1

2 Running head: Multisensory Selection

3

4

5 **Overt spatial attention modulates multisensory selection**

6

7 Anne Jensen¹, Simon Merz¹, Charles Spence², & Christian Frings¹

8

9 ¹ University of Trier,

10 ² University of Oxford

11

12 Words count (9015)

13

14 Corresponding author:

15 Anne Jensen

16 University of Trier

17 D-54286 Trier

18 jensen@uni-trier.de

1

2

Authors' Note

3

4 Anne Jensen, Simon Merz, and Christian Frings, University of Trier, Department of Psychol-
5 ogy, D-54286, Germany (Email: jensen@uni-trier.de), and Charles Spence, Crossmodal Re-
6 search Laboratory, Department of Experimental Psychology, University of Oxford, South
7 Parks Road, Oxford OX1 3UD, United Kingdom.

8 The research reported in this article was supported by a grant of the Deutsche Forschungsge-
9 meinschaft to Christian Frings and Charles Spence (FR 2133/5-3). Correspondence concern-
10 ing this article should be addressed to Anne Jensen, University of Trier, Department of Psy-
11 chology, D-54286, Germany (Email: jensen@uni-trier.de).

Abstract

In daily life, signals from the different senses are often integrated in order to enhance multisensory perception. However, an important, yet currently still controversial, topic concerns the need for attention in this integration process. To investigate this question, we turned to the processing of multisensory distractors. Note that multisensory target processing is typically confounded with attention as people attend to the stimuli that they respond to. We therefore designed a multisensory flanker task in which the target and distractor stimuli were both multisensory and the congruency between the features (auditory and visual) was varied orthogonally. In addition, we manipulated whether distractor or target was within the focus of participants' gaze (i.e., was overtly attended). Importantly, distractor congruency effects were modulated by this manipulation. Fixating the distractor led to crossmodal congruency effects between the visual and auditory feature dimensions (e.g., a visually incongruent distractor interfered more if it was also auditorily incongruent with the target), while congruency effects were independent of each other when the distractor was not fixated (i.e., visual interference was not modulated by auditory interference in this case). These results suggest that distractors outside the focus of overt attention are processed at the level of features whereas those distractors presented centrally (i.e., at fixation) are processed as a configuration of features. Taken together, these results can be taken to suggest that the multisensory integration of irrelevant stimuli depends on the focus of spatial attention.

Keywords: Multisensory perception, Multisensory selection, distractor processing, attention, gaze

1

2

Public Significance Statement

3

This study provides novel insight into the processing of multisensory distractors.

4

When simultaneously presenting two multisensory stimuli in a selection task, the components

5

of the irrelevant stimulus are not integrated unless they are attended. This suggests major role

6

of attention for integration in multisensory selection.

Introduction

Each and every day, our senses are bombarded with many different unisensory signals, that can be seen, heard, felt, tasted, or smelled. To give this flood of multisensory information meaning, and to avoid cognitive overload, people filter the incoming stream of signals and preferentially select just those stimuli, objects, or events that are likely to be relevant to them. As most of the daily events that we are confronted with are multisensory in nature, the question arises as to how we go about sorting and grouping the incoming information. Which unisensory features, in other words, are treated as belonging together? Multisensory stimuli that we want to respond to have a high chance of being integrated as our attention is allocated to these ‘target’ stimuli. But what about those multisensory objects (events, or combinations of stimuli) that are irrelevant to us or even distract us while we try to focus on the targets? Thus far, it turns out that surprisingly little is known about multisensory processing of distractor stimuli.

To date, research on the topic of multisensory integration has primarily focused on the integration of target features. We review the evidence on this kind of integration below and discuss moderators such as (spatial) attention, focusing on the case of audiovisual integration. We then introduce a multisensory variant of Eriksen and Eriksen’s (1974) classic flanker task in order to analyze multisensory distractor processing and the possible influences of overt attention thereon.

Multisensory target processing

To date, research on the topic of multisensory integration (MI) has primarily been focused on target information processing. Behaviorally, MI often leads to the facilitation of the

processing of the integrated stimuli (although there can be competition as well, e.g., Hirst, Cragg, & Allen, in press; Sinnett, Soto-Faraco, & Spence, 2008; see Spence, Parise, & Chen, 2011, for a review) in comparison to feature-based processing. On a statistical level, this effect can be seen in the interaction between the features. For instance, in settings with multisensory or unisensory feature-combinations as targets and/or distractors, congruency effects (defined as the difference between incongruent and congruent *trials*) of one feature vary as a function of the congruency of the other feature(s) (e.g., see Lavie, 1997; Zmigrod, Spapé, & Hommel, 2009)¹. But what are the necessary conditions for MI to occur – and especially for audiovisual integration? It has been argued that temporal and spatial co-occurrence are driving forces underlying MI (see Spence, 2007; Stein & Meredith, 1990, 1993; Stein & Stanford, 2008; Zmigrod & Hommel, 2011) although spatial co-occurrence would mostly only appear to be necessary if space is somehow relevant to a participant's task (Girard, Collignon, & Lepore, 2010; see Spence, 2013, for a review). Temporal co-occurrence, on the other hand, would seem to be a prerequisite for audiovisual integration (Kolewijn, Bronkhorst, & Theeuwes, 2010; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). Enhanced MI is seen with synchronized sensory signals (Harrar, Harris, & Spence, 2016; see also Vatakis & Spence, 2010). However, MI is still found for signals that are slightly desynchronized, providing that both signals fall within the window of multisensory integration (what is often referred to as the 'temporal binding window'; see Colonius & Diederich, 2004; Wallace & Stevenson, 2014). The width of this temporal binding window, however, has been shown to change as a function of different task demands and types of stimuli (Spence & Squire, 2003; Vatakis, Maragos, Rodomagoulakis, & Spence, 2012; see Vatakis & Spence, 2010, for a review). Finally, there are other factors such as temporal correlation (Parise, Spence, & Ernst, 2012), perceptual grouping (see Spence, 2015, for a review) and higher cognitive factors such as the

¹ Note that in the cited studies and in our experiments, 'congruent' means that the distractor feature is perceptually identical to the relevant target feature and associated with the same response as this feature.

‘unity assumption’ that also contribute to the MI of audiovisual stimuli (see Chen & Spence, 2017, for a review).

In the present study, temporally synchronized and correlated audiovisual signals are used as target and distractor stimuli. The features of each audiovisual stimulus are presented at the same time, in the same ‘rhythm’ (i.e., a steady sound signal) and from the same spatial location. Yet the target and distractor will be segregated both spatially and temporally thus making space task-relevant and so presumably triggering the spatial rule (see Spence, 2013, for a review).

The role of attention in multisensory integration

The role of attention in multisensory integration is much more ambiguous. Several published studies have demonstrated that attention is needed for successful integration at the same time as other researchers have published findings suggesting that MI is automatic and independent of attention (Bertelson, Vroomen, De Gelder, & Driver, 2000; Caclin, Soto-Faraco, Kingstone, & Spence, 2002; Helbig & Ernst, 2008; Santangelo, Ho, & Spence, 2008; Santangelo & Spence, 2007; Van der Burg et al., 2008; Vroomen, Bertelson, & De Gelder, 2001; Wahn, Murali, Sinnett, & König, 2017; Zimmer & Macaluso, 2007). The influence of attention on MI can especially be seen in dual-task paradigms in which attention is occupied by a secondary task. For instance, Santangelo and Spence (2007) investigated the attentional capture of unisensory (visual and auditory) and multisensory (audiovisual) cues under different conditions of perceptual load. In the low load condition (simply fixating on a stationary fixation point), all peripheral cues successfully captured spatial attention. In the high load condition (a centrally presented rapid serial visual presentation [RSVP] task), only the multisensory cues successfully captured participant’s attention. Furthermore, using two visual or two auditory signals as cues, which presumably contained as much information (and e.g.,

transients) as the multisensory cue, did not show this attentional capture, thus ruling out the possibility that the multisensory advantage merely resulted from probability summation. These findings suggest that multisensory signals may be preattentively integrated and therefore better disengage spatial attentional resources from the target stimulus, especially when processing resources are already occupied.

