four target categories in working memory. Several authors (Garavan, 1998; McElree, 2006; Nee & Jonides, 2011; Oberauer, 2002; Olivers, Peters, Houtkamp, and Roelfsema (2011), however, have made important distinctions between different states of working memory with only a single object as the focus of attention. Thus, accordingly spoken word primes might place the corresponding category in the foreground of working memory, metaphorically speaking. The second step is that the prioritized working memory item increases visual sensitivity to the corresponding visually presented item (Desimone & Duncan, 1995; Olivers, Meijer, & Theeuwes, 2006). That is, the prioritized working memory item is an active template that directly resonates with corresponding visual input (Olivers et al., 2011).

Given this rationale, it can be hypothesized that the search for the prioritized item should be more efficient. This can be seen by reduced RTs for congruently primed items. Moreover, although we did not manipulate display size (i.e., the number of stimuli in the displays) in our present experiments, we can provide preliminary evidence for the hypothesis that the search is more efficient: If this is the case, not only mean RTs should be affected by priming an object, but also the means of the individual SDs of raw RTs, since flatter search slopes lead to a decrease in mean RTs *and* SDs. In *Appendix A,* we present supplemental analyses of our data that indeed find this effect for all our experiments. Of course, it is up to future research to test this additional hypothesis more directly.

Inattentional deafness research

A research topic that is touched but not addressed by our research is research on inattentional deafness (see, e.g., Macdonald & Lavie, 2011; see Murphy & Greene, 2015, 2017, for applications in a driving context), which – at least at first sight – seems to be in conflict to our results. Inattentional deafness refers to the missing of an auditory stimulus when participants are engaged in visual tasks. Tests of inattentional deafness mimic studies of inattentional blindness (e.g., Simons & Chabris, 1999) in that they present a single noticeable (or even salient), but unexpected and task irrelevant stimulus (e.g., a tone). Dependent variable is whether participants had noticed or not noticed the stimulus (but see Molloy, Griffiths, Chait, & Lavie, 2015, for a study assessing event-related potentials). Detection rate was dramatically lower given high perceptual load conditions compared to low load conditions (Macdonald & Lavie, 2011; Raveh & Lavie, 2015).

Of course, in our present experiments we did not manipulate perceptual load. Thus, we do not know for sure whether effects in the present experiments would have been even larger in a low perceptual load condition. However, with reference to our former experiments (Mahr & Wentura, 2014), this seems unlikely. Nevertheless, there are two arguments that make the results of inattentional deafness research and the present ones compatible.

First, the dependent variable in our research is the response time to the *target* stimuli. Thus, in principle it is conceivable that participants cannot subjectively report on the auditory *prime* event, although an effect of the prime on target latencies might be observable (compare, e.g, to studies in the visual domain on non-conscious priming, e.g., Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003).

Second and more important, Tellinghuisen et al. (2016) recently already elaborated on the *prima facie* discrepancy between their (and our) earlier results (i.e., larger cross-modal Stroop-like effects given high peceptual load compared to low load; Mahr & Wentura, 2014; Tellinghuisen & Nowak, 2003) and results of on inattentional deafness research. They made clear that task-relatedness of distractor/prime stimuli seems to be the decisive factor. The salient event in inattentional deafness research is always realized by a task-irrelevant stimulus whereas distractor/prime stimuli in our research correspond directly to the stimuli of the targer set.

We have to admit, however, that one aspect of inattentional deafness research points to a critical aspect of our studies (and the majority of other studies who transfer paradigms from basic research into more applied contexts). Critical incidents in road traffic are – fortunately – rather rare. Hence, it is an open question whether we can directly infer from our results that a sudden auditory prime which is presented closely to such an incident really helps (see in general Ho, Gray, & Spence, 2014; for an analogue problem in visual attention studies, see Wolfe, Horowitz, & Kenner, 2005). This must be left for future research.

