



Audiovisual semantic interference and attention: Evidence from the attentional blink paradigm

Erik Van der Burg^{a,*}, Sanne G. Brederoo^a, Mark R. Nieuwenstein^b, Jan Theeuwes^a, Christian N.L. Olivers^a

^aVrije Universiteit Amsterdam, The Netherlands

^bExperimental Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 23 March 2009

Received in revised form 29 January 2010

Accepted 30 January 2010

Available online 21 February 2010

PsycINFO classification:

2346

Keywords:

Attention

Multisensory processing

Attentional blink

ABSTRACT

In the present study we investigate the role of attention in audiovisual semantic interference, by using an attentional blink paradigm. Participants were asked to make an unspeeded response to the identity of a visual target letter. This target letter was preceded at various SOAs by a synchronized audiovisual letter-pair, which was either congruent (e.g. hearing an “F” and viewing an “F”) or incongruent (e.g. hearing an “F” and viewing a “Z”). In Experiment 1, participants were asked to match the members of the audiovisual letter-pair. In Experiment 2, participants were asked to ignore the synchronized audiovisual letter-pairs altogether and only report the visual target. In Experiment 3, participants were asked to identify only one of the audiovisual letters (identify the auditory letter, and ignore the synchronized visual letter, or vice versa). An attentional blink was found in all three experiments indicating that the audiovisual letter-pairs were processed. However, a congruency effect on subsequent target detection was observed in Experiments 1 and 3, but not in Experiment 2. The results indicate that attention to the semantic contents of at least one modality is necessary to establish audiovisual semantic interference.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

We live in a multimodal environment and receive an enormous amount of information via the different senses. This information often interacts when presented in close temporal or spatial proximity (see Welch & Warren, 1980, for a review). For instance, when watching television, sounds seem to come from the actors lip-movements, instead of coming from the location of the loudspeakers – an effect known as the ventriloquism illusion (e.g. Alais & Burr, 2004; Thomas, 1941). Whereas these studies reported a visual effect on audition, other studies have reported the reverse, an auditory effect on vision (e.g. Vroomen & De Gelder, 2000). For instance, the presentation of a visual event is perceived as being brighter when that visual event is accompanied by an auditory signal (Stein, London, Wilkinson, & Price, 1996). More recently, Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008b) have shown that events that are hard to distinguish visually capture attention when they are synchronized with an auditory signal.

In addition to many studies reporting multisensory interactions when using simple stimuli (e.g. flashes in combination with beeps), quite a few studies have reported multisensory interactions on the level of more semantic representations (Calvert, Campbell, &

Brammer, 2000; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; McGurk & MacDonald, 1976; Molholm, Ritter, Javitt, & Foxe, 2002; Soto-Faraco, Navarra, & Alsius, 2004; Van Atteveldt, Formisano, Goebel, & Blomert, 2004). For instance, McGurk and MacDonald (1976) reported that the auditory percept of a phoneme is influenced by the observation of the speaker's lip-movements. Soto-Faraco et al. (2004) found evidence that this McGurk effect is an automatic process and occurs prior to attentional selection. In line with the idea that semantic information is automatically integrated, Calvert et al. (2000), and more recently, Van Atteveldt et al. (2004) reported effects of audiovisual letter congruency (i.e. same vs. different audiovisual letter-pairs) in the human brain, even though participants listened and viewed the letters passively (i.e. there was no task involved).

In a follow-up study, Van Atteveldt, Formisano, Goebel, and Blomert (2007) replicated their 2004 findings, by showing a congruency effect when participants viewed and listened to the letter-pairs passively. Surprisingly, however, no congruency effect was observed, neither in terms of brain activity, nor in terms of behavioral data, when participants were asked to actively respond to the audiovisual letter-pairs. These findings suggest that when the task requires participants to attend both modalities, the automatic integration of multisensory information ceases to exist. Interestingly, such a conclusion is inconsistent with earlier findings indicating that some attention to both modalities is a prerequisite to achieve audiovisual integration in the first place (Fujisaki,

* Corresponding author. Address: Cognitive Psychology, Vrije Universiteit, Van der Boerhorststraat 1, 1081BT Amsterdam, The Netherlands. Tel.: +31 20 598 6744. E-mail address: e.van.der.burg@psy.vu.nl (E. Van der Burg).

Koene, Arnold, Johnston, & Nishida, 2006; Talsma, Doty, & Woldorff, 2007), although it should be noted that these studies did not use semantic stimuli. In contrast with Van Atteveldt et al. (2007), other studies have reported a semantic congruency effect when participants were asked to attend and actually respond to the presentation of the audiovisual stimuli (see e.g. Blau, Van Atteveldt, Formisano, Goebel, & Blomert, 2006; Laurienti et al., 2004). For instance, Laurienti et al. reported that participants were faster when they responded to congruent stimuli (e.g. hearing the word “red”, and seeing a red circle) than when they responded to incongruent stimuli (i.e. hearing the word “red”, and seeing a blue circle).

All in all then, the role of attention in audiovisual semantic interference (i.e. interference between two modalities based on the semantic content) is rather inconsistent and mixed. Here we sought to further investigate whether synchronously presented audiovisual letters automatically interfere at a semantic level. Can participants ignore semantic information from one modality (e.g. audition) while simultaneously processing semantic information presented to another modality (e.g. vision)? In the present study, we use the performance on an attentional blink task (Raymond, Shapiro, & Arnell, 1992) as an indirect measure of the effects of semantic congruency across vision and audition under different attentional circumstances.

