Running head: Spatial Attention in Multisensory Distractor Processing

**Multisensory distractor processing
is modulated by spatial attention**

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

¹ Department of Psychology, University of Trier, Germany

² Department of Experimental Psychology, University of Oxford, United Kingdom

Word count : 7447 words (without figure captions & tables, references
and appendix)

Address for correspondence : Simon Merz
Department of Psychology, Cognitive Psychology
Universitätsring 15
54286 Trier
Germany
Email : merzs@uni-trier.de
Phone : +49 651 / 201 2907

Abstract

In order to respond to multisensory information, inputs from different sensory modalities must be processed and combined. Recently, overt spatial attention was shown to be a crucial factor modulating the processing of irrelevant audiovisual multisensory stimuli. Here we investigate the processing of task-irrelevant visuotactile features in a multisensory flanker interference task incorporating visuotactile target and distractor stimuli. The congruency between the target and distractor features was varied orthogonally. Across three experiments, overt spatial attention as well as the spatial separation between the distractor and the target were varied systematically. When fixating the distractor, the processing of the visual and tactile distractor features was not independent. Manipulating overt spatial attention as well as the spatial separation between the target and distractor impacted multisensory distractor processing. These results are consistent with those approaches emphasizing the role of attention in multisensory processing specifically in relation to the cognitive load or selection difficulty of the task situation.

Keywords: Multisensory selection, spatial attention, distractor processing

1

2

Statement of public significance

3

This study provides insight into the modulation of multisensory distractor processing by

4

spatial attention. In order to elicit multisensory processing, the visual and tactile feature of the

5

target as well as the distractor were presented in close spatiotemporal proximity. Our results

6

clearly demonstrate that irrelevant multisensory stimuli are combined and not processed

7

independently if overt spatial attentional resources are directed toward them.

Introduction

The world in which we live is fundamentally multisensory. In every waking moment, signals impinging on our senses may be seen, felt, heard, smelled, or even tasted. But which are the important preconditions such that those unisensory signals are combined rather than being treated as independent? Such multisensory processing has primarily been investigated by focusing on the processing of multisensory targets, i.e., multisensory information that is both attended and responded to. To combine and integrate the multisensory information, spatial as well as temporal co-occurrence have been suggested to constitute important preconditions for multisensory integration (e.g., Spence, 2007; Stein & Meredith, 1990, 1993; Stein & Stanford, 2008; Zmigrod & Hommel, 2011; for a discussion concerning temporal co-occurrence, see Harrar, Harris, & Spence, 2017; Vatakis & Spence, 2010; and for discussion of the many conditions under which spatial co-occurrence does not matter, see Spence, 2013).

Attention's role in multisensory target processing is still a hotly-debated topic (see Jensen, Merz, Spence, & Frings, 2019, for discussion). On the one hand, multisensory processing often appears to be automatic (e.g., Bertelson, Vroomen, De Gelder, & Driver, 2000; Santangelo & Spence, 2007). On the other, attention sometimes appears to be an important factor modulating multisensory processing (e.g., Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007). This seemingly conflicting evidence has been addressed by recent frameworks emphasizing selection difficulty, or cognitive load and salience, as key factors determining the influence of attention for multisensory processing (e.g., De Meo, Murray, Clarke, & Matusz, 2015; Navarra, Alsius, Soto-Faraco, & Spence, 2010; Talsma, 2015; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010; Tang, Wu, & Shen, 2016). But what about multisensory distractors, that is, perceptual information which is in no way relevant, or might even interfere, with current goals?

In a recent study, we reported evidence suggesting that overt spatial attention influences distractor processing in a multisensory flanker task with audiovisual stimuli (Jensen et al., 2019). Specifically, combinations of visual and auditory features (varying in terms of their color and pitch) were used as stimuli, and were presented on/via two loudspeakers. The participants were tasked with responding to specific combinations of audiovisual features presented from one loudspeaker, i.e., the multisensory target, while audiovisual stimuli from the other loudspeaker (35° spatial separation) had to be ignored. The audiovisual stimuli which had to be ignored were defined as the multisensory distractors. The general pattern of results observed was that reaction times (RT) were fastest (slowest) when both features of the distractor matched (did not match) both features of the target, the congruent-congruent (incongruent-incongruent) condition. In those conditions with partial feature overlap, where one distractor feature matched the target feature while the other distractor feature did not (i.e., the congruent-incongruent conditions), RTs were somewhere in-between these two extremes. Critically, the participants' overt spatial attention was manipulated by varying whether the target or the distractor was presented at fixation (with the other stimulus being presented laterally). This experimental manipulation influenced the processing of the distractor features.

In our view, distractor processing describes the complete processing of the distractor stimulus, from early perceptual stages through to late response selection stages, as in Pashler's (1994, 1998) account of the three stages of processing (1: perceptual processing; 2: central processing, e.g. response selection; 3: response execution). Hereby, the features of the distractor can be processed independently, that is, the two distractor features are processed separately and were never combined throughout these different processing stages. In contrast, multisensory distractor processing indicates that the two features were combined and interacted somewhere along these stages of information processing. In our audiovisual study (Jensen et al., 2019), the distractor features were combined and not processed independently when the distractor was

presented at fixation. More precisely, the size of the congruency effect (that is, the difference between congruent and incongruent trials) of one feature was influenced by the congruency of the other feature. In other words, a significant interaction was evidenced. However, no interaction was evidenced when the target was presented at fixation. Under the latter conditions, the visual and auditory features of the multisensory distractor were processed independently.

These findings help to provide some initial insights into the processing of multisensory distractors and suggest that multisensory processing is modulated by overt spatial attention. It should, however, be noted that Jensen et al.'s (2019) study focused solely on audiovisual stimulus combinations. The question therefore arises as to whether these findings can be generalized to other modality combinations. The literature from different fields, such as cognitive psychology, perception, and attention research suggest that a simple generalization across the different modalities and, more importantly for present purposes, modality combinations can by no means be assured. There appears to be a general imbalance between the senses, with a dominance of one modality over the others often being reported (see Posner, Nissen, & Klein, 1976; Sinnett, Spence, & Soto-Faraco, 2007; Spence, 2009, for reviews and discussions of the Colavita visual dominance effect; see Shams, Kamitani, & Shimojo, 2000, for the double flash illusion), which might be explained by differences in uncertainty / variance between the senses (Bayesian decision theory, Ernst & Banks, 2002; Ernst & Bühlhoff, 2004). Further, modality-specific processing differences have been evidenced, for example, in the change blindness / detection paradigm (e.g. Auvray, Gallace, Tan, & Spence, 2007; Gallace, Tan, & Spence, 2006; Gregg & Samuel, 2008; Simons & Rensink, 2005). Moreover, different modality pairings (such as audiovisual, visuotactile, and audiotactile) have been shown to elicit somewhat distinct patterns of crossmodal temporal adaptation and recalibration (e.g., similar long-term adaption to temporal shifts, but very different short-term inter-trial adaptation effects; Alais, Ho, Han, & Van der Burg, 2017; see also Van der Burg, Orchard-Mills, & Alais, 2014).

Furthermore, dual-task performance has been shown to be significantly modulated by the input / output modality pairing (Hazeltine, Ruthruff, & Remington, 2006).

In the negative priming task, a selection task closely related to the flanker task used in this study, modality comparisons have revealed some interesting findings. When selection is based on target identity, recent research has demonstrated that similar mechanisms underlie the effect in the visual, auditory and tactile modality (Frings, Bader, & Spence, 2008; Mayr & Buchner, 2007; though effect sizes are noticeably larger in the tactile modality, Frings, Amendt, & Spence, 2011; see Frings, Schneider, & Fox, 2015, for a review). In contrast, changing the criterion for target selection from a target's identity to its spatial location results in different mechanisms accounting for the visual (response inhibition; Buckolz, Goldfarb, & Khan, 2004; Fitzgeorge, Buckolz, & Khan, 2011), auditory (feature mismatch; Möller, Mayr, & Buchner, 2013), and tactile (location inhibition; Frings, Mast, & Spence, 2014; Wesslein, Spence, Mast, & Frings, 2015) negative priming effects. This broad variety of examples from different cognitive and attentional paradigms reveals that a generalization of previous results for different modalities cannot be assured, nor can it be assumed, *a priori*.