The meaning for an applied context

Our experiments clearly show the usefulness of brief auditory prime signals to detect visual items. These primes facilitate target detection even if the probability of congruent pairs is at chance level, thereby corroborating the argument that the effect is basically of an automatic character (in the sense of involuntary, non-strategic).

For the driving context, we considered it especially relevant (a) whether improvements in reaction time actually occur due to a semantic match between prime and target, (b) whether attentive listening is a necessary precondition, (c) whether the effects occur immediately, (d) whether auditory presentation duration influences effects, and (e) whether the effects would hold for aspects of complex motor performance like swerving and braking.

Based on our findings, we recommend speech warnings due to their general potential to automatically affect visual processing. If semantic representations of spoken words and visual objects match, considerable and immediate benefits can be expected, whereas in the case of a mismatch, interference seems to be less pronounced (but see *Discussion* of Experiment 3). The robust performance increase that we found for compatible speech warnings compared to no warnings or unspecific warnings (equivalent to a master alert) points to an important safety benefit, making successful real-world application a viable proposition.

Moreover, we argue that the application of time-compressed speech warnings might improve responses for warnings that are reliably compatible with the actual hazard. In this regard, we further suggest that time compression might especially improve longer, polysyllabic warnings, but further research would be needed regarding this additional hypothesis.

Concerning the transferability of our results to real-world critical road scenarios, we recommend that future work focus on four major points: First, a larger set of target objects would considerably reduce warning frequency and expectancy, resulting in a more realistic driving scenario and attentional setting. Second, more spoken denotations with different levels of phonetic or semantic similarity should be investigated for purposes of further specification. Third, a direct comparison between auditory information regarding detection of objects on the one hand and action suggestions on the other could provide useful insights into cross-modal processes and further clarify which warning type should be used under which circumstances. Fourth, effects should be tested with more rare presentation of primes and critical incidents.

Of course, further research is needed to explore whether benefits of compressed auditory warnings are restricted to young participants who are native speakers of the language of the warning messages. Thus, studies with older participants or non-native speakers of the language are needed. Finally, a further step would be to incorporate compressed auditory warnings into more complex and realistic driving simulations than the one used in our experiments.

Summing up, our results support the application of speech warnings in time-critical road situations when drivers need to be informed immediately about a current safety issue. We assume that speech warnings have the potential to reduce accident severity or even prevent accidents from occurring as compared to no warnings or non-specific auditory alerts.

References

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version retrieved 2011 from http://www.praat.org/).

Campbell, J. L., Richard, C. M., Brown, J. L., & McCallum, M. (2007). *Crash warning system interfaces: Human factors insights and lessons learned. (National Highway Traffic Safety Administration Technical Report No. DOT HS 810 697)*. Washington, DC: U.S. Department of Transportation.

Cao, Y., Mahr, A., Castronovo, S., Theune, M., Stahl, C., & Müller, C. (2010). Local danger warnings for drivers: The effect of modality and level of assistance on driver reaction. In C. Rich & Q. Yang (Eds.), *Proceedings of the 15th International Conference on Intelligent User Interfaces* (pp. 239–248). New York: ACM.

Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1554-1568.

Chen, Y.-C., & Spence, C. (2013). The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. *Multisensory Research, 26*, 371-386.

Chen, Y.-C., & Spence, C. (2017). Dissociating the time courses of the cross-modal semantic priming effects elicited by naturalistic sounds and spoken words. *Psychonomic Bulletin & Review*, Advance online publication.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics, 41*, 393-401.

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition, 31*, 353-359.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Elliott, E. M., Cowan, N., & Valle-Inclan, F. (1998). The nature of cross-modal color–word interference effects. *Perception & Psychophysics, 60*, 761-767.

Elliott, E. M., Morey, C. C., Morey, R. D., Eaves, S. D., Shelton, J. T., & Lutfi-Proctor, D. A. (2014). The role of modality: Auditory and visual distractors in Stroop interference. *Journal of Cognitive Psychology, 26*, 15-26.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*, 263-276.