1.1. The attentional blink paradigm as an indirect measure of congruency effects

The attentional blink paradigm is often used to study the temporal limitations of the human brain in processing multiple stimuli. The attentional blink is the term that refers to an impairment in detecting the second of two visual targets (T1 and T2) when it appears within roughly half a second after the first. The attentional blink has been found within modalities (Duncan, Martens, & Ward, 1997; Tremblay, Vachon, & Jones, 2005; Vachon & Tremblay, 2005; Vachon & Tremblay, 2006), and also between modalities (Arnell & Jenkins, 2004; Arnell & Jolicoeur, 1999; Arnell & Larson, 2002; Potter, Chun, Banks, & Muckenhaupt, 1998). The latter version, however, turns out to be more difficult to find (Duncan et al., 1997; Hein, Parr, & Duncan, 2006), and we have recently suggested that a cross-modal attentional blink may depend on the complexity of the stimuli (Van der Burg, Olivers, Bronkhorst, Koelewijn, & Theeuwes, 2007). Specifically, an auditory T1 may interfere only with a visual T2 if both are processed at a semantic level. This is because a semantic processor may supersede the different modalities. In the Van der Burg et al. study, no audiovisual attentional blink was found with T1s that are specific to the auditory modality (i.e. different beeps).

Whereas most studies reported an attentional blink when T1 and T2 were separated by distractor elements (e.g. letters), recently, Nieuwenstein, Potter, and Theeuwes (2009) reported an attentional blink within the visual modality when the targets were separated by a blank interval (i.e. in the absence of distractors between T1 and T2). Moreover, Nieuwenstein, Van der Burg, Theeuwes, Wyble, and Potter (2009) reported an attentional blink when a single target (i.e. the former T2) was preceded by a single distractor letter, indicating that irrelevant events can trigger an attentional blink as well. The experimental design used in the Nieuwenstein, Van der Burg, et al. (2009) study makes it possible to re-investigate whether audiovisual stimuli from different senses automatically integrate at a semantic level. For instance, one might expect that an unmasked audiovisual letter-pair can trigger an attentional blink for an upcoming visual target, and that the magnitude of the attentional blink depends on the semantic content of the synchronized audiovisual events. In other words, similar audiovisual letters would lead to faster T1 processing and thus to a reduced attentional blink for an upcoming second target compared to the condition that the audiovisual letters are different.

In the present study, participants were asked to detect a visual target letter (vT2), which was preceded by an audiovisual letter-pair (avT1). The members of avT1 were either congruent or incongruent (e.g. hearing an “F” while seeing an “F”, or hearing an “F” while seeing a “Z”). Furthermore, the relevance of processing the two members of avT1 was manipulated. Experiment 1 set one boundary condition by asking participants to match the two members of the audiovisual letter-pair and investigate the effects of semantic congruency on the detection of the subsequent visual target letter (vT2). As participants needed to attend to both modalities to perform the task, semantic congruency effects between the two were expected, and indeed occurred: vT2 detection improved when a congruent avT1 was presented. Experiment 2 then set the other boundary condition by asking participants to fully ignore avT1. Thus no attention should be paid to the semantic content of the letter-pair, while vT2 still had to be detected. The data will show that avT1 then still generated an attentional blink (indicating processing of the avT1 pair), but that the congruency effect indeed disappeared. Finally, in Experiment 3, participants were asked to look or listen to only one of the members of avT1, and to ignore the other. If semantic information stemming from the ignored modality can be ignored, no congruency effect should be found. If this information is automatically processed, a congruency effect should emerge. In Experiment 3, vT2 detection improved when the preceding avT1 was congruent compared to the condition that avT1 was incongruent.

2. Experiment 1

In the present experiment, participants viewed and listened to the presentation of two synchronized letters, which were either congruent (e.g. hearing an “F” and seeing an “F”) or incongruent (e.g. hearing an “F” and seeing a “Z”). Subsequently, after a specified stimulus onset asynchrony (SOA), participants viewed another target letter (A, B, or C), which was immediately followed by a mask. Participants were asked to report whether the synchronized auditory and visual letters were similar or different (avT1), and to identify the second target letter (vT2; an A, B or C). Fig. 1 presents an example trial.

We expected that avT1 would trigger an attentional blink despite the fact that it is not followed by a mask (Nieuwenstein, Potter, et al., 2009; Nieuwenstein, Van der Burg, et al., 2009). If attending to the audiovisual stimuli overrides multisensory interference, in line with Van Atteveldt et al. (2007), no semantic congruency effect on vT2 performance is expected. In contrast, theories that predict that audiovisual integration depends on attention (Talsma et al., 2007) would be supported by the presence

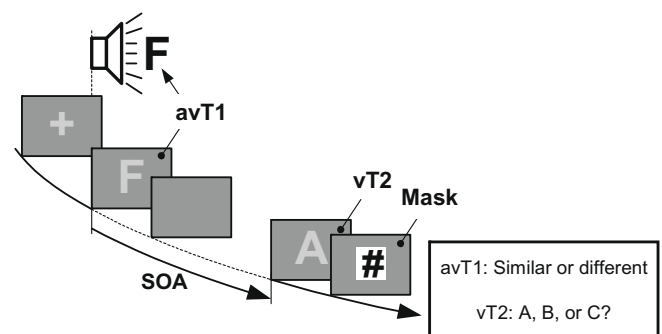


Fig. 1. Example trial of Experiment 1. Participants were asked whether the synchronized auditory and visual letters were similar or different (avT1), and to identify the second target letter (vT2; an A, B or C). The stimulus onset asynchrony (SOA) between avT1 and vT2 was manipulated. Participants were required to make unspeeded responses, after the presentation of the mask.

of an audiovisual semantic congruency effect, because attention was assigned to both the auditory and visual letter in order to perform the task.

2.1. Method

2.1.1. Participants

Nine students (four female; mean age 21.3 years, range: 18–25 years) participated in the experiment, and were paid €7 an hour.