On the other hand, a number of studies have also demonstrated the need of attention for successful integration (Alsius, Möttönen, Sams, Soto-Faraco, & Tiippana, 2014; Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007; Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006; Gibney et al., 2017; Hyun, Woodman, & Luck, 2009; Talsma, Doty, & Woldorff, 2007; Talsma & Woldorff, 2005; van Ee, van Boxtel, Parker, & Alais, 2009). For example, two studies investigating the audiovisual integration of speech under dual-task conditions (Alsius et al., 2005, 2007) documented impaired multisensory integration when the participants' processing resources were taxed. These studies used the McGurk effect (McGurk & MacDonald, 1976), a crossmodal illusion arising from the integration of mismatching visual and auditory input. While completing a secondary task that was either in one of the task-relevant modalities (vision or audition; Alsius et al., 2005) or in an irrelevant modality (touch; Alsius et al., 2007), the McGurk effect, and thus audiovisual speech integration, declined significantly. The suggestion was that the secondary task pulled the participant's attention away from the video (despite the fact that the participants were still fixating it) thus leaving insufficient processing resources for the integration of visual and auditory information. These results therefore suggest that the strength of MI depends on the amount of top-down attention that is voluntarily (i.e., endogenously) directed to the task by the participants.

The controversy concerning the relationship between attention and MI is addressed in frameworks that define moderating factors such as stimulus complexity, competition between

stimuli (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010), and perceptual load (Navarra, Alsius, Soto-Faraco, & Spence, 2010; see also Koelewijn, Bronkhorst, & Theeuwes, 2010). According to Talsma and colleagues, MI is modulated by top-down attention in environments where the competition between stimuli is high, or where the stimuli themselves are complex. In these situations, the participant's intentions and goals determine what is being integrated and attended stimuli are integrated first (for example, with audiovisual speech integration as the McGurk effect can be modulated in a top-down manner; see Vatakis & Spence, 2010). On the other hand, when the stimuli are simple, and the competition between stimuli is low, MI precedes attentional selection and operates in a bottom-up fashion. In this case, preattentive integration drives attention to the stimuli. Similarly, perceptual load moderates the relation between MI and attention (Navarra et al., 2010). According to Perceptual Load Theory (Lavie, 1995, 2005, 2010), processing resources are fully used until the capacity limit is reached. Under low load conditions, all stimuli are automatically processed and integrated because the limit has not yet been reached. With increasing load, however, task-relevant stimuli are processed and integrated first and the integration of irrelevant stimuli depends on the remaining processing resources.

Taken together, stimulus competition and perceptual load may modulate whether attention determines MI or whether instead MI operates in an automatic fashion. Thus, if the criteria for MI are fulfilled, attended features (of targets) are integrated regardless of these moderators while features outside the focus of attention (e.g., distractors) are probably not.

Multisensory distractor processing

Crucially, the studies that have been presented so far have all focused on the integration of target features (i.e., stimulus features that are somehow directly task relevant) and the variation in the amount of attention devoted to them. The features that are integrated are al-

ways more or less in the focus of attention because the participants have always been required to respond to them. To the best of our knowledge, to date, there has been no investigation of multisensory distractor integration in a truly multisensory selection task, in which participants have to select a multisensory target against an interfering multisensory distractor. Thus far, it is, in fact, unclear whether the integration of distractor features occurs when they are spatially and temporally aligned and correlated, i.e., when important criteria for MI are fulfilled. One possibility is that integration occurs automatically in a bottom-up manner as in crossmodal studies with unisensory targets. Here, multisensory distractors capture attention more effectively than unisensory distractors in noisy environments most likely due to preattentive integration (Mast, Frings, & Spence, 2015, 2017; Santangelo, Ho, & Spence, 2008; Santangelo & Spence, 2007). Alternatively, however, attention may be needed for the integration of distractor features. Perceptual load in the target task modulates the amount of processing resources available for the distractor (Lavie, 1995, 2005, 2010) and the integration of the multisensory target already requires resources (similar to high load conditions), so the distractor features might not be integrated as the attentional focus is mainly on the target.

In the case that this latter idea holds true and the distractor is not integrated because it receives insufficient attentional resources, varying attention to the distractor might be expected to modulate multisensory integration. Lavie (1997) investigated this question in a unisensory context. She used a flanker task with visual feature conjunctions and manipulated attention to the distractor(s). Targets were composed of two visual features (colored shapes, e.g., a green circle) and flanked by two distractor objects that contained two response-relevant features (either congruent or incongruent with the target) and two neutral features. The relevant features were thereby either conjoint in one distractor object or distributed across the two objects. Further, spatial attention was manipulated in this study by either holding the target location constant or variable. In Experiments 1 and 2, the target was always presented from

the central location, while in Experiment 3, the target location was unpredictable, thus requiring attention to be spread across all three locations. It turned out that there was no difference in the magnitude of congruency effects between conjoint and disrupted distractors when the target location was fixed while a significantly larger congruency effect was reported for conjoint distractors when the location varied. These findings demonstrate that attention modulates distractor integration in unisensory selection; when attended to, distractor features are integrated, while without attention, distractors are processed at the feature-level.

Next to attentional spreading, there are several other ways in which to manipulate attention. Unisensory studies have, for example, demonstrated that fixation or solely the orientation of our head and body guide spatial attention and thereby have an impact on the interference produced by distractors (Driver & Grossenbacher, 1996; Honoré, Bourdeaud'hui, & Sparrow, 1989; Macaluso, Driver, van Velzen, & Eimer, 2005; Macaluso, Frith, & Driver, 2002; Pierson, Bradshaw, Meyer, Howard, & Bradshaw, 1991; Tipper et al., 1998; Wesslein, Frings, & Spence, 2015). For instance, Driver and Grossenbacher manipulated their participants' gaze in a tactile version of the flanker task. Participants' head orientation and gaze was either aligned to the relevant target hand, the irrelevant distractor hand, or a point straight ahead in the middle of the two hands. When facing the relevant hand or straight ahead, these researchers documented less interference by the distractor in a condition in which the hands were placed far apart in comparison to when they were placed close together. When facing the irrelevant hand though, distance was found to have no effect. These findings remained the same when the participants in another experiment were blindfolded, demonstrating that it was not vision *per se*, but rather the direction of gaze that was the crucial factor. So, it would appear that head orientation guided spatial attention: turning towards the distractor made ignoring it more difficult as it received more attentional resources regardless of whether or not covert attention was there (that is supposed to be on the target) or whether vision was prevented.

The present study

In the present study, we investigated whether multisensory distractors are processed as feature-compounds (i.e., features are not processed independently) or whether instead they are only processed on a feature-based (i.e., unisensory) level. Against the background of the literature on multisensory integration and distractor processing, our hypothesis was that overt visual spatial attention might be an important factor for the integration of irrelevant distractor features.

In two experiments, a multisensory flanker task was used with audiovisual stimuli as both targets and distractors. Multisensory target stimuli were created by mapping specific combinations of a visual and an auditory feature on to a particular response. In order to respond correctly, the participants had to integrate the target features (e.g., a red light-high pitch combination required a specific response). In addition, the participants had to select multisensory targets against the background of a multisensory distractor that also had one visual and one auditory feature that were varied orthogonally to each other. Further, the distractor features were also varied orthogonally with respect to the target features. Crucially, we manipulated overt spatial attention by varying the distractor location: In Experiment 1, the distractor was presented laterally and the target was presented at fixation whereas in Experiment 2 the distractor was presented in the center of participant's gaze (at fixation) and the target was presented laterally instead.

This experimental design allowed us to test whether distractor feature dimensions were processed independently and whether multisensory distractor processing was modulated by distractor position relative to the direction of a participant's gaze. When the distractor feature dimensions are not processed independently, congruency effects of one feature should vary with the congruency of the other feature. For instance, the visual feature could elicit a

larger congruency effect when the auditory feature is congruent (i.e., perceptually identical to the auditory target feature) than when it is incongruent (i.e., at a statistical level we expect the interaction of visual and auditory feature congruency to be significant). We interpret this interaction as evidence for multisensory integration as both feature dimensions influence each other. Note, however, that we do not have further predictions concerning the exact pattern of interaction (i.e., we had no predictions about over- or underadditivity as this was not our main focus). By contrast, if each feature produces a congruency effect independently of the other feature's congruency (i.e., we obtain main effects of congruency but no interaction), it could be argued that multisensory distractors are processed at the level of individual features and integration is therefore weak.