Hillyard, S. A., Störmer, V. S., Feng, W., Martinez, A., & McDonald, J. J. (2016). Cross-modal orienting of visual attention. *Neuropsychologia, 83*, 170-178.

Ho, C., Gray, R., & Spence, C. (2014). To what extent do the findings of laboratory-based spatial attention research apply to the real-world setting of driving? *IEEE Transactions on Human-Machine Systems, 44*, 524-530.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied, 11*, 157-174.

Ho, C., & Spence, C. (2006). Verbal interface design: Do verbal directional cues automatically orient visual spatial attention? *Computers in Human Behavior, 22*, 733-748.

Ho, C., Tan, H. Z., & Spence, C. (2006). The differential effect of vibrotactile and auditory cues on visual spatial attention. *Ergonomics, 49*, 724-738.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Hillsdale, NJ: Erlbaum.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 451-468.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*, 75-82.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*, 314-334.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition, 7*, 166-174.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 110*, 14196-14201.

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics, 73*, 1780-1789.

Mahr, A., & Wentura, D. (2014). Time-compressed spoken word primes crossmodally enhance processing of semantically congruent visual targets. *Attention, Perception, & Psychophysics, 76*, 575-590.

Math, R., Mahr, A., Moniri, M. M., & Müller, C. (2012). OpenDS: A new open-source driving simulator for research. In A. L. Kun, N. L. Boyle, B. Reimer & A. Riener

(Eds.), *Adjunct proceedings of the 4th international conference on automotive user interfaces and interactive vehicular applications* (pp. 7-8). Portsmouth, NH: ACM.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 46* (Vol. 46, pp. 155-200). San Diego: Elsevier Academic Press Inc.

McKeown, D., & Isherwood, S. (2007). Mapping candidate within-vehicle auditory displays to their referents. *Human Factors, 49*, 417-428.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. *The Journal of Neuroscience, 35*, 16046-16054.

Mordkoff, J. T. (2012). Observation: Three reasons to avoid having half of the trials be congruent in a four-alternative forced-choice experiment on sequential modulation. *Psychonomic Bulletin & Review, 19*, 750-757.

Murphy, G., & Greene, C. M. (2015). High perceptual load causes inattentional blindness and deafness in drivers. *Visual Cognition, 23*, 810-814.

Murphy, G., & Greene, C. M. (2017). Load theory behind the wheel; perceptual & cognitive load effects. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, Advance online publication.

Nee, D. E., & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: Evidence for a 3-state model of memory. *Neuroimage, 54*, 1540-1548.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Ngo, M. K., Pierce, R. S., & Spence, C. (2012). Using multisensory cues to facilitate air traffic management. *Human Factors, 54*, 1093-1103.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97*, 316-333.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 411-421.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1243-1265.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15*, 327-334.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160-174.

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics, 77*, 483-492.

Roelofs, A. (2005). The visual-auditory color-word Stroop asymmetry and its time course. *Memory & Cognition, 33*, 1325-1336.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*, 1059-1074.

Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences, 1191*, 182-200.

Spence, C. (2012). Drive safely with neuorgonomics. *The Psychologist, 25*, 664-667.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Tellinghuisen, D. J., Cohen, A. J., & Cooper, N. J. (2016). Now hear this: Inattentional deafness depends on task relatedness. *Attention Perception & Psychophysics, 78*, 2527-2546.

Tellinghuisen, D. J., & Nowak, E. J. (2003). The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics, 65*, 817-828.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, 100*, 6275-6280.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (pp. 95-116). New York: Guilford.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435*, 439-440.