2.1.2. Apparatus and stimuli

The experiment was run in a dimly lit, air-conditioned cubicle. Participants were seated at a distance of approximately 80 cm from the monitor (120 Hz refresh rate) and wore Sennheiser (HD202) head phones. The visual stimuli consisted of the letters A, B, C, F, L, R, S, X, Y and Z. All letters were capitals and were presented in a gray 48-point Arial font (9.34 cd/m², 0.8° width, 1.0° height) at the center of a black background (<0.05 cd/m²). The auditory stimuli were the letters: F, L, R, S, X, Y and Z. The letters were spoken in Dutch in a male voice and were compressed, without altering pitch, to a duration of 90 ms (stereo, 16 bit, 44.1 kHz sample rate). The first target (avT1) was a combination of an auditory and a visual letter (F, L, R, S, X, Y and Z). On half of the trials, avT1 was congruent (e.g. hearing a “Z” and seeing a “Z”), and on the remaining trials, avT1 was incongruent (e.g. hearing an “F” and seeing a “Z”). The second target (vT2) was a visual letter A, B, or C. The mask was a “#” (black, <0.05 cd/m², 0.8° height, 0.8° width), on a white background (76.7 cd/m²). AvT1 and vT2 were never identical to avoid priming (Koelewijn, Van der Burg, Bronkhorst, & Theeuwes, 2008), or repetition blindness effects (e.g. Kanwisher, 1987).

2.1.3. Design and procedure

Each trial began with the presentation of a fixation cross (0.8° height, 0.8° width) for 600 ms at the center of the screen followed by the presentation of an auditory and visual letter, the combination of which represents avT1. The visual letter was presented for 88 ms, and the auditory letter for 90 ms. The audiovisual onsets were synchronized. vT2 then followed at four possible stimulus onset asynchronies (SOAs; 100, 200, 300 or 700 ms), and was immediately followed by a mask for a duration of 100 ms. After the termination of the mask, participants were asked to make an unspeeded response to avT1 by pressing the g- or n-key when the auditory and visual vT1 were the same, or different, respectively, and to make an unspeeded response to vT2 by pressing the 1, 2, 3 key when they saw A, B, or C, respectively. vT2 identity was balanced, and mixed within blocks. Participants received instructions prior to the experiment, emphasizing accuracy. Participants first performed one practice block of 75 random trials (i.e. congruent and incongruent trials) in which they only practiced the vT2 task. During practice, the SOA was fixed (700 ms) and the duration of the vT2 was adjusted, such that participants' overall vT2 accuracy approached 80% correct. The overall mean duration for vT2 was subsequently used in the main experiment. Overall mean vT2 presentation time was 47.1 ms, ranging from 32 to 64 ms. After the practice block, participants completed six blocks of 48 trials each. SOA and congruency were randomly mixed within blocks. Participants received feedback about their overall mean accuracy after each block.

2.2. Results and discussion

The results of Experiment 1 are presented in Fig. 2. Data from practice blocks were excluded. avT1 and vT2 accuracy were subjected to a repeated-measures Univariate Analysis of Variance (AN-

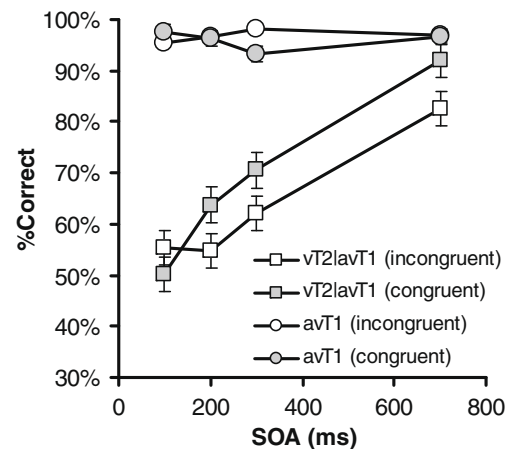


Fig. 2. Results of Experiment 1. Mean percentage correct vT2 (given that avT1 was correctly identified) and mean percentage correct avT1, as a function of congruency, and stimulus onset asynchrony (SOA). The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (Loftus & Masson, 1994). The confidence intervals are those for the congruency \times SOA interaction effect.

OVA) with SOA and congruency as within-subject variables. The reported values for p are those after a Huynh–Feldt correction for sphericity violations, with alpha set at .05.

2.2.1. avT1 accuracy

Overall avT1 accuracy was 96.3%. The ANOVA on avT1 accuracy revealed no significant main effects of SOA and congruency ($F_s < 1$ and $ps > .6$). The two-way interaction between SOA and congruency was reliable, $F(3, 24) = 4.8$, $MSE = .001$, $p < .01$, and further examined by separate two-tailed t -tests for each SOA condition. When the SOA was 300 ms, avT1 performance was worse when avT1 was congruent (93.2%) than when avT1 was incongruent (98.1%), $t(8) = 2.5$, $p < .05$. No such congruency effect was observed for the other SOA conditions ($ts < 1.4$ and $ps > .19$).

2.2.2. vT2 accuracy

Overall vT2 accuracy was 66.4%. There was a trend towards a reliable congruency effect, as vT2 accuracy was overall higher when the auditory and visual letters were congruent (69.1%) than when the auditory and visual letters were incongruent (63.6%), $F(1, 8) = 4.8$, $MSE = .011$, $p = .059$. Furthermore, vT2 accuracy increased as the SOA increased, $F(3, 24) = 13.8$, $MSE = .052$, $p = .001$. More importantly, the two-way interaction between SOA and congruency was significant, $F(3, 24) = 6.1$, $MSE = .004$, $p < .005$, and further examined by separate two-tailed t -tests for each SOA condition. These analyses revealed reliable congruency effects when the SOA was 200, 300, and 700 ms ($ts > 2.4$, $ps < .05$), indicating that vT2 performance was better when the auditory and visual letters (avT1) were congruent than when the auditory and visual letters were incongruent. In contrast, no such congruency effect was observed for the shortest SOA condition, $t(8) = 1.1$, $p = .287$.