Experiment 1

Method

Participants. In the literature, and based on our own laboratory research, flanker effects typically elicit medium effects sizes (irrespective of the modality in which the stimuli are presented). To make sure that we at least observed a unisensory flanker effect, we therefore planned our study with a medium effect size in mind (d_z around .5). Given an alpha-level of .05 and a desired power of at least $1-\beta > .80$, we aimed for a minimum of 27 participants (power analyses were run with G-power 3.1.7; Erdfelder, Faul, & Buchner, 1996). Thirty students from the University of Trier took part in the study in return for course credit or were paid 6€ (~7.5\$). One participant was excluded due to extremely high error rates in the catch trials presumably indicating a lack of understanding of the task (this participant was an outlier with respect to the errors in catch trials of the sample, Mean = 17%, SD = 11%, outlier with 76% errors). Another participant was excluded due to extremely high error rates (again this participant was an outlier with respect to error rates of the sample, Mean = 4%, SD = 3%, outlier

1 with 30% errors). A total of 28 participants (4 male) aged from 20 to 35 years (Mean = 24,
 2 SD = 6) were used for analyses. All of the participants had normal or corrected-to-normal
 3 vision and audition.²

4
 5 *Figure 1.* Experimental set-up used in Experiment 1 and 2 (right) showing the three
 6 speaker positions (central, lateral left, and lateral right), and a sample timeline (in ms) in Ex-
 7 periment 1 (left). Note that a LED was affixed in front of each loudspeaker.

8
 9 *Apparatus and materials.* The participants were seated in a dark soundproofed chamber in
 10 front of a 24" CRT computer screen with a refresh rate of 60 Hz. The distance to the screen
 11 (63 cm) was kept constant by means of a chin-rest with forehead support. The instructions
 12 were presented in grey letters against a black background in the middle of the screen. Combi-
 13 nations of three different tone pitches (low presented at 440 Hz, medium presented at 880 Hz,

² All of the analyses were also conducted with the full sample (N=30). Note that the results showed exactly the same pattern of results as the analyses reported in the main text.

and high presented at 1760 Hz) and three different colors, namely Red (HSV: 0, 100, 12.55), Green (HSV: 120, 100, 12.55), and Blue (HSV: 240, 100, 12.55) were used as stimuli. These stimuli were presented over two loudspeakers (Logitech S-120 2.0 OEM, black) and two LEDs affixed on the lower part of the speakers. One speaker was placed in the center of participant's gaze just below the screen, the other one was standing laterally on the right or left side. There was a distance of 40 cm between the speakers (this corresponds to a visual angle of 35°). See Figure 1 for a bird's-eye view of the experimental set-up. The participants had to fixate the speaker in the center throughout the whole experiment. Responses were given by means of foot pedals using Chronos®. The experiment was conducted with the E-Prime software (Version 2.0) and the data were analyzed with IBM SPSS Statistics (Version 23).

Design. The experiment consisted of a 3×3 repeated measurement design with the variables Visual Distractor Feature Congruency (congruent, neutral, and incongruent with respect to the visual target feature) and Auditory Distractor Feature Congruency (congruent, neutral, and incongruent with respect to the auditory target feature). Response times as well as error rates were used as dependent variables.

Procedure. The participants were tested individually in a soundproof and dimly-lit chamber. They sat 63 cm from the display with their head resting on a chinrest. They were instructed to fixate the speaker placed in the center of gaze throughout the experiment. The participant's feet were placed on foot-pedals. The participant's task was to identify the target via foot-pedal responses. There were two multisensory target stimuli, a red light combined with a high pitch, which required the participants to depress the left foot pedal, and a blue light combined with a low pitch, which required the participants to depress the right foot pedal. The participants were instructed to respond to these target stimuli presented at the central speaker as rapidly and accurately as possible while they were told to ignore the stimuli presented laterally to either their right or left side (i.e., the distractors). In order to ensure that the

participants processed both target features, and did not just respond to one sensory feature, catch stimuli with reversed target feature assignment (i.e., red light and low pitch or blue light and high pitch) were also presented from the target speaker. The participants were instructed not to respond to these stimuli thus turning the task into a go-/no go-task. These catch trials accounted for 20% of all trials. To boost the motivation to respond as rapidly and accurately as possible, the participants were told that all trials on which they responded erroneously, or else responded too slowly, would be attached at the end of each block. In fact, only catch trials to which the participants responded erroneously were repeated in the end. Therefore, the number of trials completed by each participant varied (due to the difference in the number of catch trials) while the number of target trials remained constant for all participants.

		Color		
Pitch		Congruent (red)	Neutral (green)	Incongruent (blue)
Target	Congruent (high)
	Neutral (middle)
	Incongruent (low)
	No distractor

Figure 2. The ten different distractor conditions in Experiment 1 and 2 in regard of the congruency of each the visual (line) and the auditory (column) feature exemplary for the red-color-high-pitch-target. The identity of the feature is shown in parentheses.

Distractor stimuli were created as combinations of a visual and an auditory feature out of three possible visual and auditory features respectively (two target features and one response-neutral feature each). A green light and a medium pitch served as neutral features that were not mapped to a response. We implemented these neutral features as they emphasize the irrelevance of the distractor to the participants that now is not only supporting or interfering. Additionally, there were trials in which no distractor was presented. The visual and auditory feature of the distractors could either be congruent (i.e., perceptually identical), neutral, or incongruent to the corresponding feature of the current target and were varied orthogonally to each other. Figure 2 depicts an example of possible distractors for a red light high pitched target. The ten different distractor conditions were conducted with an equal number of trials (32 trials throughout the experimental phase). The side of the distractor (left, right) was manipulated within-participants in two blocks with 200 trials each (plus the incorrect catch trials) while the order (left first, right first) was balanced between-participants. An additional break was offered in the middle of each block.

The exact stimulus sequence of each trial (see Figure 1) was as follows: a cross was presented at the bottom of the screen just above the central speaker indicating the start of the trial, followed by the onset of the distractor after 250 ms. The target was presented 100 ms after the onset of the distractor³. Both stimuli then lasted for 1000 ms and the participants had to respond within 1500 ms of distractor onset. Feedback was only given for false or missing responses and was displayed for 500 ms. In order to reduce any after-effects from the previous trial on the present trial, an eraser stimulus consisting of white noise and a white flash (HSV: 0, 0, 100) was included in the trial sequence. This stimulus was presented after a 500 ms black screen and appeared on both loudspeakers. The stimulus lasted for 500 ms and was followed by a Response-Stimulus-Interval (RSI) of 500 ms.

³ We presented the distractor 100 ms before the target to maximize congruency effects. Congruency effects, unisensorially as well as crossmodally, increase when the distractor precedes the target and peak at about 100 ms SOA (Flowers & Wilcox, 1982; Shore, Barnes, & Spence, 2006).

Before the experiment, instructions were shown on the screen. The participants were encouraged to ask the experimenter (who was in the control room next to the experimental room during the whole experiment), if anything was unclear. The experiment started with a learning phase in which the stimulus-response mapping was presented and practiced for 8 trials. After that, a first practice phase was run with 40 trials in which the target and catch trials both accounted for 50% of all trials and were presented on the target speaker while there were no stimuli on the distractor speaker. The second practice phase with 100 trials was exactly the same as the experimental phase (20% catch trials) except for the fact that no catch trials were attached in the end. In total, the experiment lasted for about 45 minutes.

Results

First, we checked whether participants worked through the task properly by examining the catch trials. We excluded the catch trials that were attached in the end of each block after producing an error. On average, participants responded correctly (meaning that they did not react) to 83% of these trials (range from 60-98%). This shows that they processed both features on the target speaker.

For the analysis of the reaction time (RT) data, only those trials in which the participants responded correctly to the target were considered. All trials in which the RT was below 200 ms, as well as those trials with an RT that was 1.5 interquartile ranges above the third quartile of each participant's individual RT distribution (Tukey, 1977), were excluded from the data analysis as well. In total, 6.2% of the trials were excluded from the analysis due to these restrictions. Table 1 depicts the mean RT and error data.

Table 1. Mean RT (in milliseconds; error rates in % in parentheses) as a function of visual and auditory distractor feature congruency in Experiment 1.