Table 1

Mean reaction times (RTs in ms; error rates in parentheses) in Experiment 1 as a function of compression levels and semantic congruency conditions

	Semantic congruency condition		
Duration	Neutral	Congruent	Incongruent
100%	1095 (2.1)	1001 (5.0)	1154 (4.4)
50%	1140 (5.0)	1041 (5.5)	1166 (3.7)
30%	1163 (4.0)	1108 (5.5)	1201 (4.4)

Table 2

Mean reaction times (in ms; error rates in parentheses) as a function of semantic congruency conditions (Experiment 2), response type, and contingency (Experiment 3)

	Semantic congruency condition			
	Neutral Silence	Neutral Word	Congruent	Incongruent
Experiment 2				
	1078 (4.3)	1074 (2.8)	1015 (3.0)	1078 (4.3)
Experiment 3				
No Contingency				
Braking	1069 (1.1)	1089 (1.1)	992 (0.3)	1108 (0.6)
Steering	1002 (1.7)	1048 (2.7)	966 (1.9)	1013 (2.6)
High Contingency				
Braking	1054 (0.5)	1067 (0.0)	914 (0.3)	1100 (1.8)
Steering	1043 (2.7)	1065 (1.6)	939 (1.0)	1119 (4.4)

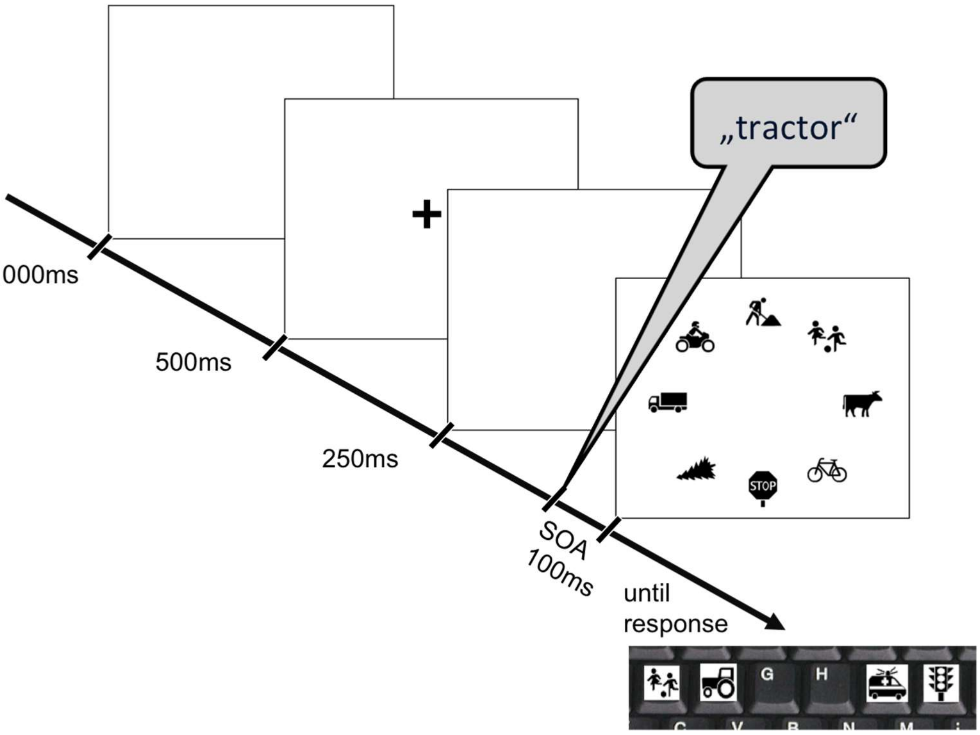


Figure 1. Trial sequence (here an example of an incongruent trial) from Experiment 1. The prime word ("tractor") is presented via headphones, starting 100 ms prior to the visual target (here: children; see top/right). Task was to categorize the target stimulus (which was always from the set "children", "tractor", "ambulance", "traffic light" and placed randomly at one of the eight possible locations) in a given display by key-press (see the key assignment at the bottom of the figure; keys G and H are only depicted to indicate the location of the response keys on a QWERTY keyboard). Filler icons were randomly placed at the remaining locations.