In everyday life, there is a strong connection between vision and touch (Shopland & Gregory, 1964), with visual input often helping us to identify and specify the exact location from which a tactile stimulus was felt (see Gallace & Spence, 2014, for a review). This strong interconnection and interplay between these two spatial senses has generated a reasonably large body of empirical research investigating the modalities separately and their crossmodal influences in various tasks (e.g., Driver & Grossenbacher, 1996; Honoré, Bourdeaud'hui, & Sparrow, 1989; Soto-Faraco, Ronald, & Spence, 2004; Spence, Lloyd, McGlone, Nicholls, & Driver, 2000; Spence, Pavani, & Driver, 2004; Spence, Pavani, Maravita, & Holmes, 2008; Zmigrod, Spapé, & Hommel, 2009; see Wesslein, Spence, & Frings, 2014, for a review). For instance, Driver and Grossenbacher investigated the influence of overt spatial attention (i.e.,

the orientation of a participant's head and gaze) in a tactile flanker interference task. When facing the relevant hand (i.e., the hand from which the response relevant target stimuli were presented), increasing the distance to the irrelevant hand (i.e., the hand at which the distractor stimuli was presented) led to a decrease in distractor interference. By contrast, increasing the distance between the hands while focusing on the irrelevant hand had no influence on tactile distractor processing. This result therefore underlines the importance of overt spatial attention to the processing of tactile distractors.

The present study

In the present study, a variant of Eriksen and Eriksen's (1974) classic flanker task was used in order to analyze distractor processing for the visuotactile stimulus combination. For this study, participants held a multisensory cube in each hand. The participants' task was to respond to a specific combination of visual (different color) and tactile (different amplitude) target features (e.g., a red light combined with a weak vibration), presented from one location. In order to ensure that the participants responded to the specific combination of the presented target, catch trials with reversed feature assignments were also presented. Furthermore, the multisensory distractor was also composed of a visual and tactile feature, which were varied orthogonally to each other and also with respect to the target features. This is comparable to our previous study (Jensen et al., 2019), only that instead of tactile target feature (used here), an auditory target feature (different pitch) was used there. Importantly, the direction of participant's gaze as well as the spatial separation between the distractor and target were manipulated in order to investigate the role of spatial attention on multisensory distractor processing.

Based on our previous findings (Jensen et al., 2019), as well as the evidence concerning the influence of overt spatial attention in studies of tactile distractor processing (Driver &

Grossenbacher, 1996; Soto-Faraco et al., 2004), we expected to find a modulation of multisensory distractor processing following the manipulation of participants' spatial attention. More precisely, we expected to observe the fastest (slowest) RTs for the congruent-congruent (incongruent-incongruent) conditions. For the partial feature overlap conditions, we expected RTs to lie somewhere in-between the other two conditions. When the participants fixated the distractor, it was expected that the distractor features would be combined, and hence not processed independently. That is, the size of the congruency effect (the difference between faster congruent and slower incongruent trials) of one feature should be influenced by the congruency of the other feature. On a statistical level, a significant ordinal interaction between visual and tactile feature congruency was predicted. This interaction was expected to diminish, if not dissipate entirely, when the distractor is oriented away from the focus of participants' overt spatial attention (i.e., gaze) while fixating the target.

To foreshadow the results, multisensory processing is evidenced when the distractor is presented in the center of the participant's gaze (Experiment 1). Moving the distractor away from the center of the participant's gaze while presenting the target there led to a significant decrease in multisensory processing toward unisensory processing (Experiments 2 and 3). Spatial attention impacted multisensory distractor processing via the combined manipulation of overt spatial orientation (see also Jensen et al., 2019) and the spatial separation between the target and distractor.

Experiment 1

Method

Participants

Flanker effects in the literature and from our own laboratories have typically been found to have 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 27 participants at a minimum (power analyses were run with G-power 3.1.7; Erdfelder, Faul, & Buchner, 1996). Thirty-two students from the University of Trier took part in the study in return for course credit¹. One participant was excluded due to extremely high error rates (ER) in the catch trials, possibly indicating a lack of understanding of the task (participant: 53.8% errors; sample mean: 13.8% errors, sample SD = 12.7% errors). A total of 31 participants (5 male) aged from 18 to 29 years (mean = 21.4, SD = 2.5) were used for analyses. All of the participants had normal or corrected-to-normal vision and no sensory impairment by self-report.

Apparatus and materials

The participants were seated in a dark soundproofed chamber in front of a 24" CRT computer screen with a refresh rate of 60 Hz. The distance to the screen (63 cm) was kept constant with the aid of a chin and forehead rest. Instructions were presented in grey letters on a black background in the middle of the screen. Combinations of three different vibration intensities (strong – about 200 μm peak-to-peak amplitude; medium – about 56 μm peak-to-peak amplitude; weak – about 15 μm peak-to-peak amplitude, frequency of about 270 Hz) and three different colors, namely Red (HSV: 0, 100, 12.55), Green (HSV: 120, 100, 12.55), and

¹ Data from thirty-two participants were collected to allow for possible exclusions, but furthermore, the number of participants had to be a multiple of four to balance our between-participants factors, Target Block Order (left vs. right first) and Target Feature Mapping (strong vibration mapped with red light and weak vibration mapped with blue light vs. strong vibration mapped with blue light and weak vibration mapped with red light).

Blue (HSV: 240, 100, 12.55) were used as stimuli. Further, an ‘eraser’ stimulus consisting of a white flash (HSV: 0, 0, 12.55) and five brief vibrations (about 128 μ m peak-to-peak amplitude) was used to prevent after-effects from the previous trial on the current trial. The stimuli were presented over two custom-made cubes with two LEDs and two tactors each (Model C-2, Engineering Acoustic, Inc.; controlled via the serial interface; for a detailed description of the cubes, see the Appendix). On each cube, the upper LED as well as the upper tactor were used. Note that they were located very close to each other on each cube. The cubes were held in the participant’s hands. One cube was held centrally in front of them just below the screen, and participants’ gaze was directed toward this central cube. The other one was held laterally, 35° on the right or left side (identical to Jensen et al., 2019). On the desk, three lines were prepared to assist the experimenter in positioning the participant’s hands at the correct visual angle. Therefore, the actual spatial separation between the distractor and target varied between participants. For example, at a holding distance of 40 cm (30 cm), this corresponds to a spatial separation of about 25 cm (19 cm) between the target and distractor. The participants were instructed to fixate the central cube throughout the course of the experiment. Responses were given using footpedals using Chronos®. The relevant distances and the experimental set-up are shown in Figure 1. The experiment was conducted with the E-Prime software (Version 2.0) and the data were analyzed with IBM SPSS Statistics (Version 23).

Figure 1: Experimental set-up and the results of Experiments 1-3. The birds-eye view depicts the orientation of the head, arm position and subsequent location of the two cubes used to present the target (T) and distractor (D) stimuli (after half of the experimental trials, the arm position were changed; for more details, see the text): For the results figures, error bars represent standard errors. Note that the line diagrams depicts the mean RTs of the 2 (Visual Congruency) x 2 (Tactile Congruency) MANOVA while the bar diagram depicts the interaction term of this MANOVA.

Procedure

The participants were tested individually in a sound-proof and dimly-lit chamber. They sat 63 cm from the display with their head resting on a chin-and-forehead rest and were

1 instructed to fixate the cube held in their direction of gaze throughout the experiment (for a
2 schematic illustration of the experimental set-up, see Figure 1). The participant's feet were
3 placed on footpedals. Their task was to identify the target via footpedal responses. The feature
4 mapping was balanced across participants. That is, two multisensory target stimuli were used,
5 a red (blue) light combined with a strong vibration, which required the depression of the left
6 footpedal, and a blue (red) light combined with a weak vibration, which required the depression
7 of the right footpedal. The participants were instructed to respond to these target stimuli
8 presented from the lateral cube as rapidly and accurately as possible, while they were also
9 instructed to ignore the stimuli on the central cube (i.e., the distractors). To assure that the
10 participants responded to a combination of visual and tactile features, and not just to an
11 individual feature, catch trials with a reversed target feature assignment (i.e., red (blue) light
12 and weak vibration or blue (red) light and strong vibration) were also presented. The
13 participants were instructed not to respond to these stimuli thus making the task a go-/no-go-
14 task. These catch trials accounted for 20% of all trials. To boost participants' motivation to
15 react as rapidly and as accurately as possible, the participants were told that all trials on which
16 they reacted erroneously, or else too slowly, would have to be redone throughout the
17 experiment. In fact, only catch trials to which the participants erroneously responded were
18 repeated after a quarter, half, three quarters, or at the end of each block. Therefore, the total
19 number of trials conducted by each participant varied (due to different numbers of catch trials)
20 while the number of target trials remained equal for all participants.