Auditory distractor feature	Visual distractor feature			Mean
	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	
<i>Congruent</i>	699 (1.3)	719 (1.2)	740 (1.3)	720 (1.3)
<i>Neutral</i>	777 (5.1)	783 (4.1)	815 (4.4)	792 (4.5)
<i>Incongruent</i>	799 (7.8)	802 (6.6)	820 (7.4)	807 (7.3)
Mean	759 (4.8)	768 (4.0)	792 (4.4)	

Reaction times. A 3 (Visual Distractor Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Auditory Distractor Feature Congruency: congruent vs. neutral vs. incongruent) multivariate analysis of variance (MANOVA) with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(2, 26) = 21.00$, $p < .001$, $\eta_p^2 = .62$; that is, the participants responded more rapidly when the visual target and distractor feature were congruent than when they were incongruent. The main effect of Auditory Distractor Feature Congruency was significant, $F(2, 26) = 43.02$, $p < .001$, $\eta_p^2 = .77$; that is, participants responded most rapidly when the auditory target and distractor feature were congruent and slowest when they were incongruent. Critically, there was no significant interaction between these two variables, $F(4, 24) = 1.213$, $p = .331$, $\eta_p^2 = .168$, thus suggesting that the two sensory distractor features were processed *independently* of one another.⁴

⁴ In order to account for any sequential effects that may arise due to the fact that the participants were supposed not to respond on 20% of the trials, we excluded the first trial of each block and all trials following a catch trial and reran the analysis. In the RT data, we obtained two significant main effects (visual: $F(2, 26) = 19.79$, $p < .001$, $\eta_p^2 = .60$; auditory: $F(2, 26) = 48.89$, $p < .001$, $\eta_p^2 = .79$) and a non-significant interaction, $F(4, 24) = 1.213$, $p = .331$, $\eta_p^2 = .168$.

We further checked in an additional $2 \times 2 \times 3 \times 3$ MANOVA with Distractor Side (left vs. right) as within- and Order (left vs. right first) as the between-participants variable, whether the two-way interaction of Auditory Distractor Feature Congruency $\times$ Visual Distractor Feature Congruency was further specified by a higher-order interaction. Yet, neither the three-way nor the four-way interactions were significant (all $ps > .594$).

Figure 3. Mean RT (in ms) as a function of visual and auditory distractor feature congruency (without neutral condition) in Experiments 1 (left) and 2 (right). The error bars depict the standard error of the mean.

The 3×3 MANOVA is, however, somewhat hard to interpret as both factors have three levels (see Jonides & Mack, 1984, for a review on the difficulty of design-

2.217, $p = .097$, $\eta_p^2 = .27$. For the error rates, we obtained a significant effect of the auditory feature, $F(2, 26) = 13.63$, $p < .001$, $\eta_p^2 = .51$, but again, no effect of the visual feature, $F(2, 26) = 1.94$, $p = .164$, $\eta_p^2 = .13$. In contrast to the original pattern, we obtained a marginally significant interaction, $F(4, 24) = 2.66$, $p = .057$, $\eta_p^2 = .31$. The mean RT and error rates are depicted in the Appendix Table 3.

ing/interpreting neutral conditions). Thus, we ran a 2×2 MANOVA without the neutral condition and obtained a significant main effect of Visual Distractor Feature Congruency, $F(1, 27) = 24.59, p < .001, \eta_p^2 = .48$, and of Auditory Distractor Feature Congruency, $F(1, 27) = 45.92, p < .001, \eta_p^2 = .63$. Crucially, the interaction was still not significant, $F(1, 27) = 3.31, p = .08, \eta_p^2 = .11$, suggesting that each modality produced congruency effects *independent* of the other one (see Figure 3).⁵ (Alternatively, however, one may compute costs and benefits using the neutral condition as a baseline; see Appendix for this analysis-strategy.)

Errors. For the error rates, we conducted the same 3×3 MANOVA as for the RT data. There was a main effect of Auditory Distractor Feature Congruency, $F(2, 26) = 14.42, p < .001, \eta_p^2 = .53$, but no main effect of Visual Distractor Feature Congruency, and no interaction, $ps > .302$. In the additional $2 \times 2 \times 3 \times 3$ MANOVA with Distractor Side (left vs. right) as the within- and Order (left vs. right first) as the between-participants variable, there were no significant effects regarding these two variables and our crucial two-way interaction, all $ps > .233$. The 2×2 MANOVA revealed a significant main effect of Auditory Distractor Feature Congruency, $F(1, 27) = 17.18, p < .001, \eta_p^2 = .39$, but again, no main effect of Visual Distractor Feature Congruency, and no interaction, $ps > .766$.

Discussion

In Experiment 1, multisensory targets had to be selected against a background of multisensory distractors. We used a multisensory variant of the flanker task, in which the stimuli had a visual and an auditory feature and participants had to respond to specific combinations

⁵ As for the 3×3 MANOVA, we conducted a 2×2 MANOVA with excluding the first trial of each block and all trials following a catch trial to account for sequential effects. In the RT data, we obtained two significant main effects (visual: $F(1, 27) = 31.53, p < .001, \eta_p^2 = .54$; auditory: $F(1, 27) = 63.65, p < .001, \eta_p^2 = .70$) and a non-significant interaction, $F(1, 27) = .004, p = .953, \eta_p^2 = .00$. For the error rates, we obtained a significant main effect of Auditory Distractor Feature Congruency, $F(1, 27) = 24.05, p < .001, \eta_p^2 = .47$, but no other effects, all $ps > .114$.

of features. The features of the distractors were varied orthogonally to each other and to the target features. In Experiment 1, the centrally-presented target was fixated while the distractor was presented laterally. Congruency effects were obtained for each of the two distractor features showing that the distractor interfered with target processing. The participants responded more rapidly and made fewer errors when one feature was congruent than when it was incongruent. Critically, the interaction failed to reach significance in both the RT and the error data thus indicating that the two features were rather processed independently from one another. Although several criteria for MI (spatial and temporal co-occurrence, temporal correlation) were fulfilled, the distractor features were not easily integrated. However, in the analysis where we excluded the trials following a catch trial, we found a marginally significant interaction in the error data and also a $p < .10$ in the RT data. Thus, we do not want to imply that integration is completely absent in this setting.

With respect to the literature on the influence of spatial attention on MI and distractor processing (e.g., Driver & Grossenbacher, 1996; Koelewijn et al., 2010; Lavie, 1997; Navarra et al., 2010; Talsma et al., 2010), the MI of distractors in a multisensory context might depend on spatial attention. Obviously, the distractor was processed and influenced target processing, as evidenced by the main effects of congruency. Yet, as overt and covert attention were both focused on the target, the resources remaining for distractor processing were probably not sufficient for a strong MI effect to take place, and so the interaction failed to reach significance.

To further test whether increasing resources on distractor processing leads to better integration, we changed the attentional focus. In Experiment 2, the distractor was now presented centrally in participants' gaze while the target appeared laterally. The participants were instructed to fixate the distractor throughout the whole experiment thereby allocating visual spatial attention on the distractor stimulus. If MI depends on spatial attention, the features

should be better integrated in this situation and we should find a strong interaction between visual and auditory distractor feature congruency.

Experiment 2

Method

Participants. Following the same logic concerning power calculation as in Experiment 1, we again aimed for a sample of at least 27 participants. Thirty students from the University of Trier took part in the study in return of course credit or else were paid 6€ (~ 7.5\$). One student was excluded due to extremely high error rates in the catch trials indicating a lack of understanding of the task (this participant was an outlier with respect to the errors in catch trials of the sample, Mean = 23%, SD = 15%, outlier with 95% errors). Another student was excluded due to extremely high error rates and an extremely high loss of trials (again this participant was an outlier with respect to error rates of the sample, Mean = 8%, SD = 7%, outlier with 36% errors, and the loss of trials of the sample, Mean = 7%, SD = 6%, outlier 43% loss of trials). A total of 28 participants (3 male) aged from 19 to 27 years (Mean = 23, SD = 2) were used for analyses. All of the participants had normal or corrected-to-normal vision and audition.⁶

Apparatus, materials, and procedure. The apparatus and materials used were exactly the same as in Experiment 1. Procedure, task, and trial sequence were exactly the same as in Experiment 1 except for one major change. The targets (and catch trials) were now presented at the lateral speaker (i.e., participants' right or left side) while the distractor stimuli were presented at the central speaker. Critically, the participants still had to fixate the central speaker presenting now the stimuli that they had to ignore.

⁶ Once again, all of the analyses were conducted with the full sample (N=30). Exactly the same pattern of results was obtained as reported in the main text.

Design. The experiment consisted of a 3×3 repeated measures design with the variables Visual Distractor Feature Congruency (congruent, neutral, and incongruent) and Auditory Distractor Feature Congruency (congruent, neutral, and incongruent). Response times as well as error rates were used as the dependent variables.

Results

Once again, we first checked whether the participants worked through the task properly by examining the catch trials (again, without the attached trials). On average, the participants responded correctly in 77% of these trials (range from 48-96%) showing that participants processed both features on the relevant speaker in Experiment 2 as well.

Table 2. Mean RT (in milliseconds; error rates in % in parentheses) as a function of visual and auditory distractor feature congruency in Experiment 2.