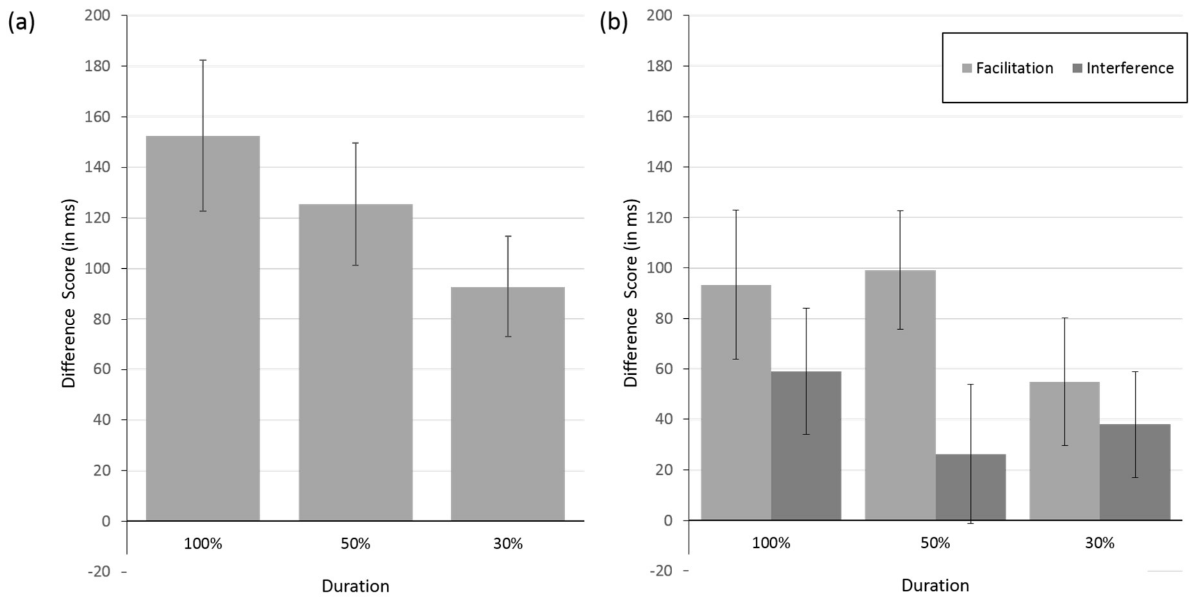


Figure 2. Reaction time (RT) differences for the three compression rates (duration) in Experiment 1; (a) Overall: RT incongruent trials – RT congruent trials; (b) facilitation: RT neutral – RT congruent; interference: RT incongruent – RT neutral. The error bars indicate ± 1 standard error mean of the overall effect.