21 Distractor stimuli were computed as combinations of three visual and three tactile
22 features. A green light and a medium vibration served as neutral features that were not mapped
23 on to a response. Additionally, there were trials without a distractor to emphasize the irrelevance
24 of the distractor to the participants' task (comparable to our previous audio-visual study, Jensen
25 et al., 2019). Therefore, ten different distractor conditions were incorporated with an equal

number of trials (32 trials throughout the experimental phase). The visual and tactile features of the distractors could either be congruent, neutral, or incongruent with respect to the corresponding feature of the current target and were varied orthogonally to each other. The side of the target cube (left, right) was manipulated within-participants in two blocks of 200 trials each (plus the incorrect catch trials) while the order (left first, right first) was balanced across participants. Additional breaks were offered after a quarter, half, and three-quarters of the target trials.

The exact stimulus sequence of each trial was as follows: A cross was presented at the bottom of the screen just above the central, focused cube indicating the start of the trial, followed by the onset of the distractor after 250 milliseconds (ms). The target was presented 100 ms after the onset of the distractor. Both stimuli then lasted for 1,000 ms and participants had to respond within 1,500 ms of distractor onset. Feedback was only given for false or missing responses and was displayed for 500 ms. After an additional 500 ms blank interval, an eraser stimulus (duration: 500 ms), consisting of a white flash (duration: 500 ms) and five brief 50 ms vibrations (each vibration was separated by a 50 ms blank interstimulus-interval) was presented from both cubes in order to reduce any aftereffects from the previous trial on the present trial. The eraser stimulus was followed by a Response-Stimulus-Interval (RSI) of 500 ms before the start of the next trial.

Instructions for the experiment were provided on the screen. The experiment started with a learning phase of 8 trials in which the stimulus-response mapping was learned. After that, a first practice phase was run with 16 trials in which target and catch trials both accounted for 50% of all trials and were presented on both cubes. The second practice phase with 50 trials was exactly the same as the experimental phase except for the fact that no erroneous responses in the catch trials were repeated. During the learning and practice phases, feedback about the response (correct, false or missing response) was presented at the bottom of the screen, in close

proximity to the focused cube (in this experiment, the distractor). In the experimental phase, only feedback for false or missing responses was presented. Further, the experimenter was in the experimental room to ensure the correct positioning of the cube. The experiment lasted for about 50 minutes.

Design & Analysis

The experiment consisted of a 3×3 repeated measures design with the variables *Visual Distractor Feature Congruency* (congruent, neutral, or incongruent) and *Tactile Distractor Feature Congruency* (congruent, neutral, or incongruent). A 3×3 design is, however, somewhat hard to interpret as both factors have three levels (see Jonides & Mack, 1984, for a review on the difficulty of designing/interpreting the data from the neutral condition). These trials were still incorporated into the task in order to emphasize the irrelevance of the distractor and to make the task comparable to our previous audio-visual study (Jensen et al., 2019). A 2×2 multivariate analysis of variance (MANOVA)² without the neutral condition is reported in the main text, but the full model as well as control analysis, accounting for Distractor Side, Distractor Block Order, and Target Feature Mapping is reported in Appendix B. RTs as well as ER were used as dependent variables. The raw data is available for open access (Merz, 2019).

Results

We first checked whether participants performed the task appropriately by examining the catch trials. On average, participants responded correctly on 86.25% of these trials (SD = 12.71%) showing that they successfully selected multisensory targets. For the analysis of the

² Note that all repeated-measures designs are inherently multivariate and the MANOVA has the advantage that sphericity cannot influence the results (e.g., see Tabachnick, Fidell, & Ullman, 2007).

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, 12.45% of the trials were excluded from the analysis due to these restrictions (see Table 1 for the mean RT and ER data).

Table 1: Mean reaction times (in ms) and error rates (with standard errors in brackets) of Experiments 1-3 for each Visual (congruent vs. incongruent) and Tactile (congruent vs. incongruent) Distractor Feature Congruency combination.

Visual distractor feature congruency	Response times (RTs)		Error rates (ERs)	
	Tactile distractor feature congruency Congruent	Tactile distractor feature congruency Incongruent	Tactile distractor feature congruency Congruent	Tactile distractor feature congruency Incongruent
Experiment 1				
Congruent	737 (16.8)	781 (16.3)	4.54 (0.90)	8.77 (2.01)
Incongruent	782 (17.0)	799 (17.1)	3.83 (1.02)	7.56 (1.95)
Experiment 2				
Congruent	732 (21.4)	765 (20.8)	4.39 (0.80)	7.23 (1.61)
Incongruent	759 (18.7)	781 (18.8)	3.52 (0.65)	5.96 (1.11)
Experiment 3				
Congruent	778 (14.9)	798 (16.0)	4.31 (0.91)	10.13 (2.01)
Incongruent	788 (15.5)	807 (16.1)	4.53 (0.94)	8.73 (1.94)

RT

A 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) $\times$ 2 (Tactile Distractor Feature Congruency: congruent vs. incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(1, 30) = 36.94$ $p < .001$, $\eta_p^2 = .55$, with participants responding fastest when the visual target and distractor feature were congruent (759 ms) as compared to when they were incongruent (791 ms). The main effect of Tactile Distractor Feature Congruency was significant, $F(1, 30) = 46.42$, $p < .001$, $\eta_p^2 = .61$.

That is, participants responded more rapidly when the vibrotactile target and distractor features were congruent (760 ms) than when they were incongruent (790 ms). Critically, there was also a significant interaction between these two variables, $F(1, 30) = 12.93, p = .001, \eta_p^2 = .30$ (see Figure 1), suggesting multisensory distractor processing. More precisely, the mean RTs in the partial feature overlap conditions (tactile congruent – visual incongruent and tactile incongruent – visual congruent) were slower than the RTs for the congruent-congruent condition, $t(30) = 8.08, p < .001$, but also faster than the RTs for the incongruent - incongruent condition, $t(30) = -3.51, p = .001$, indicating an ordinal interaction.

ER

For the ER, the same 2×2 MANOVA was conducted as for the RT data. There was a main effect of Tactile Distractor Feature Congruency, $F(1, 30) = 5.72, p = .023, \eta_p^2 = .16$, with fewer errors in the congruent (4.19%) than in the incongruent condition (8.17%). Neither the main effect of Visual Distractor Feature Congruency, $F(1, 30) = 1.61, p = .215$, nor the interaction, $F(1, 30) = 0.22, p = .646$, were significant.

Discussion

In Experiment 1, multisensory targets had to be selected against a background of multisensory distractors. A multisensory variant of the flanker interference task, in which the stimuli had a visual and a tactile feature, was used. Participants had to respond to specific combinations of visuotactile features. The distractor features were varied orthogonally to each other and to the target features. In Experiment 1, the centrally-presented distractor was fixated while the target was presented laterally. Congruency effects were documented for each of the two distractor features individually, thus showing that the distractor interfered with target processing. The participants responded more rapidly and made fewer errors when a feature was

1 congruent than when it was incongruent. Critically, an ordinal interaction was evidenced in the
2 RTs, indicating that the distractor features were combined and not processed independently.
3 Critically, we evidenced the fastest RTs for the congruent-congruent trials, the slowest RTs for
4 the incongruent-incongruent trials, and mean RTs in the partial feature overlap conditions that
5 were somewhere in-between, but still significantly different from these two extreme conditions.
6 This pattern of results is important because it excludes an alternative explanation for our data
7 which might explain an interaction without any multisensory processing. This alternative
8 account would assume that for the congruent-congruent condition, an additional step of
9 localizing the target against the distractor does not have to be performed, given that the
10 distractor and target are perceptually identical in this condition³. This account would not assume
11 any difference between the incongruent-incongruent and the partial feature overlap conditions.
12 As these conditions differed significantly, the alternative of simple discrimination facilitation
13 in the congruent-congruent condition cannot explain the present data set.