Visual distractor feature					
Auditory	distractor	Congruent	Neutral	Incongruent	Mean
feature					
Congruent		719 (2.6)	766 (2.6)	795 (5.5)	760 (3.5)
Neutral		799 (7.3)	801 (5.9)	831 (5.6)	810 (6.3)
Incongruent		850 (19.6)	848 (17.2)	853 (16.2)	850 (17.7)
Mean		790 (9.8)	805 (8.6)	826 (9.1)	

For the analysis of the reaction time (RT) data, only those trials in which the participants responded correctly to the target were considered. All trials in which the RT was shorter than 200 ms, as well as those trials with an RT that was 1.5 interquartile ranges above the

third quartile of each participant's individual RT distribution (Tukey, 1977), were excluded from the data analysis. In total, 10.7% of the trials were excluded from the analysis due to these restrictions. Table 2 depicts the mean RT and error data.

Reaction times. A 3 (Visual Distractor Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Auditory Distractor Feature Congruency: congruent vs. neutral vs. incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(2, 26) = 19.41, p < .001, \eta_p^2 = .60$, that is, participants responded most rapidly when the visual target and distractor feature were congruent and slowest when they were incongruent. The main effect of Auditory Distractor Feature Congruency was significant, $F(2, 26) = 34.47, p < .001, \eta_p^2 = .73$; that is, participants responded fastest when the auditory target and distractor feature were congruent and slowest when they were incongruent. Critically, there was a significant interaction between the Visual Distractor Feature Congruency and the Auditory Distractor Feature Congruency, $F(4, 24) = 11.49, p < .001, \eta_p^2 = .66$, suggesting that distractor features were *integrated*.⁷

In the additional $2 \times 2 \times 3 \times 3$ MANOVA with Distractor Side (left vs. right) as the within- and Order (left first vs. right first) as the between-participants variable, there were no significant effects regarding these two variables and our crucial two-way interaction, all $ps > .117$.

As for Experiment 1, we followed the 3×3 MANOVA with a 2×2 MANOVA without the neutral condition in order to directly assessing the congruency effects and basically

⁷ Again, to account for sequential effects, we conducted the same analysis with excluding the first trial of each block and all trials following a catch trial. In the RT data, we obtained two significant main effects (visual: $F(2, 26) = 13.35, p < .001, \eta_p^2 = .51$; auditory: $F(2, 26) = 30.56, p < .001, \eta_p^2 = .70$) and a significant interaction, $F(4, 24) = 13.35, p < .001, \eta_p^2 = .69$. For the error rates, we obtained a significant main effect of the auditory feature, $F(2, 26) = 9.72, p = .001, \eta_p^2 = .43$, and a significant interaction, $F(4, 24) = 7.87, p < .001, \eta_p^2 = .57$, but again, no effect of the visual feature, $F(2, 26) = 1.446, p = .254, \eta_p^2 = .10$. The mean RT and error rates are depicted in the Appendix Table 5.

obtained the same pattern as in the original analysis. Both main effects were significant, $F(1, 27) = 34.10, p < .001, \eta_p^2 = .56$, for Visual Distractor Feature Congruency, and $F(1, 27) = 64.99, p < .001, \eta_p^2 = .71$, for Auditory Distractor Feature Congruency, and we obtained a significant interaction, $F(1, 27) = 47.00, p < .001, \eta_p^2 = .64$, suggesting that in Experiment 2 distractor features' processing was depending each other (see Figure 3).⁸

Errors. For the error rates, the same 3×3 MANOVA was conducted as for the RT data. There was a main effect of Auditory Distractor Feature Congruency, $F(2, 26) = 10.61, p < .001, \eta_p^2 = .45$, and also a significant interaction term, $F(4, 24) = 3.33, p = .027, \eta_p^2 = .36$, but no main effect of Visual Distractor Feature Congruency, $F(2, 26) = .981, p = .388$. In the additional $2 \times 2 \times 3 \times 3$ MANOVA with Distractor Side (left vs. right) as the within- and Order (left first vs. right first) as the between-participants variable, there were again no significant interactions between the interesting effects and these two control variables, all $ps > .075$. The 2×2 MANOVA without the neutral condition revealed a significant main effect of Auditory Distractor Feature Congruency, $F(1, 27) = 17.23, p < .001, \eta_p^2 = .39$, a significant interaction, $F(1, 27) = 7.89, p = .009, \eta_p^2 = .23$, and again no visual main effect, $F(1, 27) = .046, p = .832, \eta_p^2 = .002$.

Comparison between Experiments. As we obtained different patterns of results in Experiments 1 and 2 regarding the interaction between Visual and Auditory Distractor Feature Congruency, the patterns of results were compared in a further analysis with Experiment as an additional between-participant variable. A 3 (Visual Distractor Feature Congruency: congru-

⁸ As for the 3×3 MANOVA, we conducted a 2×2 MANOVA with excluding the first trial of each block and all trials following a catch trial to account for sequential effects. In the RT data, we obtained two significant main effects (visual: $F(1, 27) = 42.59, p < .001, \eta_p^2 = .61$; auditory: $F(1, 27) = 42.62, p < .001, \eta_p^2 = .61$) and a non-significant interaction, $F(1, 27) = 1.434, p = .242, \eta_p^2 = .05$. For the error rates, the pattern was basically the same. We obtained two significant main effects (visual: $F(1, 27) = 18.32, p < .001, \eta_p^2 = .40$; auditory: $F(1, 27) = 12.44, p = .002, \eta_p^2 = .32$) and a non-significant interaction, $F(1, 27) = 2.726, p = .110, \eta_p^2 = .09$. These results contrast our main analyses, but note that we have an enormous loss of trials (68% in comparison to the original 3×3 MANOVA), which significantly reduces the power of this analysis.

ent vs. neutral vs. incongruent) $\times$ 3 (Auditory Distractor Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 2 (Experiment: Exp. 1 distractor laterally vs. Exp. 2 distractor centrally) multivariate analysis of variance (MANOVA) with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant three-way-interaction, $F(4, 51) = 3.34, p = .017, \eta_p^2 = .21$, showing that the interaction of Visual and Auditory Distractor Feature Congruency differs significantly between Experiments 1 and 2.⁹

To compare the analyses without the neutral condition, we further a conducted a 2 (Visual Distractor Feature: congruent vs. incongruent) $\times$ 2 (Auditory Distractor Feature: congruent vs. incongruent) $\times$ 2 (Experiment) MANOVA with Pillai's trace as the criterion and mean RT as the dependent variable. Again, we documented a significant three-way-interaction, $F(1, 54) = 12.27, p = .001, \eta_p^2 = .19$, showing that the interaction of Visual and Auditory Distractor Feature Congruency differs significantly between Experiments 1 and 2, also without the neutral condition. Together, these results suggest that sensory distractor features are processed depending each other when they are presented at fixation but processed more independently when they are presented laterally.

Discussion

In Experiment 2, the multisensory distractor now was presented at the central speaker that participants had to fixate while they had to respond to the laterally-presented multisensory target. Once again, we found the two main effects of congruency for the auditory and visual distractor features in the RT data and an auditory congruency effect in the error data. In contrast to Experiment 1, however, we found a strong significant interaction between the two

⁹ In the $3 \times 3 \times 2$ analysis where we excluded the trials following a catch trial, we also obtained a significant three-way-interaction, $F(4, 51) = 3.44, p = .015, \eta_p^2 = .21$, allowing the same conclusions.

distractor features in the RT as well as in the error data. In Experiment 2, the processing of one feature interacted with the other feature's processing. Comparing the results of Experiments 1 and 2 revealed a significant difference between the interaction effects in these two experiments. Taken together, the results of the two experiments reported here demonstrate that multisensory distractor processing is modulated by overt spatial attention. When the distractor is presented laterally, its sensory features are processed more independently of one another while they do highly interact when the distractor is presented in the center of gaze (i.e., at fixation).

General Discussion

In two experiments, we set out to investigate multisensory distractor processing and its possible modulation by overt visual spatial attention in a multisensory selection situation. We implemented multisensory stimuli in a flanker task (Eriksen & Eriksen, 1974; see also Chan, Merrifield, & Spence, 2005) where participants had to classify audiovisual targets presented at a specific target location. To make people process both target features, trials with reversed feature assignment (i.e., catch trials) were added in which the participants had to withhold their responses (thereby forcing participants always to process both modalities together in order to determine the correct target response). At the same time, however, irrelevant multisensory distractor stimuli were presented from a different location. Critically, the direction of overt visual spatial attention to the distractor location was manipulated between the two experiments: the distractor was either presented centrally (at fixation) or laterally on either the left or right side. This manipulation significantly influenced performance. When the distractor was presented laterally, we found significant congruency effects of visual and auditory distractor features, respectively, that were independent of each other (i.e., the interaction failed to reach statistical significance). However, when the distractor was presented at fixation in the

center of participants' gaze instead, a significant interaction of the two distractor feature dimensions was observed suggesting feature integration.