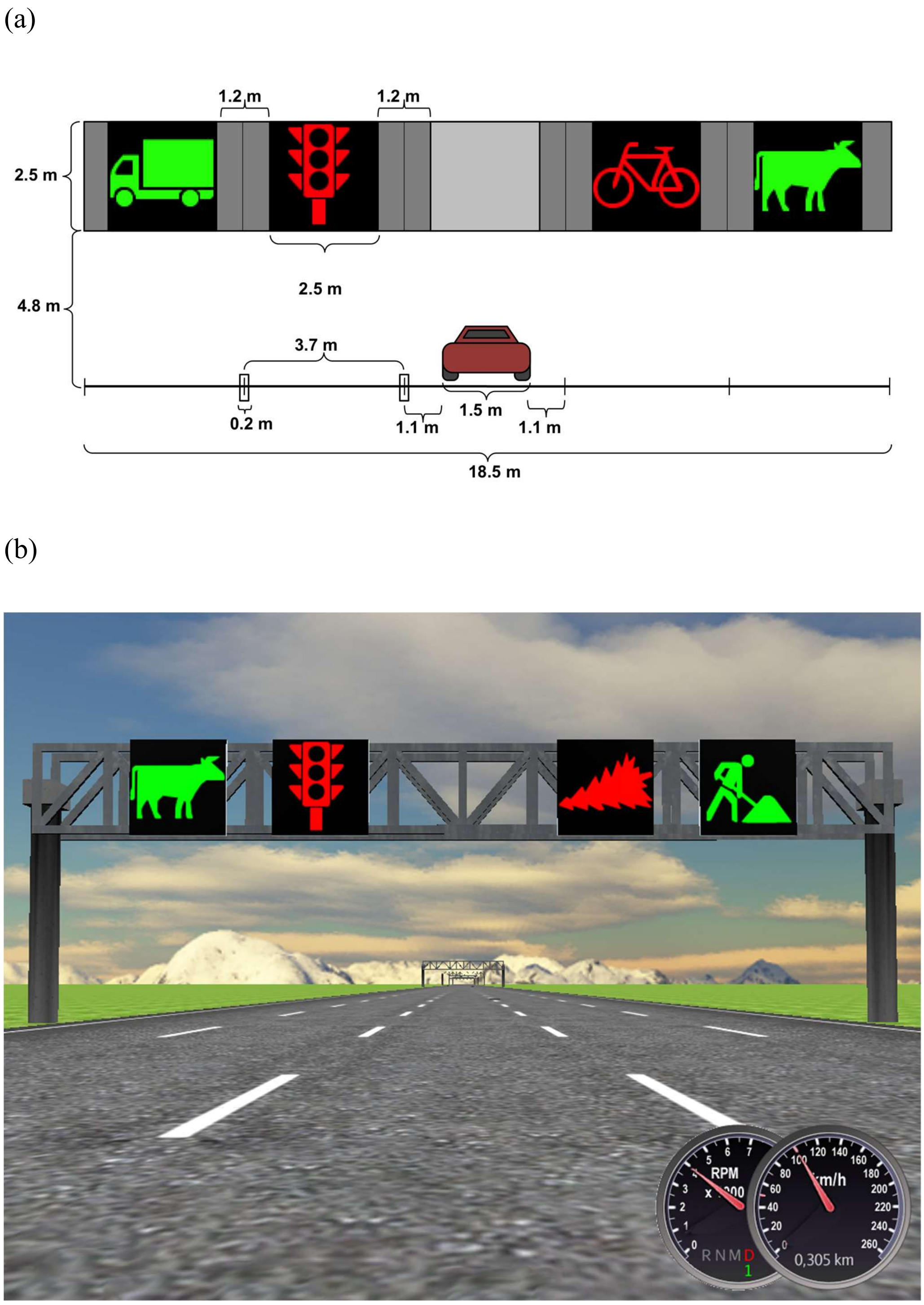


Figure 3. (a) Schematic cross-section of the three-dimensional simulation environment including dimensions of the car, the roadway, and the gantry road signs. (b) Screenshot of the track and an overhead gantry sign recorded during a simulation run (Exp. 2 and 3).