The present findings are clear. Processing two letters from different sensory modalities (avT1) causes an attentional blink pattern for an upcoming vT2, corroborating many studies showing within and between modality attentional blinks (e.g. Arnell & Jolicoeur, 1999; Raymond et al., 1992). Whereas most studies reported an attentional blink when avT1 was followed by distractors, the present study revealed an attentional blink when T1 was followed by a blank interval (see also Nieuwenstein, Potter, et al., 2009; Nieuwenstein, Van der Burg, et al., 2009; Vachon & Tremblay, 2008). Furthermore, vT2 accuracy was better when the members of avT1 were congruent rather than incongruent. Inconsistent with

Van Atteveldt et al. (2007), a congruency effect was observed when participants were asked to respond to the congruence of the audiovisual letters. These results are consistent with Laurienti et al. (2004) who have reported an audiovisual congruency effect when participants were also required to respond to the congruence of audiovisual stimuli. Whereas we observed a clear congruency effect when the SOA was long (i.e. >100 ms), no congruency effect was observed for the shortest SOA (100 ms). The absence of a congruency effect in the latter condition might be explained by the temporal window of multisensory integration (see e.g. Van der Burg et al., 2008b; Van Wassenhove, Grant, & Poeppel, 2007). In other words, the temporal window of multisensory integration is known to be rather lengthy and as a result it could be possible that the auditory content was matched up with vT2 instead of with the synchronized visual vT1, explaining the absence of a congruency effect when the SOA between avT1 and vT2 was short, but not when the SOA between avT1 and vT2 was long (SOA > 100 ms).

The finding of a congruency effect is in line with Talsma et al. (2007) who have proposed that attention to both modalities is a prerequisite to establish multisensory integration. In the next two experiments we further investigate the automaticity of the observed audiovisual congruency effect on vT2 performance, by manipulating the relevance to process the semantic contents of the audiovisual (avT1) events.

3. Experiment 2

In Experiment 1, we found that the semantic congruency of an attended audiovisual letter-pair affected subsequent detection of a visual target. This may not be too surprising given that the congruency of the two members was directly relevant to the task (though see Van Atteveldt et al., 2007). The aim of the second experiment was to investigate how automatic these congruency effects are. In contrast to Experiment 1, the avT1 was now completely irrelevant to the task. Participants were asked to report only the identity of a single visual target letter (A, B, or C), and to ignore the synchronized audiovisual events preceding the target event. If the semantic content stemming from the two sensory modalities is automatically integrated (Van Atteveldt et al., 2004, 2007), then we expect to see a similar congruency effect as in Experiment 1. In contrast, if semantic interference depends on the task-relevance of the semantic content, then we expect no audiovisual congruency effect here.

It may be argued that an ignored first target (here avT1) should not generate an attentional blink in the first place (Raymond et al., 1992) and hence no congruency effects would be expected anyway. However, there are several reasons to expect an attentional blink under the present conditions: (i) the second visual target (vT2) is a letter, and so are the members of avT1. Studies have shown that objects contingent with the participant's attentional set will capture attention and generate an attentional blink (e.g. Folk, Leber, & Egeth, 2002; Maki & Mebane, 2006); (ii) if the second target is made sufficiently sensitive, an attentional blink can be observed even after an irrelevant event (Nieuwenstein, Van der Burg, et al., 2009). In any case, the data here will reveal a full-blown attentional blink.

3.1. Method

3.1.1. Participants

Fifteen new students (six female; mean age: 20.5 years; ranging from 18 to 24 years) participated in the present experiment.

The present experiment was identical to Experiment 1, except that participants were required to only execute one task, identifying the target letter A, B, or C (i.e. the former vT2 task). Participants

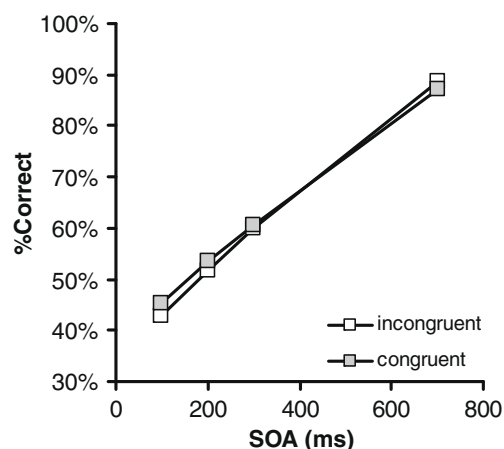


Fig. 3. Results of Experiment 2. Mean percentage correct, as a function of distractor similarity, and stimulus onset asynchrony (SOA). The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (Loftus and Masson, 1994). The confidence intervals are those for the congruency \times SOA interaction effect.

were asked to ignore the synchronized audiovisual events and to focus on the visual target letter. The congruency between the synchronized audiovisual distractor letters was manipulated within blocks. On half of the trials, the distractors were similar (congruent trials), and on the remaining trials the distractors were different (incongruent trials). Practice was identical to practice in the previous experiment and determined vT2 duration. Overall mean vT2 presentation time was 48.0 ms, ranging from 32 to 72 ms. After practice, participants performed six experimental blocks, of 48 trials each. During experimental blocks, SOA and distractor similarity were randomly mixed within blocks.

3.2. Results and discussion

The results of Experiment 2 are presented in Fig. 3. Target accuracy was subjected to an ANOVA with SOA, and congruency as within-subject variables.

3.2.1. vT2. accuracy

Overall accuracy was 61.2%. There was a significant effect of SOA, $F(3, 42) = 73.9$, $MSE = .015$, $p < .001$, as the target performance increased with increasing SOA. The effect of congruency as well as the two-way interaction between SOA and congruency were far from significant, $F_s < 1$, $p_s > .4$.