14 In fact, the observed pattern of data was similar to the audiovisual data (Jensen et al.,
15 2019), as the difference between congruent-congruent trials and incongruent-congruent trials
16 was larger than the difference between congruent-congruent trials and incongruent-congruent
17 trials. This suggests multisensory processing and stands in line with our previous findings in
18 the case of the audiovisual modality combination (Jensen et al., 2019, Experiment 2). In order
19 to investigate whether overt spatial orientation was critical for multisensory processing to occur,
20 the target was presented at the focus of participants' gaze in Experiment 2. Note that the distance
21 between the target and the distractor was kept at 35°, ensuring an appropriate comparison with
22 Experiment 1.

³ We thank an anonymous reviewer for bringing this alternative explanation to our attention.

Experiment 2

Method

Participants

A total of 32 students (6 male), aged from 18 to 24 years (mean = 20.3, SD = 1.5) from the University of Trier took part in the study in return of course credit. All of the participants had normal or corrected-to-normal vision and no sensory impairment by self-report.

Apparatus, materials, design, and procedure

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

Results

We first checked again whether participants performed the task appropriately by examining the catch trials. On average, participants responded correctly in 84.22% of these trials (SD = 12.91%) showing that they successfully selected multisensory targets. For the analysis of the 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, 8.8% of the trials were excluded from the analysis due to these restrictions.

RT

A 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) $\times$ 2 (Tactile Distractor Feature Congruency: congruent vs. incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(1, 31) = 24.04, p < .001, \eta_p^2 = .44$; that is, participants responded more rapidly when the visual target and distractor features were congruent (749 ms) than when they were incongruent (770 ms). There was a significant main effect of Tactile Distractor Feature Congruency, $F(1, 31) = 34.75, p < .001, \eta_p^2 = .53$, with participants responding more rapidly when the tactile target and distractor features were congruent (746 ms) than when they were incongruent (773 ms). Critically, there was no interaction between these two variables, $F(1, 31) = 1.45, p = .238$ (see Figure 1), thus suggesting independent, unisensory processing of the distractors.

ER

For the ER, we conducted the same 2×2 MANOVA as for the RT data. There was a main effect of Tactile Distractor Feature Congruency, $F(1, 31) = 6.71, p = .014, \eta_p^2 = .18$, but no main effect of Visual Distractor Feature Congruency, $F(1, 31) = 1.56, p = .221$, nor any interaction, $F(1, 31) = 0.12, p = .730$.

Between-Experiment Analysis

A 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) $\times$ 2 (Tactile Distractor Feature Congruency: congruent vs. incongruent) $\times$ 2 (Experiment: 1 vs. 2) MANOVA with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed significant main effects of Visual, $F(1, 61) = 61.16, p < .001, \eta_p^2 = .50$, as well as Tactile Distractor Feature Congruency, $F(1, 61) = 80.45, p < .001, \eta_p^2 =$

.57, neither of which were further modulated by Experiment, $F(1, 61) < 1.94$, $p > .168$. The interaction between Visual and Tactile Distractor Feature Congruency was significant, $F(1, 61) = 10.85$, $p = .002$, $\eta_p^2 = .15$. However, this interaction was not further modulated by Experiment, $F(1, 61) = 2.27$, $p = .137$, thus suggesting no statistical difference in distractor processing between the two experiments.

Discussion

In Experiment 2, the multisensory target was fixated, and the distractor cube was not fixated. Once again, we found the two main effects of congruency for the tactile and visual distractor features in the RT data and a vibrotactile congruency effect in the error data. In contrast to Experiment 1, though, the interaction between the two distractor features in the RT data was not significant. In Experiment 2, the processing of one feature did not seem to interact significantly with the processing of the other. However, the comparison of the results of Experiments 1 and 2 failed to reveal a significant modulation of the interaction between experiments. Across the two experiments, an overall significant interaction between the congruencies of the two distractor features was evidenced. This result indicates that the simple change of overt spatial attention from the distractor toward the target was not quite enough to induce the more independent, unisensory information processing in the visuotactile modality combination. This stands in contrast to the results of our previous audiovisual study, in which this manipulation was strong enough to induce a significant difference in distractor processing.

In Experiment 3, the spatial separation between the distractor and the target, which was still presented at the center of participants' gaze, was doubled. Looking at unisensory tactile distractor processing, increasing the spatial separation between target and distractor while focusing on the target diminished unisensory processing (Driver & Grossenbacher, 1996).

Similarly, spatial separation has been shown to influence unisensory visual distractor processing (e.g., Eriksen & Eriksen, 1974). In particular, increasing the spatial distance between the distractor and target leads to decreased distractor processing (e.g., Eriksen & Eriksen, 1974; Fox, 1998; Miller, 1991), indicating similar influences of spatial separation on unisensory visual and tactile distractor information processing. In Experiment 3, we investigate whether this modulation translates to the case of multisensory distractor processing.

Experiment 3

Method

Participants

Thirty-two students from the University of Trier took part in the study in return for course credit. One participant was excluded due to extremely high ER in the catch trials indicating a lack of understanding of the task (participant: 88.8% errors; sample mean: 14.8% errors, SD = 10.6% errors). One participant was excluded from the analysis due to extremely high ER (participant: 66.9% errors; sample mean: 6.6% errors, SD = 5.5% errors). One participant was excluded because RTs were shorter than 200 ms for all trials in the second half of the experiment, and so RTs could not be meaningfully assessed. A total of 29 participants (10 male) aged from 18 to 38 years (mean = 23.5, SD = 4.1) were used for analyses. All of the participants had normal or corrected-to-normal vision and no sensory impairment by self-report.

Apparatus, materials, design, and procedure

The apparatus, materials procedure, task, and trial sequence were exactly the same as in Experiment 2 with the sole exception that the distance between the fixed target and the

distractor was doubled (70°). Further, the lateral distractor cube was placed on the desk and only touched by the participant at the upper tactor location with the corresponding index finger (see Figure 1)⁴. Please note, the screen was only used to present the instructions as well as the cross to indicate a new trial and the feedback during the learning and practice phase, the latter two were presented on the bottom left / right corner of the screen, depending on the direction of focus.

Results

Once again, we first checked whether the participants had performed the task appropriately by examining the catch trials. On average, participants responded correctly in 85.22% of these trials (SD = 10.63%) showing that they successfully selected multisensory targets.

For the analysis of the RT data, only those trials in which the participants responded correctly to the target were considered. All of those 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, 11.69% of the trials were excluded from the analysis due to these restrictions.

RT

A 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) × 2 (Tactile Distractor Feature Congruency: congruent vs. incongruent) MANOVA with Pillai's trace as the

⁴ This change did not alter the perception of the stimuli themselves (colored LED / vibrations of the tactors), as the tactile vibration was still presented to the left / right index finger (please note that the signal was only presented at that location and that our multisensory cube features strong damping qualities – see the Appendix) and also the visibility of the distractor LED was unchanged.

criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed a significant main effect of Tactile Distractor Feature Congruency, $F(1, 28) = 13.74, p = .001, \eta_p^2 = .33$; that is, participants responded more rapidly when the tactile target and distractor features were congruent (783 ms) than when they were incongruent (803 ms). Descriptively, participants responded more rapidly in the visual congruent trials (788 ms) than in the visual incongruent trials (798 ms), but the main effect of Visual Distractor Feature Congruency was not significant, $F(1, 28) = 2.86, p = .102$. Critically, there was no interaction between these two variables, $F(1, 28) = 0.02, p = .902$ (see Figure 1), thus suggesting that both sensory distractor features were processed *independently* of each other.