These findings suggest that MI of an irrelevant stimulus is modulated by overt visual spatial attention. On the basis of previous studies (Driver & Grossenbacher, 1996; Honoré et al., 1989; Macaluso et al., 2002, 2005; Pierson et al., 1991; Tipper et al., 1998), we know that gaze can guide the focus of visual spatial attention. We used this property to vary the amount of spatial attention allocated to the distractor. The main effects of the two distractor features found in both experiments reveal that the distractor is always processed at the level of separate features and that each feature produces interference based on its congruency. How strong the features' processing is further interacting, however, is modulated by visual spatial attention.

When not at the center of gaze, the distractor features do not interact while still individually producing reliable interference effects on their own. In this condition, overt and covert spatial attention are both presumably focused on the target, so that the distractor may not receive enough processing resources for multisensory integration to take place. As has been presumed since the 1980's, attention might be a precondition for integrating features (starting with visual feature integration theory, Treisman & Gelade, 1980). Yet, of course, nowadays we know that the story is rather more complex than initially suggested. Assuming that similar mechanisms account for intramodal and crossmodal binding (e.g., Zmigrod, Spapé, & Hommel, 2009), it would seem that only when the task is complex are attentional resources needed for feature integration (e.g., Talsma et al., 2010). Given that multisensory selection requires participants to integrate auditory and visual target features in order to compute the correct motor response, we argue that our distractors do not attract enough attention and hence integration processes are too weak when they are not in the focus gaze.

If the distractor was fixated, however, the feature dimensions interacted, that is, visual congruency effects varied as a function of auditory congruency between the distractor and target (and the other way round). While covert attention was still on the target, overt attention was now directed toward the distractor as participants were oriented toward and fixated the distractor. As a result, we found feature integration as attentional resources were sufficient now.

In view of the marginally significant interaction in the error data in Experiment 1 after excluding the trials following a catch trial, we cannot say whether distractor integration is completely absent when participant's focus is on the target. We rather propose a continuous approach and argue that the impact of attention on integration is not simply absent or present but instead is a continuous one. This view is supported by the three-way-interaction found in the comparison between experiments that constitutes an *ordinal* interaction. This means that the pattern in Experiment 1 does not point into the opposite direction and one can only say that integration in Experiment 2 is stronger than in Experiment 1. The data do not warrant concluding that focusing a stimulus is a precondition for multisensory integration. To test this idea in future research, a manipulation of the distance between target and distractor in the sense of an attentional gradient would be helpful.

Further, when a neutral condition is included, thinking about the origins of the effects, namely facilitation or interference, seems to be the next step. Yet, in our design, considering this issue is rather complex. Based on our previous experience, the neutral condition seems to fall rather unpredictably in-between congruent and incongruent and therefore does not constitute a highly valid measure to compute costs and benefits. One reason for this is that the neutral features had a different frequency in the experiment and are probably treated differently than the features with response association. Nevertheless, we implemented a neutral condition as this emphasizes the irrelevance of the distractor to the participants. (We did run an addi-

1 tional analysis where we compute modality-specific costs and benefits despite the difficulty to
2 interpret them; see Appendix.)

3 The present findings are in line with previous studies on unisensory selection (Driver
4 & Grossenbacher, 1996; Lavie, 1997). Lavie's investigation as well as ours found attention to
5 modulate the integration of distractor features. In contrast to Lavie, we used overt orienting
6 instead of attentional spreading across different locations as the attention manipulation and
7 one might wonder why we did not adopt a similar approach. The reason lies in the multisen-
8 sorial character of our study: the auditory modality differs in many characteristics from the
9 visual modality. In the sense of the modality-appropriateness-hypothesis, the spatial resolu-
10 tion in the auditory modality is much worse than in the visual modality while auditory tem-
11 poral resolution is much better than the visual one (Aschersleben & Bertelson, 2003; Ward,
12 McDonald, & Lin, 2000; Welch & Warren, 1980). Adding a second distractor and/or varying
13 the target location would have made the task incredibly difficult and probably impossible to
14 solve as distinguishing between target and distractor already was hard in this task. These mo-
15 dality differences can also explain the asymmetry in the congruency effects between auditory
16 and visual modality. Auditory congruency effects were larger in both experiments compared
17 to visual congruency effects. Previous studies showed that congruency effects are typically
18 larger for the modality that is not dominating in the current task (Kreutzfeldt, Stephan, Sturm,
19 Willems, & Koch, 2015; Lukas, Philipp, & Koch, 2010, 2014). As in our task target and dis-
20 tractor were predominantly segregated by location and vision is dominant, largest congruency
21 effects were found for the auditory modality.

22 Our study provides new insights into the limits of multisensory processing. Previous
23 research in this field primarily focused on target processing. As targets are responded to and
24 therefore always receive attention, investigating the role of attention on MI only on target
25 processing is simply not sufficient. Even when the attentional demands are high, the target is

intentionally attended to; this is not the case for distractors. Investigating ignored multisensory objects can improve the investigation on the influence of attention on MI, and the circumstances determining MI (see Wesslein, Spence, & Frings, 2014, for an analogue argument in the tactile modality). Our results indicate that that attention is the crucial factor for successful integration in a multisensory selection situation. Although the distractor features were temporally and spatially aligned with each other, temporally correlated and spatially and temporally segregated from the target features, they were not simply integrated when the distractor was not presented in the center of participant's gaze. Regarding the relationship between MI and attention as discussed in recent frameworks on MI (Koelewijn et al., 2010; Navarra et al., 2010; Talsma et al., 2010), attention is thought to modulate MI in noisy environments. Additionally, effects of perceptual load are more pronounced when target and distractor are spatially segregated (see Lavie, 2005). Multisensory selection with spatially segregated targets and distractors both comprising features from two modalities might therefore represent not a simple but rather a complex and perceptually loaded situation and thus, attention is required for integration in this case.

Concerning the processing stage at which MI takes place in this context, we argue that late integration is likely. The literature differentiates between early and late MI (Calvert & Thesen, 2004; De Meo, Murray, Clarke, & Matusz, 2015; Koelewijn et al., 2010). De Meo and colleagues (2015) define a multisensory process that occurs within the first 100 milliseconds (ms) as early-latency multisensory integration. This integration process is a rather automatic, stimulus-driven, bottom-up process independent of top-down attention. Late MI, on the other hand, enacts at a later point in time and is modulated by top-down attentional resources. Given that we used a flanker task in which distractors interfered with targets at a response and perceptual level, it seems clear that our results only hold true as an instance of late MI – fitting to the assumption that only late MI is influenced by attention (Koelewijn et al., 2010).

Moreover, at later processing stages, a distinction between different pathways can be made. Like for the visual system (Goodale & Milner, 1992), there has also been a distinction between “what/how”- and “where”-pathway for the auditory, tactile and multisensory domain (e.g., Arnott, Binns, Grady, & Alain, 2004; Chan & Newell, 2008; Renier et al., 2009; Werner & Noppeney, 2010). As the task used in the present study based on the identity of the stimuli, processing might have occurred primarily on the “what”-pathway (although target and distractor were segregated by space, so the “where”-pathway must also be involved). Our results therefore also fit with the assumption that integration on the “what”-pathway only occurs with attention (Spence, 2013).

Yet, one might wonder whether our gaze manipulation only affected participants’ visual spatial attention or also perceptual acuity which, in turn, might have led to a stronger signal for the focused stimulus. One could argue that the distractor is not integrated outside fixation because the signals are simply too weak. This, however, would indicate that signal strength is a prerequisite for MI and not attention. We think this alternative explanation is less plausible for two reasons. First, the law of inverse effectiveness suggests that maximal integration is found when one or both of the unisensory signals are weakly effective on their own (Stein & Meredith, 1990, 1993). Therefore, one would expect weaker stimuli (the distractor when it is not in the focus) to provoke more integration, which is not the case. Second, if signal strength did play a role here, one would expect that, besides the interaction, the congruency main effects of the visual and auditory feature would be affected as well. However, the main effects for visual or auditory congruency (as the difference between incongruent and congruent trials) did not vary as a function of the experiment (visual: $D = 4$ ms, $t(54) = .446$, $p = .658$; auditory: $D = 2$ ms, $t(54) = .165$, $p = .869$) thus suggesting that the interference of the visual and the auditory feature *per se* was identical in both conditions.