The present results are clear. Whereas Experiment 1 revealed a clear audiovisual congruency effect on vT2 performance, the present experiment yielded no such audiovisual congruency effect when the synchronized events were irrelevant to the task. Note that the ignored audiovisual letter-pair did generate an attentional blink of the same magnitude as in Experiment 1, indicating that the avT1 was processed. This is consistent with automatic contingent attentional capture accounts (e.g. Folk et al., 2002; Nieuwenstein, Van der Burg, et al., 2009). However, the fact that its semantic content was not task-relevant led to the absence of a congruency effect, suggesting that cross-modal semantic interference only occurs when the semantic information is relevant to the task of the participant.

4. Experiment 3

In Experiment 1, participants were asked to attend to both members of avT1, with an audiovisual semantic congruency effect as a result. In Experiment 2, participants were asked to ignore avT1

altogether, resulting in the absence of a congruency effect. Having set the two boundary conditions, we can now turn to the question whether participants can exclusively attend to the semantic information stemming from just one modality. In other words, can they filter for a particular sense when processing semantic content? Or does semantic content from an irrelevant modality interfere when semantic content of the relevant modality becomes relevant? In the present experiment participants were asked to either report the identity of the auditory letter (and to ignore the synchronized visual letter, aT1), or to report the identity of the visual letter (and to ignore the synchronized auditory letter, vT1 condition) as the first target. In this way, we made one modality irrelevant to the task. If semantic information from an irrelevant modality interferes with semantic information from a relevant modality, then we expect to replicate the results of Experiment 1, by showing an audiovisual congruency effect on vT2 performance. In contrast, when attention to *both* the auditory and the visual letter is a prerequisite to find semantic interference (e.g. Talsma et al., 2007), then no audiovisual congruency effect is expected.

4.1. Method

4.1.1. Participants

Fourteen new students (one male; mean age: 20.6 years; ranging from 18 to 29 years) participated in the present experiment.

The experiment was identical to Experiment 1, except that participants were asked to report the identity of the target letter from a predefined modality as a/vT1 task. On half of the trials, participants were asked to report the auditory target letter (aT1 condition) and to ignore the synchronized visual distractor letter. On the other half of the trials, participants were asked to report the visual target letter as T1 task (vT1 condition), and to ignore the synchronized auditory distractor letter. The target-distractor congruency was manipulated as before. On half of the trials, the target letter was congruent with the synchronized distractor letter, and on the remaining trials, the letters were incongruent. After each trial, participants made an unspeeded response to the a/vT1 target letter by pressing the corresponding key on the keyboard. Participants performed eight blocks of 48 trials each. a/vT1 modality was varied between blocks, presented in alternating and counterbalanced order. Practice was identical to Experiment 1, and determined the vT2 presentation duration. Overall mean vT2 presentation time was 38.0 ms, ranging from 24 to 56 ms.

4.2. Results and discussion

The results of Experiment 3 are presented in Fig. 4. a/vT1 and vT2 accuracy were subjected to an ANOVA with SOA, T1 modality, and congruency as within-subject variables.

4.2.1. a/vT1 accuracy

Overall a/vT1 accuracy was 96.3%. There was a reliable congruency effect on a/vT1 performance, $F(1, 13) = 6.8$, $MSE = .012$, $p < .05$, as participants made overall more errors when a/vT1 was incongruent with the distractor in the other modality (94.4%) than when a/vT1 was congruent with the distractor in the other modality (98.2%). The ANOVA revealed a trend towards a reliable SOA effect as performance was better when the SOA was 200, 300, or 700 ms (SOA = 200 ms, 96.8%; SOA = 300 ms, 96.5%; SOA = 700 ms, 96.6%) than when vT2 immediately followed T1 (SOA = 100 ms; 95.2%), $F(3, 39) = 2.7$, $MSE = .001$, $p = .063$. No other effects were reliable (all $F_s < 2.2$ and $p_s > .126$).

4.2.2. vT2 accuracy

Overall vT2 accuracy was 63.2%. vT2 performance increased with increasing SOA, $F(3, 39) = 63.1$, $MSE = .051$, $p < .001$. There was a reliable congruency effect, $F(1, 13) = 15.3$, $MSE = .016$, $p < .005$, as overall vT2 performance was better when a/vT1 was congruent with the distractor in the other modality (66.5%) than when a/vT1 was incongruent with the distractor in the other modality (59.9%). Furthermore, there was a significant two-way interaction between SOA and congruency, $F(3, 39) = 4.0$, $MSE = .008$, $p < .005$. The interaction was further examined by separate two-tailed *t*-tests for each SOA condition. These analysis yielded a reliable congruency effect on vT2 performance when the SOA was 200 ms, and 300 ms [$t(13) = 3.2$, $p < .01$, and $t(13) = 5.0$, $p < .001$, respectively], but not when the SOA was 100, or 700 ms [$t(13) = .1$, $p = .9$, and $t(13) = 1.7$, $p = .1$, respectively]. More specifically, when the SOA was 200 or 300 ms, vT2 performance was better when a/vT1 was congruent with the distractor in the other modality than when a/vT1 was incongruent with the distractor in the other modality (SOA was 200 ms: 48.9% vs. 59.4%; SOA was 300 ms: 56.8% vs. 68.9%). Importantly, this congruency effect on vT2 performance was independent of the to be reported T1 modality, as the main effect of T1 modality, and none of the interactions including T1 modality were reliable (all $F_s < 1.7$ and $p_s > .17$).