ER

For the ER, we conducted the same 2×2 MANOVA as for the RT data. There was a main effect of Tactile Distractor Feature Congruency, $F(1, 28) = 8.69, p = .006, \eta_p^2 = .24$, but no main effect of Visual Distractor Feature Congruency, $F(1, 28) = 0.50, p = .486$, nor any interaction, $F(1, 28) = 1.03, p = .319$.

Between-Experiments Analysis

Experiment 2 vs. 3: A 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) $\times 2$ (Tactile Distractor Feature Congruency: congruent vs. incongruent) $\times 2$ (Experiment: 2 vs. 3) MANOVA with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(1, 59) = 19.72, p < .001, \eta_p^2 = .25$, and Tactile Distractor Feature Congruency $F(1, 59) = 44.97, p < .001, \eta_p^2 = .43$. Interestingly, the two factors did not interact, $F(1, 59) = 0.85, p = .362, \eta_p^2 = .01$, and none of the effects were modulated by Experiment, all $F(1, 59) < 3.23, ps > .077$. This indicates independent, unisensory distractor

processing of both features in the experiments in which the participants focused visually on the target.

Experiment 1 vs. 3: A 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) $\times$ 2 (Tactile Distractor Feature Congruency: congruent vs. incongruent) $\times$ 2 (Experiment: 1 vs. 3) MANOVA with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(1, 58) = 29.18$, $p < .001$, $\eta_p^2 = .34$, modulated by Experiment, $F(1, 58) = 8.62$, $p = .005$, $\eta_p^2 = .13$, indicating a decrease in distractor interference from Experiment 1 to 3. The main effect of Tactile Distractor Feature Congruency was significant, $F(1, 58) = 52.99$, $p < .001$, $\eta_p^2 = .48$, and not modulated by Experiment, $F(1, 58) = 2.65$, $p = .109$. The interaction between these two factors was significant, $F(1, 58) = 5.96$, $p = .018$, $\eta_p^2 = .09$, and crucially, this interaction was modulated by Experiment, $F(1, 58) = 5.07$, $p = .028$, $\eta_p^2 = .08$. This support the finding of a change from multisensory distractor processing (Experiment 1) to independent, unisensory distractor processing (Experiment 3).

Discussion

In Experiment 3, the multisensory target was presented from the centrally-located cube while the distractor was presented from the lateral cube, as in Experiment 2. In comparison to Experiment 2, the spatial distance between the distractor and the center of participants' gaze was doubled (70°). The main effect of congruency for the tactile distractor features was significant in both the RT and ER data. This stands in line with previous evidence concerning unisensory tactile distractor processing. That is, distractor interference effects were still reported even with spatial separations of about 65 to 74 cm (e.g., at a holding distance of 40 cm, this corresponds to a visual angle of about 78° to 86° ; Driver & Grossenbacher, 1996,

Experiment 1a & b, far condition) or even 100 cm (corresponds to a visual angle of about 102°; Soto-Faraco et al., 2004, Experiment 4, hands far condition).

The single analysis of Experiment 3 suggests that no visual distractor interference occurred in Experiment 3 (no significant main effect). In contrast, the between-experiments analysis of Experiments 2 and 3 indicates an overall interference of the visual distractor feature (i.e., a significant main effect was observed), but no significant modulation by Experiment (please note that RTs to visual congruent trials were still 9 ms faster than for visual incongruent trials in Experiment 3, see Table 1). These results therefore imply that some interference of the visual distractor feature in Experiment 3 was still there, although clearly reduced in comparison with Experiment 1. This is supported by the significant interaction between Visual Distractor Congruency and Experiment in the combined analysis of Experiments 1 and 3. As expected, the interaction between the two distractor features in the RT data was not significant. The between-experiments analysis indicated a significant modulation of this interaction between Experiments 1 and 3. That is, overt spatial orientation as well as spatial separation between target and distractor are necessary components to induce a shift from multisensory (Experiment 1) to unisensory processing (Experiment 3). This further underlines the importance of spatial attention in multisensory distractor processing (see also Spence, Pavani, Maravita, & Holmes, 2008).

General Discussion

In this study, a visuotactile flanker task was used to investigate multisensory distractor processing in a multisensory selection task. The participants had to select, as well as respond to, visuotactile targets in the presence of visuotactile distractors. Importantly, overt spatial attention (i.e., fixation) as well as the spatial separation between the target and the distractor

were varied systematically. Reducing the amount of spatial attention devoted toward the distractor by focusing on the target combined with increasing the spatial distance between the distractor and the target ultimately led to a significant shift from multisensory (Experiment 1) to a more independent unisensory processing of the distractor features (Experiment 3).

The results of the three experiments reported here therefore underline the importance of overt spatial attention for multisensory distractor processing. This represents a successful replication of the central findings of our previous audiovisual study (Jensen et al., 2019) as well as the findings from unisensory studies (e.g., Driver & Grossenbacher, 1996; Eriksen & Eriksen, 1974; Fox, 1998; Gallace, Soto-Faraco, Dalton, Kreukniet, & Spence, 2008; Miller, 1991; Soto-Faraco et al., 2004). When the distractor was placed at the center of the participants' gaze (Experiment 1), the visual and tactile feature of the distractor were not processed independently of one another. What this means, in practice, is that a significant interaction was evidenced between Visual and Tactile Distractor Feature Congruency. Here, overt spatial attention was directed toward the distractor. Accordingly, sufficient spatial attention was directed toward the distractor to induce multisensory distractor processing. Placing the target in the center of the participant's gaze, combined with the increase of spatial separation between the distractor and the target, reduced spatial attention toward the distractor and led to the independent processing of the distractor features. Thus, these results indicate a close relationship between spatial attentional resources and multisensory distractor processing.

Overall, the size of the tactile congruency effects were larger than the size of the visual congruency effects in all three experiments (see Table 1). Previous studies have indicated that congruency effects are typically larger for the modality that is not dominating in the current task (Kreutzfeldt, Stephan, Sturm, Willmes, & Koch, 2015; Lukas, Philipp, & Koch, 2010, 2014). In our task, target and distractor were predominantly segregated by location and vision

tends to dominate for spatial localization, therefore largest congruency effects were expected and ultimately found for the tactile modality.

One might be tempted to argue that the decrease of the interference of the visual distractor from Experiments 1 to 3 is the reason for the shift of a significant interaction in Experiment 1 to no interaction in Experiment 3. In other words, one might be tempted to conclude that the processing of the distractor would not have been different between the experiments, but the influence of the features of the distractor were reduced. However, we would like to argue that this explanation seems unlikely, as multisensory processing has been evidenced with strong and, in particular, with weak unisensory signals (Diederich & Colonius, 2004; Senkowski, Saint-Amour, Höfle, & Foxe, 2011). In fact, the idea of inverse effectiveness in multisensory processing (the law of inverse effectiveness / superadditivity, e.g. Holmes & Spence, 2005; Stein & Meredith, 1990, 1993; Stein & Stanford, 2008) suggests a relationship between the strength of the individual unisensory signals and the subsequent multisensory response (though see Spence & Driver, 1999). Specifically, weaker unisensory signals tend to elicit stronger multisensory integration and processing (but see the controversial discussion surrounding the evidence of inverse effectiveness, e.g. Holmes, 2009; Stanford & Stein, 2007; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). Therefore, we would like to conclude that the processing of the distractor is modulated by spatial attention from Experiment 1 to 3.

The present study extends the findings of our previous audiovisual distractor processing study (Jensen et al., 2019). In line with our previous results, we once again identify overt spatial orienting as an important moderator of multisensory distractor processing, this time for visuotactile stimuli. This constitutes a successful replication and indeed generalization of our previous results. But overt spatial attention is not the necessary precondition to modulate distractor processing as previously hypothesized. That is, if the distractor is still too close to the

target, presented at the center of gaze, multisensory distractor processing still occurs. The spatial separation between the target and distractor has to be large enough (in our experimental set-up, somewhere between 35° and 70° visual angle for the visuotactile modality combination) to significantly modulate multisensory processing. Decreasing spatial attention toward the distractor by focusing the target combined with increasing spatial separation led to a significant shift in multisensory distractor processing.