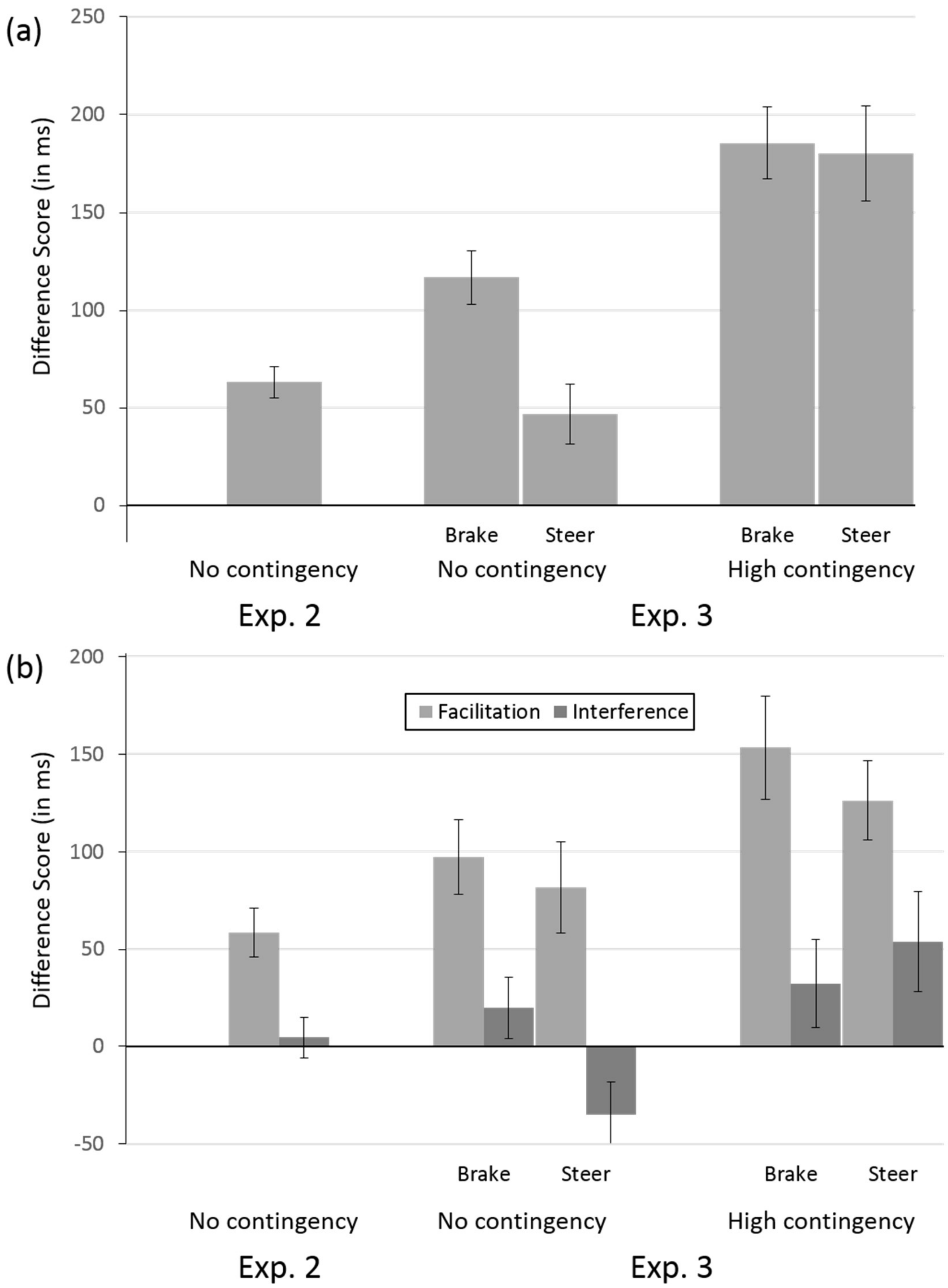


Figure 4. Reaction time (RT) differences in Experiments 2 and 3; (a) Overall: RT incongruent trials – RT congruent trials; (b) facilitation: RT neutral – RT congruent; interference: RT incongruent – RT neutral. The error bars indicate ± 1 standard error mean of the overall effect.

Appendix A

Analyses of standard deviations of individual raw RTs

Table A1 shows the means of the individual standard deviations (SD) of raw RTs for all experiments (see *General Discussion* for the hypothesis related to these variables). For succinctness, we present only analyses of the difference scores SD(incongruent) minus SD(congruent).