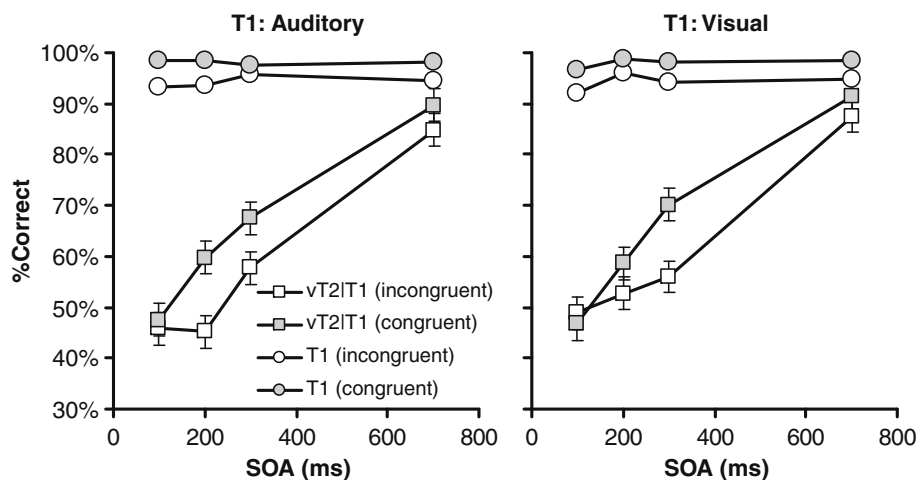


Fig. 4. Results of Experiment 3. Mean percentage correct vT2 (given that a/vT1 was correctly identified) and mean percentage correct a/vT1, as a function of T1 modality, congruency, and stimulus onset asynchrony (SOA). The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). The confidence intervals are those for the congruency \times SOA interaction effect.

The present findings are clear. Viewing or listening to the identity of a visual or auditory target, resulted in a clear attentional blink pattern for a second visual target. Furthermore, the present study replicated Experiment 1, by showing a congruency effect on vT2 performance between the target (a/vT1) and the synchronized irrelevant distractor in the other modality. Thus, even though participants were required to ignore the irrelevant distractor letter in the other modality, the synchronization of audiovisual events affected the detection for an upcoming target letter. It is interesting to note that the size of the congruency effect was unaffected by whether participants were asked to identify both (Experiment 1) or only one of the audiovisual letters (Experiment 3), as the observed congruency effect was overall 5.5% and 6.6%, in Experiment 1 and 3, respectively. Moreover, we observed that this congruency effect was independent of the to be reported T1 modality, as the T1 modality \times congruency interaction and the three-way interaction were not reliable, $F < 1$ and $F = 1$, $p > .5$, respectively. Even though participants were asked to respond to visual targets, the presentation of conflicting auditory semantic information did lead to some interference.

However, one might argue that participants were not able to ignore the irrelevant distractor in the other modality, but actually adopted a strategy to use the semantic information of the synchronized distractor letter to improve, or facilitate T1 identification. This alternative explanation is possible, however, we see this strategy as unlikely. For instance, if the observed congruency effect is due to active processing of both the audiovisual letters, then it is unclear why we were not able to find a congruency effect on T1 performance in Experiment 1, in which participants were required to report T1 congruency. Therefore, we suggest that the observed T1 congruency effect in the present experiment is due to interference of the to-be-ignored distractor letter in the irrelevant modality instead of a strategy to use the semantic information of the synchronized distractor letter to improve, or facilitate T1 identification.

5. General discussion

An unmasked audiovisual event generated an attentional blink pattern in all three experiments, regardless of whether this event was relevant to the task, indicating that T1 was processed in all experiments. This finding is consistent with earlier studies showing that an unmasked T1 can induce an attentional blink in the visual modality (Nieuwenstein, Potter, et al., 2009; Vachon & Tremblay, 2008), and auditory modality (Vachon & Tremblay, 2008), and that even an irrelevant distractor can induce an attentional blink pattern (Folk et al., 2002; Maki & Mebane, 2006; Nieuwenstein, Van der Burg, et al., 2009; Olivers & Meeter, 2008). The result also has implications for theories claiming that adequate masking of T1 is required in order for an attentional blink to occur (e.g. Martin & Shapiro, 2008; Olivers & Meeter, 2008; Raymond et al., 1992). With regard to the observed attentional blink, it is unclear whether this attentional blink represents a visual attentional blink (i.e. vT1 affects vT2), an auditory–visual attentional blink (i.e. aT1 affects vT2), or perhaps a combination of both (i.e. avT1 affects vT2). The presence of a cross-modal attentional blink is consistent with the notion that the attentional blink is due to an amodal central processing limitation (Arnell & Jolicoeur, 1999).

More important for the present purpose, T2 performance was affected by the congruency of the members of the preceding audiovisual event when either one (Experiment 3) or both (Experiment 1) the letters were relevant to the task, but not when neither was relevant (Experiment 2). Thus, even though we observed a clear attentional blink in all experiments, suggesting that the synchronized audiovisual events captured attention, this did not lead to semantic interference per se (Remington & Folk, 2001). In other

words, audiovisual semantic interference was observed only when participants were asked to respond to at least one of the modalities of the synchronized audiovisual event. This finding suggests that the integration of semantic information provided by different modalities is not fully automatic, but depends on semantic information being relevant to the participant – even if only the semantic information from one modality is relevant. In other words, the integration of information stemming from different modalities is partially automatic: if one is looking for the identity of a visual letter, the identity of a spoken letter is automatically processed too, and vice versa. We suggest that the auditory and visual semantic information is processed in a single, supramodal semantic system in which auditory and visual input is converted into the same neural code. Integration and interference then occurs directly within this single network. This also explains why we (Van der Burg et al., 2007) observed an attentional blink between the senses (audition and vision) when we used semantic stimuli (i.e. a to-be-identified letter) for the auditory and visual target, but not when we used simple auditory beeps for the auditory target in combination with a semantic visual target.