The present results further support an idea of a similar multisensory distractor processing mechanism across the senses. In this study, the same experimental set-up (trial procedure, attention manipulation, as well as visual angle between distractor and target) was used as in our previous audiovisual study (Jensen et al., 2019). Overall, the pattern of results is similar to those reported in our previous study. Distractor features are not processed independently if overt spatial attention resources are directed toward the distractor. These results stand in line with recent frameworks describing the role of attention in multisensory processing (e.g., Navarra et al., 2010; Talsma et al., 2010). That is, multisensory processing only occurs at the attended locations in situations with competing stimulus features (Talsma et al., 2010). In the present study, each of two multisensory cubes were used to present two stimulus features (one visual and one tactile feature, respectively). Covert attention was directed toward the multisensory target cube via instruction (*attend and respond to the stimuli of the left/right cube*). The features of the multisensory target were combined in order to complete the task successfully, as catch trials were incorporated into the task. The features of the multisensory distractor were only combined if sufficient overt spatial attentional resources were directed toward them, once again indicating a crucial role of attention for multisensory distractor processing.

Conclusions

1 The results of the present study emphasize the importance of overt spatial attention for
2 multisensory distractor processing. Across three experiments, overt spatial orientation as well
3 as the spatial separation between the distractor and target were systematically varied in order
4 to investigate multisensory distractor processing. The combined manipulation of overt spatial
5 orientation and spatial separation significantly affected distractor processing, in line with a shift
6 from multisensory to unisensory, independent distractor processing. This result is in line with
7 the recent literature on audiovisual distractor processing (Jensen et al., 2019) as well as current
8 frameworks concerning the interplay of multisensory processing and attention (e.g., Talsma et
9 al, 2010). As such, it would appear that similar mechanisms are implicated in the processing of
10 multisensory distractors in the different spatial senses (namely, vision, touch, and audition).

References

- Alais, D., Ho, T., Han, S. E., & Van der Burg, E. (2017). A matched comparison across three different sensory pairs of cross-modal temporal recalibration from sustained and transient adaptation. *i-Perception*, 8:1-17.
- 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.
- Auvray, M., Gallace, A., Tan, H. Z., & Spence, C. (2007). Crossmodal change blindness between vision and touch. *Acta Psychologica*, 126, 79-97.
- 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.
- Buckolz, E., Goldfarb, A., & Khan, M. (2004). The use of a distractor-assigned response slows later responding in a location negative priming task. *Perception & Psychophysics*, 66, 837-845.
- De Meo, R., Murray, M. M., Clarke, S., & Matusz, P. J. (2015). Top-down control and early multisensory processes: Chicken vs. egg. *Frontiers in Integrative Neuroscience*, 9:17.
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66, 1388-1404.

- 1 Driver, J., & Grossenbacher, P. G. (1996). Multimodal spatial constraints on tactile selective
2 attention. In T. Innui & J. I. McClelland (Eds.), *Attention and performance XVI:*
3 *Information integration in perception and communication* (pp. 209-235). Cambridge,
4 MA: MIT Press.
- 5 Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program.
6 *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- 7 Eriksen, B. A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a
8 target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.
- 9 Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a
10 statistically optimal fashion. *Nature*, 415, 429-433.
- 11 Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in*
12 *Cognitive Sciences*, 8, 162-169.
- 13 Fitzgeorge, L., Buckolz, E., & Khan, M. (2011). Recently inhibited responses are avoided for
14 both masked and nonmasked primes in a spatial negative priming task. *Attention,*
15 *Perception, & Psychophysics*, 73, 1435-1452.
- 16 Fox, E. (1998). Perceptual grouping and visual selective attention. *Perception &*
17 *Psychophysics*, 60, 1004-1021.
- 18 Frings, C., Amendt, A., & Spence, C. (2011). When seeing doesn't matter: Assessing the after-
19 effects of tactile distractor processing in the blind and the sighted. *Journal of*
20 *Experimental Psychology: Human, Perception and Performance*, 37, 1174-1181.
- 21 Frings, C., Bader, R., & Spence, C. (2008). Selection in touch: Negative priming with tactile
22 stimuli. *Perception & Psychophysics*, 70, 516-523.
- 23 Frings, C., Mast, F., & Spence, C. (2014). Tactile spatial negative priming occurs without
24 feature mismatch. *Attention, Perception, & Psychophysics*, 76, 2305-2314.

- 1 Frings, C., Schneider, K. K., & Fox, E. (2015). The negative priming paradigm: An update and
2 implications for selective attention. *Psychonomic Bulletin & Review*, 22, 1577-1597.
- 3 Gallace, A., Soto-Faraco, S., Dalton, P., Kreukniet, B., & Spence, C. (2008). Response
4 requirements modulate tactile spatial congruency effects. *Experimental Brain Research*,
5 191, 171-186.
- 6 Gallace, A., & Spence, C. (2014). *In touch with the future: The sense of touch from cognitive*
7 *neuroscience to virtual reality*. Oxford, UK: Oxford University Press.
- 8 Gallace, A., Tan, H. Z., & Spence, C. (2006). The failure to detect tactile change: A tactile
9 analogue of visual change blindness. *Psychonomic Bulletin & Review*, 13, 300-303.
- 10 Gregg, M. K., & Samuel, A. G. (2008). Change deafness and the organizational properties of
11 sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 34,
12 974-991.
- 13 Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of
14 perceived simultaneity. *Experimental Brain Research*, 235, 763-775.
- 15 Hazeltine, E., Ruthruff, E., & Remington, R. W. (2006). The role of input and output modality
16 pairings in dual-task performance: Evidence for content-dependent central interference.
17 *Cognitive Psychology*, 52, 291-345.
- 18 Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some
19 statistical considerations. *Brain Topography*, 21, 168-176.
- 20 Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity.
21 *Current Biology*, 15, R762-R764.
- 22 Honoré, J., Bourdeaud'hui, M., & Sparrow, L. (1989). Reduction of cutaneous reaction time by
23 directing eyes towards the source of stimulation. *Neuropsychologia*, 27, 367-371.

- 1 Jensen, A., Merz, S., Spence, C., & Frings, C. (2019). Overt spatial attention modulates
2 multisensory selection. *Journal of Experimental Psychology: Human, Perception and*
3 *Performance*, 45, 174-188.
- 4 Kreutzfeldt, M., Stephan, D. N., Sturm, W., Willmes, K., & Koch, I. (2015). The role of
5 crossmodal competition and dimensional overlap in crossmodal attention switching.
6 *Acta Psychologica*, 155, 67-76.
- 7 Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of*
8 *Experimental Psychology: Human, Perception and Performance*, 21, 451-468.
- 9 Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive*
10 *Sciences*, 9, 75-82.
- 11 Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions*
12 *in Psychological Science*, 19, 143-148.
- 13 Lukas, S., Philipp, A. M., & Koch, I. (2010). Switching attention between modalities: Further
14 evidence for visual dominance. *Psychological Research*, 74, 255-267.
- 15 Lukas, S., Philipp, A. M., & Koch, I. (2014). Crossmodal attention switching: Auditory
16 dominance in temporal discrimination tasks. *Acta Psychologica*, 153, 139-146.
- 17 Mayr, S., & Buchner, A. (2007). Negative priming as a memory phenomenon. *Journal of*
18 *Psychology*, 215, 35-51.
- 19 Merz, S. (2019). Data for 'Multisensory distractor processing is modulated by spatial attention'
20 [Data file]. Available from doi:<https://doi.org/10.23668/psycharchives.2452>
- 21 Miller, J. (1991). The flanker compatibility effect as a function of visual angle, attentional
22 focus, visual transients, and perceptual load: A search for boundary conditions.
23 *Perception & Psychophysics*, 49, 270-288.