In Experiment 1, a one-factorial 3 (duration: 100% vs. 50% vs. 30%) repeated measures MANOVA with the difference scores as the dependent variable yielded a significant constant effect, $F(1, 18) = 10.39$, $p = .005$, $\eta_p^2 = .366$. SDs for the congruent condition were smaller than SDs for the incongruent condition. This contrast was not moderated by duration, $F < 1$. In Experiment 2, the mean difference score was significantly above zero, $F(1, 24) = 7.46$, $p = .012$, $\eta_p^2 = .237$. That is, SDs for the congruent condition were again smaller than SDs for the incongruent condition. Finally, for Experiment 3 a 2 (response type: braking vs. steering) × 2 (contingency: no vs. high) mixed factors MANOVA for repeated measures with the difference scores as the dependent variable yielded a significant constant effect, $F(1, 44) = 19.91$, $p < .001$, $\eta_p^2 = .312$. As in Experiments 1 and 2, SDs for the congruent condition were smaller than SDs for the incongruent condition. Additionally, there was a main effect of response type, $F(1, 44) = 6.21$, $p = .017$, $\eta_p^2 = .124$. However, this moderation of the basic difference is of ordinal type only: the constant effect was significant for braking, $F(1, 44) = 22.70$, $p < .001$, $\eta_p^2 = .340$, and steering, $F(1, 44) = 4.88$, $p = .032$, $\eta_p^2 = .100$. (For the sake of completeness: $F(1, 44) = 2.16$, $p = .149$, $\eta_p^2 = .047$ for the main effect of contingency; $F < 1$ for the interaction of response type and contingency.)

Table A1

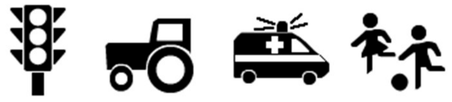
Mean individual standard deviations of raw RTs (in ms) in Experiments 1 to 3 as a function of compression levels and semantic congruency conditions

	Semantic congruency condition			
	Neutral Silence	Neutral Word	Congruent	Incongruent
Experiment 1				
100%	–	382	358	417
50%	–	449	388	426
30%	–	411	424	453
Experiment 2	247	238	227	244
Experiment 3				
No Contingency				
Braking	262	276	256	284
Steering	202	221	212	222
High Contingency				
Braking	263	270	212	265
Steering	211	230	187	209

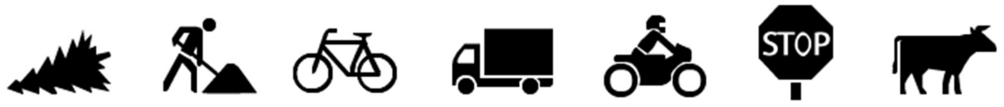
Appendix B

Visual stimuli used in Experiment 1

Target stimuli

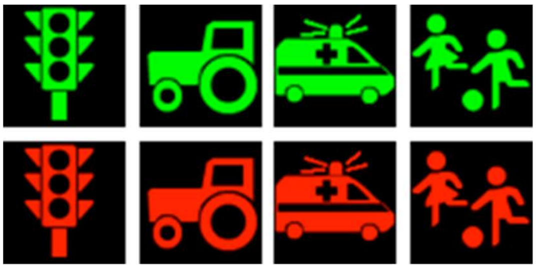


Distractor stimuli

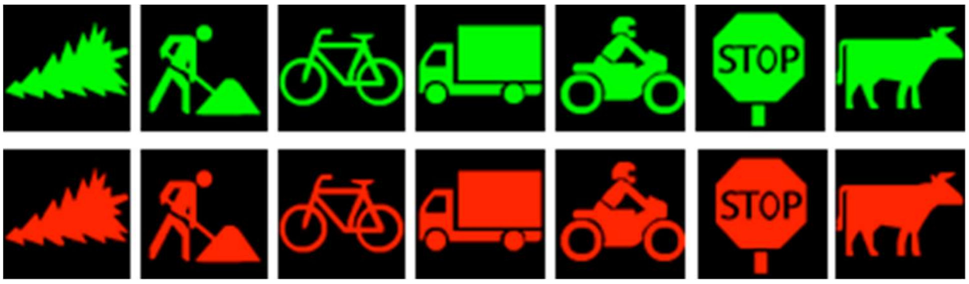


Visual stimuli used in Experiments 2 and 3

Target stimuli



Distractor stimuli



Stimuli for the practice phase