5.1. Memory encoding account

The observed congruency effect in Experiment 1 can be explained in terms of a memory load effect (Akyürek & Hommel, 2005; Akyürek, Hommel, & Jolicoeur, 2007). For instance, Akyürek and colleagues reported that T2 performance is susceptible to the memory load. In other words, they have shown that T2 performance decreased as the memory load increased. With regard to Experiment 1, one might argue that participants stored one letter in memory in the case of a congruent trial and two different letters in the case of an incongruent trial, explaining why T2 suffered more when T1 was incongruent than when T1 was congruent. However, this memory load account cannot explain the observed attentional blink in Experiment 2, since there was no need to store the letters in memory, as participants were asked to ignore the two synchronized audiovisual events. Moreover, when comparing the congruent conditions between the different experiments, overall T2 performance was even worse in Experiment 2 (61.2%) in which participants were asked to ignore the synchronized letters than in Experiment 1 (69.1%) or Experiment 3 (66.1% when T1 was an auditory task; 66.7% when T1 was a visual task) on congruent trials in which participants stored just one relevant letter. This suggests that the attentional blink was independent of the number of letters stored in memory (see also Nieuwenstein & Potter, 2006; Olivers, Van der Stigchel, & Hulleman, 2007). We suggest that the observed congruency effects in Experiments 1 and 3 are due to the semantic content of the synchronized audiovisual letters.

The observed congruency effects are consistent with the “auditory–visual Stroop effect” (see e.g. Blau et al., 2006; Cowan & Barron, 1987; Elliott, Cowan, & Valle-Inclan, 1998; Laurienti et al., 2004; see Stroop, 1935, for the original visual–visual Stroop effect). For instance, Laurienti et al. observed faster responses when participants were asked to make a speeded response to congruent audiovisual color stimuli than when responding to incongruent audiovisual stimuli. In the present study, we observed a similar enhanced performance in an unspeeded task, in which naming of visually presented letter is impaired by the simultaneous auditorily presented letter. Here too the semantic content stemming from the auditory modality could not be ignored when the semantic content delivered by the visual modality was important. The present study shows that this type of interference is bidirectional, that it is modulated by the task-relevance of semantic information, and that it is just as strong as when both modalities are important.

The present results are inconsistent with those of Van Atteveldt et al. (2004, 2007). Whereas they observed no reliable effect of con-

gruency when participants responded to the congruency of the audiovisual letter-pair, the present Experiment 1 revealed a clear congruency effect on T2 performance when participants responded to the congruency of the audiovisual letter-pair, even though we did not observe a direct congruency effect on the T1 task itself (Experiment 1). Conversely, Van Atteveldt and colleagues observed a congruency effect when participants viewed and listened to the audiovisual stimuli passively, but not actively, whereas we found no such effect when participants passively viewed the first target (Experiment 2). Van Atteveldt and colleagues argued that an explicit task allows participants to completely overrule the congruency effect while passive viewing gives free play to automatic processes. It remains unclear why they observed a pattern of results completely opposite to ours. Obviously there are many differences between their study and ours, each of which might need investigation in the future. For instance, in the Van Atteveldt et al. (2007) study, participants were asked to make a speeded response (congruent vs. incongruent) when they saw or heard the second letter, because on most trials the audiovisual letters were never synchronized. In their case, the modality to respond to was unknown and might have reduced the congruency effect. Furthermore, they have used rather lengthy intervals between the visual and auditory events (up to 500 ms), which might have reduced the congruency effect as well (see e.g. Van der Burg et al., 2008b; Van Wassenhove et al., 2007). For instance, participants might have perceived the auditory and visual events as being two different entities in the case that the audiovisual events were separated by rather long intervals, explaining the absence of semantic interference.

5.2. The role of attention

The present study indicates that at least some attention to the semantic content of at least one modality is needed to obtain a cross-modal congruency effect on a semantic processing level. Attending to either the visual or the auditory input appears to be a prerequisite and at the same time a sufficient condition to generate cross-modal semantic interference. Presently, there are a number of studies that suggest a role for attention in multisensory integration (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007; Talsma et al., 2007). For example, it has been found that audiovisual integration of speech signals suffers when participants attend to yet another modality (Alsius et al., 2007), or when attentional demands are generally high (Alsius et al., 2005).

On the other hand, there are also a number of studies that indicate that multisensory integration occurs relatively easy and automatically (Olivers & Van der Burg, 2008; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008a; Van der Burg et al., 2008b; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2009; Vroomen & De Gelder, 2000), and even under high attentional demands (Santangelo & Spence, 2007). For instance, recently we have shown that synchronized audiovisual events capture attention in a cluttered dynamic environment (Van der Burg et al., 2008b). This effect occurred even though the synchronized audiovisual events (a beep in combination with a color change) were completely irrelevant to the task, suggesting a strong automatic, pre-attentive component for multisensory integration (Van der Burg et al., 2008a). Consistently, in an event-related potential study, Giard and Peronn et (1999) showed audiovisual interactions in the EEG signal emerging from the visual cortex as early as 50 ms post-stimulus onset. A coarse analysis of when attention does and does not play a crucial role in the integration of information from multiple senses appears to suggest that pre-attentive integration may be limited to the space and time dimensions. When lights and sounds co-occur in space and/or time, their signals are rapidly merged to create a

strong unified percept that can even capture attention (Van der Burg et al., 2008a, 2008b, 2009). In contrast, attention is necessary for the integration of *content* – that is, the integration of non-spatial and non-temporal properties, such as the semantic representations involved in speech recognition (Alsius et al., 2005, 2007). An exception is the Talsma et al. (2007) study, in which participants were presented with simple concurrent beeps and gratings, but here too the task was to attend to an orthogonal property, namely the brief softening of the light or sound in a sequence of stimuli. It is clear that more work is needed to be done. So far, from the current study we can conclude that semantic cross-modal integration requires attention to semantic content. Therefore, we suggest that the observed congruency effect is due to interference at a late semantic processing stage, instead of early multisensory integration.