- 1 Möller, M., Mayr, S., & Buchner, A. (2013). Target localization among concurrent sound
2 sources: No evidence for the inhibition of previous distractor responses. *Attention,*
3 *Perception, & Psychophysics*, 75, 132-144.
- 4 Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of attention in
5 the audiovisual integration of speech. *Information Fusion*, 11, 4-11.
- 6 Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological*
7 *Bulletin*, 116, 220-244.
- 8 Pashler, H. (1998). *Attention*. New York, NY: Psychology Press.
- 9 Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-
10 processing account of its origins and significance. *Psychological Review*, 83, 157-171.
- 11 Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of
12 perceptual load. *Journal of Experimental Psychology: Human, Perception and*
13 *Performance*, 33, 1311-1321.
- 14 Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in
15 early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage*,
16 56, 2200-2208.
- 17 Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear.
18 *Nature*, 408, 788.
- 19 Shopland, C., & Gregory, R. L. (1964). The effect of touch on a visually ambiguous three-
20 dimensional figure. *Quarterly Journal of Experimental Psychology*, 16, 66-70.
- 21 Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in*
22 *Cognitive Sciences*, 9, 16-20.
- 23 Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita
24 effect revisited. *Perception & Psychophysics*, 69, 673-686.

- 1 Soto-Faraco, S., Ronald, A., & Spence, C. (2004). Tactile selective attention and body posture:
2 Assessing the multisensory contributions of vision and proprioception. *Perception &*
3 *Psychophysics*, 66, 1077-1094.
- 4 Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science & Technology*, 28,
5 61-70.
- 6 Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain*
7 *Research*, 176, 245-258.
- 8 Spence, C. (2013). Just how important is spatial coincidence to multisensory integration?
9 Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296, 31-49.
- 10 Spence, C., & Driver, J. (1999). A new approach to the design of multimodal warning signals.
11 In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics, Vol. 4: Job*
12 *design, product design and human-computer interaction* (pp. 455-461). Hampshire:
13 Ashgate Publishing.
- 14 Spence, C., Lloyd, D., McGlone, F., Nicholls, M. E. R., & Driver, J. (2000). Inhibition of return
15 is supramodal: A demonstration between all possible pairings of vision, touch and
16 audition. *Experimental Brain Research*, 134, 42-48.
- 17 Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal
18 distractor congruency effects. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 148-
19 169.
- 20 Spence, C., Pavani, F., Maravita, A., & Holmes, N. P. (2008). Multi-sensory interactions. In M.
21 C. Lin & M. A. Otaduy (Eds.), *Haptic rendering: Foundations, algorithms, and*
22 *applications* (pp. 21-52). Wellesley, MA: AK Peters.
- 23 Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the
24 computation in context. *Neuroreport*, 18, 787-792.

- 1 Stein, B. E., Burr, D., Constantinides, C., Laurienti, P. J., Meredith, A. M., Perrault, T. J., et al.
2 (2010). Semantic confusion regarding the development of multisensory integration: A
3 practical solution. *European Journal of Neuroscience*, 31, 1713-1720.
- 4 Stein, B. E., & Meredith, M. A. (1990). Multisensory integration. Neural and behavioral
5 solutions for dealing with stimuli from different sensory modalities. *Annals of the New*
6 *York Academy of Sciences*, 608, 51-70.
- 7 Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- 8 Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the
9 perspective of the single neuron. *Nature Reviews Neuroscience*, 9, 255-267.
- 10 Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009).
11 Challenges in quantifying multisensory integration: Alternative criteria, models, and
12 inverse effectiveness. *Experimental Brain Research*, 198, 113-126.
- 13 Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5).
14 Boston, MA: Pearson.
- 15 Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of
16 the multisensory mind. *Frontiers in Integrative Neuroscience*, 9:19.
- 17 Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted
18 interplay between attention and multisensory integration. *Trends in Cognitive Sciences*,
19 14, 400-410.
- 20 Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with
21 endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, 61, 208-
22 224.
- 23 Van der Burg, E., Orchard-Mills, E., & Alais, D. (2015). Rapid temporal recalibration is unique
24 to audiovisual stimuli. *Experimental Brain Research*, 233, 53-59.

- 1 Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-
2 action, animal call, and musical stimuli. In M. J. Naumer & J. Kaiser (Eds.),
3 *Multisensory object perception in the primate brain* (pp. 95-121). New York, NY:
4 Springer.
- 5 Wesslein, A., Spence, C., & Frings, C. (2014). Vision affects tactile target and distractor
6 processing even when space is task-irrelevant. *Frontiers in Psychology*, 5:84.
- 7 Wesslein, A. K., Spence, C., Mast, F., & Frings, C. (2016). Spatial negative priming: In touch,
8 it's all about location. *Attention, Perception, & Psychophysics*, 78, 464-473.
- 9 Zmigrod, S., & Hommel, B. (2011). The relationship between feature binding and
10 consciousness: Evidence from asynchronous multi-modal stimuli. *Consciousness and*
11 *Cognition*, 20, 586-593.
- 12 Zmigrod, S., Spapé, M., & Hommel, B. (2009). Intermodal event files: Integrating features
13 across vision, audition, taction, and action. *Psychological Research*, 73, 674-684.

1

2

Authors Note

3

4

5

6

Simon Merz, Anne Jensen, and Christian Frings, University of Trier, Department of Psychology, D-54286, Germany (Email: merz@uni-trier.de). Charles Spence, Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, United Kingdom.

7

8

9

10

11

The research reported in this article was supported by a grant of the Deutsche Forschungsgemeinschaft to Christian Frings and Charles Spence (FR 2133/5-3). Correspondence concerning this article should be addressed to Simon Merz, University of Trier, Department of Psychology, D-54286, Germany (Email: merz@uni-trier.de). We thank Stephanie Blasl for the drawings incorporated in Figure 1.

Appendices

Appendix A

The multisensory cube (MC) used in this study was developed in cooperation with the Trier University of Applied Science. Key features include the ability to present visual, tactile, and auditory signals simultaneously. The MC is 70 mm³, ensuring a comfortable handling for women and men alike (based on DIN 33402). The weight of the fully-equipped MC is about 125 gram, allowing an easy and comfortable holding for the duration of an experiment. Importantly, the MC is separated in two parts, an outer shell and an inner supporting structure (see Figure A1). All of the technical devices are attached to the supporting structure. For optimal presentation, the outer shell has cut-outs at the corresponding locations of the technical devices.

Figure A1: 3D images of the self-developed multisensory cube (MC). (a) On the left, the MC with its outer shell impermeable. On the top (seen in the depiction) and the bottom (not shown in the depiction), a cut-out is incorporated to allow contact with the two integrated tactors. On the front surface, two cut-outs are incorporated in order to allow the LEDs to be seen. At the back, several small cut-outs (25 3x3 mm cut-outs, not seen in the depiction) are incorporated to allow the loudspeaker to present sound stimuli. (b) On the right, the MC with its outer shell permeable, to allow a view in the inside of the MC

and the supporting structure. On this supporting structure, all technical devices, two tactors (on the top and bottom), two LED (up front) as well as one loudspeaker (at the back) are attached.

A maximum of two visual, two tactile, as well as one auditory stimulus can be presented at any time from each cube. The two LEDs (5 mm² each) are attached on top of each other at the front of the supporting structure. A tactor (Model C-2, Engineering Acoustic, Inc., USA) is located behind each LED. Therefore, each tactor is placed in close proximity to one LED. Importantly, an AIREX Foam vibration damper (Airex foam C70, 3A Composites Core Materials Switzerland), offering the best damping qualities when compared to twelve different vibration damper systems, connects the tactors with the supporting structure. This reduces the transmission of the vibration onto the supporting structure. At the back of the supporting structure, a loudspeaker (Model K 14.25, Visaton Germany) was attached. All of the devices are accessed via a serial interface.

The damping qualities were assessed with the laser triangulation procedure (laser: optoNCDT 1401-50, Micro-Epsilon; analog-to-digital converter: NI-9234, National Instruments). This procedure assesses the distance by calculating the angle with which a laser returns to a receiving element. The simultaneous activation of both tactors with a peak-to-peak amplitude of about 418 µm each resulted in a peak-to-peak amplitude at the shell of less than 25 µm (reduction of factor 17). The activation of about 272 µm peak-to-peak amplitude at one tactor was passively assessed at the other tactor location with about 55 µm peak-to-peak amplitude (reduction of factor 5).