Conclusions

1 Multisensory selection was studied by investigating the processing of multisensory
2 distractors thereby de-confounding the possible influence of visual spatial attention. Multisen-
3 sory distractor processing was modulated by overt visual spatial attention: distractor features
4 were dependently processed when distractors were presented in the focus of attention whereas
5 distractor features were independently processed outside the focus of attention. This finding
6 suggests that attention furthers feature integration in multisensory selection and that in com-
7 plex environments, many multisensory inputs are actually processed at the level of single fea-
8 tures – our world is only multisensory where we focus our attention.

References

- Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology*, 5:727.
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high 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.
- Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, 22, 401-408.
- Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology*, 50, 157-163.
- Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, 62, 321-332.
- Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile “capture” of attention. *Perception & Psychophysics*, 64, 616-630.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, 98, 191-205.

1 Chan, J. S., Merrifield, K., & Spence, C. (2005). Auditory spatial attention assessed in
2 a flanker interference task. *Acta Acustica*, 91, 554-563.

3 Chan, J. S., & Newell, F. N. (2008). Behavioral evidence for task-dependent “what”
4 versus “where” processing within and across modalities. *Perception & Psychophysics*, 70, 36-
5 49.

6 Chen, Y.-C., & Spence, C. (2017). Assessing the role of the ‘unity assumption’ on
7 multisensory integration: A review. *Frontiers in Psychology*, 8:445.

8 Colonus, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction
9 time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, 16, 1000-
10 1009.

11 De Meo, R., Murray, M. M., Clarke, S., & Matusz, P. J. (2015). Top-down control and
12 early multisensory processes: Chicken vs. egg. *Frontiers in Integrative Neuroscience*, 9:17.

13 Driver, J., & Grossenbacher, P. G. (1996). Multimodal spatial constraints on tactile se-
14 lective attention. In T. Inui & J. I. McClelland (Eds.), *Attention and performance XVI: In-*
15 *formation integration in perception and communication* (pp. 209-235). Cambridge, MA: MIT
16 Press.

17 Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis
18 program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.

19 Eriksen, B. A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification
20 of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.

21 Flowers, J. H., & Wilcox, N. (1982). The effect of flanking context on visual classifi-
22 cation: The joint contribution of interactions at different processing levels. *Perception & Psy-*
23 *chophysics*, 32, 581-591.

1 Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search
2 for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society*,
3 273, 865-874.

4 Gibney, K. D., Aligbe, E., Eggleston, B. A., Nunes, S. R., Kerkhoff, W. G., Dean, C.
5 L., & Kwakye, L. D. (2017). Visual distractors disrupt audiovisual integration regardless of
6 stimulus complexity. *Frontiers in Integrative Neuroscience*, 11:1.

7 Girard, S., Collignon, O., & Lepore, F. (2010). Multisensory gain within and across
8 hemispaces in simple and choice reaction time paradigms. *Experimental Brain Research*, 214,
9 1-8.

10 Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and
11 action. *Trends in Neurosciences*, 15, 20-25.

12 Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of
13 modality-specific attention. *Journal of Vision*, 8:21.

14 Hirst, R., Cragg, L., & Allen, H. (in press). Vision dominates in adults but not chil-
15 dren: A meta-analysis of the Colavita effect. *Neuroscience & Biobehavioural Reviews*.

16 Honoré, J., Bourdeaud'hui, M., & Sparrow, L. (1989). Reduction of cutaneous reac-
17 tion time by directing eyes towards the source of stimulation. *Neuropsychologia*, 27, 367-371.

18 Hyun, J., Woodman, G. F., & Luck, S. J. (2009). The role of attention in the binding of
19 surface features to locations. *Visual Cognition*, 17, 10-24.

20 Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psycholog-
21 ical Bulletin*, 96, 29-44.

1 Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple
2 stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134,
3 372-384.

4 Kreutzfeldt, M., Stephan, D. N., Sturm, W., Willems, K., & Koch, I. (2015). The role
5 of crossmodal competition and dimensional overlap in crossmodal attention switching. *Acta*
6 *Psychologica*, 155, 67-76.

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

9 Lavie, N. (1997). Visual feature integration and focused attention: Response competi-
10 tion from multiple distractor features. *Perception & Psychophysics*, 59, 543-556.

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

13 Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Di-*
14 *rections in Psychological Science*, 19, 143-148.

15 Lukas, S., Philipp, A. M., & Koch, I. (2010). Switching attention between modalities:
16 Further evidence for visual dominance. *Psychological Research*, 74, 255-267.

17 Lukas, S., Philipp, A. M., & Koch, I. (2014). Crossmodal attention switching: Audito-
18 ry dominance in temporal discrimination tasks. *Acta Psychologica*, 153, 139-146.

19 Macaluso, E., Driver, J., Van Velzen, J., & Eimer, M. (2005). Influence of gaze direc-
20 tion on crossmodal modulation of visual ERPs by endogenous tactile spatial attention. *Cogni-*
21 *tive Brain Research*, 23, 406-417.

22 Macaluso, E., Frith, C. D., & Driver, J. (2002). Crossmodal spatial influences of touch
23 on extrastriate visual areas take current gaze direction into account. *Neuron*, 34, 647-658.

1 Mast, F., Frings, C., & Spence, C. (2015). Multisensory top-down sets: Evidence for
2 contingent crossmodal capture. *Attention, Perception, and Psychophysics*, 77, 1970-1985.

3 Mast, F., Frings, C., & Spence, C. (2017). Crossmodal attentional control sets between
4 vision and audition. *Acta Psychologica*, 178, 41-47.

5 McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264,
6 746-748.

7 Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of at-
8 tention in the audiovisual integration of speech. *Information Fusion*, 11, 4-11.

9 Parise, C. V., Spence, C., & Ernst, M. (2012). When correlation implies causation in
10 multisensory integration. *Current Biology*, 22, 46-49.

11 Pierson, J. M., Bradshaw, J. L., Meyer, T. F., Howard, M. J., & Bradshaw, J. A.
12 (1991). Direction of gaze during vibrotactile choice reaction time tasks. *Neuropsychologia*,
13 29, 925-928.

14 Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., &
15 Rauschecker, J. P. (2009). Multisensory integration of sounds and vibrotactile stimuli in pro-
16 cessing streams for “what” and “where”. *Journal of Neuroscience*, 29, 10950-10960.

17 Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multi-
18 sensory cues. *Psychonomic Bulletin & Review*, 15, 398-403.

19 Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention re-
20 gardless of perceptual load. *Journal of Experimental Psychology: Human, Perception and*
21 *Performance*, 33, 1311-1321.

22 Shore, D. I., Barnes, M. E., & Spence, C. (2006). Temporal aspects of the visuotactile
23 congruency effect. *Neuroscience Letters*, 392, 96-100.

1 Sinnett, S., Soto-Faraco, S., & Spence, C. (2008). The co-occurrence of multisensory
2 competition and facilitation. *Acta Psychologica*, 128, 153-161.

3 Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science & Tech-*
4 *nology*, 28, 61-70.

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

7 Spence, C. (2013). Just how important is spatial coincidence to multisensory integra-
8 tion? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296, 31-49.

9 Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The*
10 *Oxford handbook of perceptual organization* (pp. 649-664). Oxford, UK: Oxford University
11 Press.

12 Spence, C., & Ngo, M. K. (2012). Does attention or multisensory integration explain
13 the crossmodal facilitation of masked visual target identification? In B. E. Stein (Ed.), *The*
14 *new handbook of multisensory processing* (pp. 345-358). Cambridge, MA: MIT Press.

15 Spence, C., Parise, C., & Chen, Y.-C. (2011). The Colavita visual dominance effect. In
16 M. M. Murray & M. Wallace (Eds.), *Frontiers in the neural bases of multisensory processes*
17 (pp. 523-550). Boca Raton, FL: CRC Press.

18 Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory
19 cues: A review. *Hearing Research*, 258, 134-142.

20 Spence, C., & Squire, S. B. (2003). Multisensory integration: Maintaining the percep-
21 tion of synchrony. *Current Biology*, 13, R519-R521.

1 Stein, B. E., & Meredith, M. A. (1990). Multisensory integration. Neural and behav-
2 ioral solutions for dealing with stimuli from different sensory modalities. *Annals of the New*
3 *York Academy of Sciences*, 608, 51-65; discussion 65-70.

4 Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA:
5 MIT Press.

6 Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from
7 the perspective of the single neuron. *Nature Reviews Neuroscience*, 9, 255-267.