Acknowledgements

This research was supported by a Dutch Technology Foundation STW grant (07079), a division of NWO and the Technology Program of the Ministry of Economic Affairs (to Jan Theeuwes and Adelbert W. Bronkhorst), and a NWO-VIDI grant 452-06-007 (to Christian N.L. Olivers).

References

- Aky rek, E. G., & Hommel, B. (2005). Short-term memory and the attentional blink: Capacity versus content. *Memory and Cognition*, 33, 654–663.
- Aky rek, E. G., Hommel, B., & Jolicoeur, P. (2007). Direct evidence for a role of working memory in the attentional blink. *Memory and Cognition*, 35(4), 621–627.
- Alais, D., & Burr, D. (2004). The ventriloquism effect results from near-optimal bimodal integration. *Current Biology*, 14, 257–262.
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under attention demands. *Current Biology*, 15, 839–843.
- Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, 183, 399–404.
- Arnell, K. M., & Jenkins, R. (2004). Revisiting within-modality and cross-modality attentional blinks: Effects of target-distractor similarity. *Perception and Psychophysics*, 66, 1147–1161.
- Arnell, K. M., & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 630–648.
- Arnell, K. M., & Larson, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychonomic Bulletin and Review*, 9(3), 497–506.
- Blau, V., Van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2006). Visual influences on speech sound discrimination: A parametric fMRI study. *NeuroImage*, 31(S38).
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10, 649–657.
- Cowan, N., & Barron, A. (1987). Cross-modal, auditory–visual stroop interference and possible implications for speech memory. *Perception and Psychophysics*, 41(5), 393–401.
- Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, 387, 808–810.
- Elliott, E. M., Cowan, N., & Valle-Inclan, F. (1998). The nature of cross-modal color-word interference effects. *Perception and Psychophysics*, 60(5), 761–767.
- Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception and Psychophysics*, 64(5), 741–753.
- Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences*, 273(1588), 865–874.
- Giard, M. H., & Peronn et, F. (1999). Auditory–visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 11(5), 473–490.
- Hein, G., Parr, A., & Duncan, J. (2006). Within modality and cross modality attentional blinks in a simple discrimination task. *Perception and Psychophysics*, 68(1), 54–61.
- Kanwisher, N. G. (1987). Repetition blindness – Type recognition without token individuation. *Cognition*, 27(2), 117–143.
- Koelewijn, T., Van der Burg, E., Bronkhorst, A. W., & Theeuwes, J. (2008). Priming T2 in a visual and auditory attentional blink task. *Perception and Psychophysics*, 70(4), 658–666.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, 158, 405–414.

- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, 1(4), 476–490.
- Maki, W. S., & Mebane, M. W. (2006). Attentional capture triggers an attentional blink. *Psychonomic Bulletin and Review*, 13(1), 125–131.
- Martin, E. W., & Shapiro, K. L. (2008). Does failure to mask T1 cause lag-1 sparing in the attentional blink? *Perception and Psychophysics*, 70(3), 562–570.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2002). Multisensory visual–auditory object recognition in humans: A high density electrical mapping study. *Cerebral Cortex*, 14, 452–465.
- Nieuwenstein, M. R., & Potter, M. C. (2006). Temporal limits of selection and memory encoding: A comparison of whole versus partial report in rapid serial visual presentation. *Psychological Science*, 17(6), 471–475.
- Nieuwenstein, M. R., Potter, M. C., & Theeuwes, J. (2009). Unmasking the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 159–169.
- Nieuwenstein, M. R., Van der Burg, E., Theeuwes, J., Wyble, B., & Potter, M. C. (2009). Temporal constraints on conscious vision: On the ubiquitous nature of the attentional blink. *Journal of Vision*, 9(9), 1–14.
- Olivers, C. N. L., & Meeter, M. (2008). A boost and Bounce theory of temporal attention. *Psychological Review*, 115, 836–863.
- Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, 1242, 191–199.
- Olivers, C. N. L., Van der Stigchel, S., & Hulleman, J. (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research*, 71, 126–139.
- Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology-Learning Memory and Cognition*, 24(4), 979–992.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18, 849–860.
- Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science*, 12(6), 511–515.
- Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1311–1321.
- Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audio-visual speech integration: Evidence from the speeded classification task. *Cognition*, 92, B13–B23.
- Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychological analysis. *Journal of Cognitive Neuroscience*, 8, 497–506.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, 17(3), 691–701.
- Thomas, G. J. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, 28, 167–177.
- Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception and Psychophysics*, 67(2), 195–208.
- Vachon, F., & Tremblay, S. (2005). Auditory attentional blink: Masking the second target is necessary, delayed masking is sufficient. *Canadian Journal of Experimental Psychology*, 59(4), 279–286.
- Vachon, F., & Tremblay, S. (2006). Delayed masking and the auditory attentional blink – A test for retrieval competition and bottleneck models. *Experimental Psychology*, 53(3), 182–190.
- Vachon, F., & Tremblay, S. (2008). Modality-specific and amodal sources of interference in the attentional blink. *Perception and Psychophysics*, 70(6), 1000–1015.
- Van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech in the human brain. *Neuron*, 43, 271–282.
- Van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage*, 36(4), 1345–1360.
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., Koelewijn, T., & Theeuwes, J. (2007). The absence of an auditory–visual attentional blink is not due to echoic memory. *Perception and Psychophysics*, 69(7), 1230–1241.
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008a). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision*, 8(5), 1–10.
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008b). Pip and pop: Non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1053–1065.
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile-visual synchrony increases visual saliency. *Neuroscience Letters*, 450(1), 60–64.
- Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia*, 45, 598–607.
- Vroomen, J., & De Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583–1590.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3), 638–667.