Appendix B – Full model analysis

Experiment 1

RTs: A 3 (Visual Distractor Feature Congruency: congruent, neutral, or incongruent) × 3 (Tactile Distractor Feature Congruency: congruent, neutral, or incongruent) MANOVA with

Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable (see Table A1 for the mean RTs and ERs). The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(2, 29) = 42.68, p < .001, \eta_p^2 = .75$, with participants responding more rapidly when the visual target and distractor feature were congruent and slowest when they were incongruent. The main effect of Tactile Distractor Feature Congruency was significant, $F(2, 29) = 26.34, p < .001, \eta_p^2 = .65$. That is, participants responded fastest when the tactile target and distractor feature were congruent and slowest when they were incongruent. Critically, there was a significant interaction between these two factors, $F(4, 27) = 5.64, p < .001, \eta_p^2 = .46$.

We further conducted three 3×3 (Visual Distractor Feature Congruency: congruent, neutral, or incongruent) $\times$ 3 (Tactile Distractor Feature Congruency: congruent, neutral, or incongruent) $\times$ 2 MANOVAs with one of three additional factors. These additional factors were the between-participants factor of either Target Block Order (left vs. right first), Target Feature Mapping (strong vibration mapped with red light and weak vibration mapped with blue light vs. strong vibration mapped with blue light and weak vibration mapped with red light) or the within-participants factor of Target Side (left vs. right), respectively. Crucially, none of the three-way interactions was significant, all $F_s < 2.01$, all $p_s > .123$.

ER: The same 3×3 MANOVA was conducted for the ERs as for the RTs. There was a main effect of Tactile Distractor Feature Congruency, $F(2, 29) = 3.35, p = .049, \eta_p^2 = .19$, and a significant interaction, $F(4, 27) = 4.81, p = .005, \eta_p^2 = .42$, but no main effect of Visual Distractor Feature Congruency, $F(2, 29) = 0.66, p = .526, \eta_p^2 = .04$. Just as for the RT data, the three $3 \times 3 \times 2$ MANOVAs with the additional factor of either Target Side, Target Block Order, or Target Feature Mapping revealed no significant three-way interaction, all $F_s < 0.60$, all $p_s > .668$.

Experiment 2

RTs: A 3 (Visual Distractor Feature Congruency: congruent, neutral, or incongruent) $\times$ 3 (Tactile Distractor Feature Congruency: congruent, neutral, or incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable (see Table A1 for the mean RT and ER). The MANOVA revealed a significant main effect of Visual Distractor Feature Congruency, $F(2, 30) = 15.29, p < .001, \eta_p^2 = .51$; that is, participants responded more rapidly when the visual target and distractor features were congruent and slowest when they were incongruent. The main effect of Tactile Distractor Feature Congruency was significant, $F(2, 30) = 21.38, p < .001, \eta_p^2 = .59$. That is, the participants responded fastest when the tactile target and distractor features were congruent and slowest when they were incongruent. Critically, there was no significant interaction between these two variables, $F(4, 28) = 0.72, p = .583, \eta_p^2 = .10$, thus suggesting that the two sensory distractor features were processed *independently* of one another.

Just as for Experiment 1, three further 3 (Visual Distractor Feature Congruency: congruent, neutral, or incongruent) $\times$ 3 (Tactile Distractor Feature Congruency: congruent, neutral, or incongruent) $\times$ 2 MANOVAs were conducted with one of three additional factors. The three additional factors of either Target Side (within-participants: left vs. right), Target Block Order (between participants: left vs. right first), or Target Feature Mapping (between-participants: strong vibration with red light and weak vibration with blue light vs. strong vibration with blue light and weak vibration with red light) revealed no significant three-way interaction, all F s $< 1.70, p$ s $> .179$.

ERs: For the ER, the same 3×3 MANOVA was conducted as for the RT data. There was a main effect of Tactile Distractor Feature Congruency, $F(2, 30) = 3.65, p = .038, \eta_p^2 = .20$, but no main effect of Visual Distractor Feature Congruency, $F(2, 30) = 0.97, p = .391, \eta_p^2 = .06$,

and no interaction, $F(4, 28) = 0.40, p = .805, \eta_p^2 = .05$. Just as for the RT data, the three $3 \times 3 \times 2$ MANOVAs with the additional factor of either Distractor Side, Distractor Block Order, or Target Feature Mapping revealed no significant three-way interaction, all $F_s < 1.75, p_s > .167$.

Experiment 3

RTs: A 3 (Visual Distractor Feature Congruency: congruent vs. neutral vs. incongruent) $\times 3$ (Tactile Distractor Feature Congruency: congruent vs. neutral vs. incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable (see Table A1 for the mean RT and ER). The MANOVA revealed a significant main effect of Tactile Distractor Feature Congruency, $F(2, 27) = 19.95, p > .001, \eta_p^2 = .60$, with participants responding most rapidly when the visual target and distractor features were congruent and slowest when they were incongruent. The main effect of Visual Distractor Feature Congruency was not significant, $F(2, 27) = 2.10, p = .142, \eta_p^2 = .14$. Critically, there was no interaction between these two variables, $F(4, 25) = 0.98, p = .435, \eta_p^2 = .14$, suggesting that the two sensory distractor features were processed *independently* of one another.

Just as for Experiments 1 and 2, we further conducted three 3 (Visual Distractor Feature Congruency: congruent, neutral, or incongruent) $\times 3$ (Tactile Distractor Feature Congruency: congruent, neutral, or incongruent) $\times 2$ MANOVAs with one of three additional factors. These additional factors were the between-participants factor of either Target Block Order (left vs. right first), Target Feature Mapping (strong vibration mapped with red light and weak vibration mapped with blue light vs. strong vibration mapped with blue light and weak vibration mapped with red light) or the within-participants factor of Target Side (left vs. right), respectively. Comparable to the results of Experiments 1 and 2, the MANOVAs revealed no significant three-way interaction, all $F_s < 1.0, p_s > .429$.

ERs: For the ER, we conducted the same 3×3 MANOVA as for the RT data. There was a main effect of Tactile Distractor Feature Congruency, $F(2, 27) = 3.42$, $p = .047$, $\eta_p^2 = .20$, but no main effect of Visual Distractor Feature Congruency, $F(2, 27) = 0.23$, $p = .793$, and no interaction, $F(4, 25) = 0.80$, $p = .534$. As for the RT data, the three $3 \times 3 \times 2$ MANOVAs with the additional factor of either Distractor Side, Distractor Block Order or Target Feature Mapping revealed no significant three-way interaction, all $F_s < 2.11$, $p_s > .111$.

Table A1: Mean RT (in ms) and error rates (with standard errors in brackets) for Experiments 1-3 for each Visual (congruent vs. incongruent) and Tactile (congruent vs. incongruent) Distractor Feature Congruency combination.

Visual distractor feature congruency	Reaction times (RTs)			Error rates (ERs)		
	Tactile distractor feature congruency			Tactile distractor feature congruency		
	Congruent	Neutral	Incongruent	Congruent	Neutral	Incongruent
Experiment 1						
Congruent	737 (16.8)	762 (16.1)	781 (16.3)	4.54 (0.90)	4.03 (0.96)	8.77 (2.01)
Neutral	756 (16.6)	766 (18.3)	793 (17.0)	4.64 (1.00)	4.64 (1.11)	6.25 (1.86)
Incongruent	782 (17.0)	783 (16.8)	799 (17.1)	3.83 (1.02)	5.04 (1.33)	7.56 (1.95)
Experiment 2						
Congruent	732 (21.4)	744 (20.2)	765 (20.8)	4.39 (0.80)	4.00 (0.89)	7.23 (1.61)
Neutral	737 (18.8)	746 (17.9)	771 (20.6)	4.10 (0.71)	3.52 (0.86)	5.27 (1.14)
Incongruent	759 (18.7)	761 (18.1)	781 (18.8)	3.52 (0.65)	3.81 (0.92)	5.96 (1.11)
Experiment 3						
Congruent	778 (14.9)	786 (15.7)	798 (16.0)	4.31 (0.91)	6.03 (1.11)	10.13 (2.01)
Neutral	773 (16.2)	789 (14.8)	808 (15.9)	5.17 (1.07)	5.60 (1.06)	8.94 (2.20)
Incongruent	788 (15.5)	791 (15.6)	807 (16.1)	4.53 (0.94)	5.71 (1.17)	8.73 (1.94)