8 Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual
9 integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cor-*
10 *tex*, 17, 679-690.

11 Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifac-
12 eted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*,
13 14, 400-410.

14 Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integra-
15 tion: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neurosci-*
16 *ence*, 17, 1098-1114.

17 Tipper, S. P., Lloyd, D., Shorland, B., Dancer, C., Howard, L. A., & McGlone, F.
18 (1998). Vision influences tactile perception without proprioceptive orienting. *NeuroReport*, 9,
19 1741-1744.

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

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

1 Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip
2 and Pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental*
3 *Psychology: Human Perception and Performance*, 34, 1053-1065.

4 Van Ee, R., Van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory con-
5 gruency as a mechanism for attentional control over perceptual selection. *Journal of Neuro-*
6 *science*, 29, 11641-11649.

7 Vatakis, A., Maragos, P., Rodomagoulakis, I., & Spence, C. (2012). Assessing the ef-
8 fect of physical differences in the articulation of consonants and vowels on audiovisual tem-
9 poral perception. *Frontiers in Integrative Neuroscience*, 6:71.

10 Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex
11 speech, object-action, animal call, and musical stimuli. In M. J. Naumer & J. Kaiser (Eds.),
12 *Multisensory object perception in the primate brain* (pp. 95-121). New York, NY: Springer.

13 Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not
14 depend on the direction of automatic visual attention. *Perception & Psychophysics*, 63, 651-
15 659.

16 Wahn, B., Murali, S., Sinnett, S., & König, P. (2017). Auditory stimulus detection par-
17 tially depends on visuospatial attentional resources. *i-Perception*, 8:1.

18 Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory tem-
19 poral binding window and its dysregulation in developmental disabilities. *Neuropsychologia*,
20 64, 105-123.

21 Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spa-
22 tial attention orienting. *Perception & Psychophysics*, 62, 1258-1264.

1 Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory
2 discrepancy. *Psychological Bulletin*, 88, 638-667.

3 Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sen-
4 sory and association areas to audiovisual integration in object categorization. *Journal of Neu-*
5 *roscience*, 30, 2662-2675.

6 Wesslein, A., Spence, C., & Frings, C. (2014). Vision affects tactile target and distrac-
7 tor processing even when space is task-irrelevant. *Frontiers in Psychology*, 5:84.

8 Wesslein, A., Spence, C., & Frings, C. (2015). You can't ignore what you can't sepa-
9 rate: The effect of visually induced target-distractor separation on tactile selection. *Psycho-*
10 *nomic Bulletin and Review*, 22, 728-736.

11 Zimmer, U. & Macaluso, E. (2007). Processing of multisensory spatial congruency
12 can be dissociated from working memory and visuo-spatial attention. *European Journal of*
13 *Neuroscience*, 26, 1681-1691.

14 Zmigrod, S., & Hommel, B. (2011). The relationship between feature binding and
15 consciousness: Evidence from asynchronous multi-modal stimuli. *Consciousness and Cogni-*
16 *tion*, 20, 586-593.

17 Zmigrod, S., Spapé, M., & Hommel, B. (2009). Intermodal event files: Integrating fea-
18 tures across vision, audition, taction, and action. *Psychological Research*, 73, 674-684.

Appendix

Table 3. Mean RT (in milliseconds; error rates in % in parentheses) as a function of visual and auditory distractor feature congruency in Experiment 1 after excluding the first trial of each block and all trials following a catch trial.

Visual distractor feature					
Auditory	distractor	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	<i>Mean</i>
feature					
<i>Congruent</i>		687 (0.8)	709 (0.9)	733 (1.0)	710 (0.9)
<i>Neutral</i>		795 (7.0)	779 (3.4)	806 (4.4)	793 (4.9)
<i>Incongruent</i>		771 (4.6)	802 (5.7)	817 (7.3)	797 (5.9)
Mean		751 (4.1)	764 (3.4)	785 (4.2)	

Table 4. Mean RT (in milliseconds; error rates in % in parentheses) as a function of visual and auditory distractor feature congruency in Experiment 2 after excluding the first trial of each block and all trials following a catch trial.

Visual distractor feature					
Auditory	distractor	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	<i>Mean</i>
feature					
<i>Congruent</i>		714 (2.2)	757 (2.1)	796 (5.7)	755 (3.3)
<i>Neutral</i>		854 (19.3)	799 (5.2)	824 (4.5)	826 (9.7)
<i>Incongruent</i>		802 (6.8)	853 (15.4)	864 (15.4)	840 (12.5)
Mean		790 (9.4)	803 (7.6)	828 (8.5)	

Benefit-cost-analyses

Note the neutral condition typically falls in-between the congruent and incongruent conditions – yet, rather unpredictably. One reason for this unpredictability is that the neutral features do have a clearly different frequency in the experiment and are probably treated differently than the features with response association. As a result, we think it is somewhat unreliable to compute differences scores and interpret them as benefits versus costs (see also Jonides & Mack, 1984). We nevertheless thought it prudent to include a neutral condition since it may help signal to the participants that the distractor really is irrelevant (neither supporting nor interfering with performance).

For the sake of completeness, however, we also analyzed our data in terms of the benefit (facilitation) and costs (interference) of each feature as a function of the congruency of the other feature. We took the difference between the neutral minus the congruent condition on one congruency level as the benefit, and the difference between the incongruent minus the neutral condition as the cost. For instance, the benefit of the auditory feature for a congruent visual feature was computed as the difference between the visual-congruent-auditory-neutral condition minus the visual-congruent-auditory-congruent condition. The costs of the visual feature for a neutral auditory feature was computed as the difference between the visual-incongruent-auditory-neutral condition minus the visual-neutral-auditory-neutral condition. The differences so-computed in both experiments can be found in Table A1.

- 1 Table A1. Benefit and costs (RT) of the auditory distractor feature as a function of the
- 2 congruency of the visual feature in Experiments 1 and 2 (standard error in parentheses).

<i>Experiment 1</i>		Visual feature		
Auditory feature	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	Mean
<i>Benefit</i>	78 (10.8)	64 (8.8)	75 (9.4)	73 (7.6)
<i>Costs</i>	22 (10.6)	19 (7.2)	5 (9.7)	15 (6.5)
<i>Experiment 2</i>				
<i>Benefit</i>	79 (13.2)	34 (9.3)	36 (9.5)	50 (8.9)
<i>Costs</i>	52 (12.9)	47 (11.4)	23 (9.5)	40 (9.0)
<i>Experiment 1</i>				
Visual feature	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	Mean
<i>Benefit</i>	20 (8.3)	6 (8.8)	3 (8.8)	10 (5.4)
<i>Costs</i>	21 (7.3)	32 (6.8)	19 (8.3)	24 (4.1)
<i>Experiment 2</i>				
<i>Benefit</i>	47 (6.4)	2 (9.7)	-3 (11.3)	15 (5.8)
<i>Costs</i>	29 (8.6)	30 (7.1)	6 (8.9)	22 (5.0)

Auditory Feature. To compare benefits and costs for the auditory feature, we conducted a 3 (Visual Feature Congruency: Congruent vs. Neutral vs. Incongruent) $\times$ 2 (Difference: Benefit vs. Costs) MANOVA with Pillai's trace as the criterion and RT as the dependent variable. In Experiment 1, there was a main effect of Difference, $F(1, 27) = 35.17, p < .001, \eta_p^2 = .57$, with overall benefit being higher than overall costs. None of the other effects were significant, all $ps > .202$. In Experiment 2, there was a main effect of Visual Feature Congruency, $F(2, 26) = 22.93, p < .001, \eta_p^2 = .64$, revealing that both benefits and costs were higher when the visual feature was congruent to the visual target feature and declined when it was incongruent. None of the other effects were significant, all $ps > .219$. Note that this main effect of Visual Feature Congruency fits to our main finding, namely the interaction of auditory and visual features.

Visual Feature. To compare benefit and costs for the visual feature, we conducted a 3 (Auditory Feature Congruency: Congruent vs. Neutral vs. Incongruent) $\times$ 2 (Difference: Benefit vs. Costs) MANOVA with Pillai's trace as the criterion and RT as the dependent variable. In Experiment 1, there was a main effect of Auditory Feature Congruency, $F(2, 26) = 23.44, p < .001, \eta_p^2 = .64$, showing that both benefit and costs were higher when the auditory feature was congruent to the auditory target feature and declined when it was incongruent. None of the other effects were significant, all $ps > .085$. In Experiment 2, there were no significant effects, all $ps > .075$.