

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

RUNNING HEAD: MULTISENSORY SELECTION

Perception it is:

Processing level in multisensory selection

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

¹ University of Trier,

² University of Oxford

Words count (8651)

Corresponding author:

Anne Jensen

University of Trier

D-54286 Trier

jensen@uni-trier.de

Authors' Note

Anne Jensen, Simon Merz, and Christian Frings, University of Trier, Department of Psychology, D-54286, Germany, and Charles Spence, Crossmodal Research Laboratory, Department of Experimental Psychology, Anna Watts Building, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom.

The research reported in this article was supported by a grant from the Deutsche Forschungsgemeinschaft to Christian Frings and Charles Spence (FR 2133/5-3).

Correspondence concerning this article should be addressed to Anne Jensen, University of Trier, Department of Psychology, D-54286, Germany, Email: jensen@uni-trier.de.

Abstract

When repeatedly exposed to simultaneously-presented stimuli, associations between these stimuli are nearly always established, both within as well as between sensory modalities. Such associations guide our subsequent actions and may also play a role in multisensory selection. Thus, crossmodal associations (i.e., associations between stimuli from different modalities) learned in a multisensory interference task might affect subsequent processing. The aim of the present study was to investigate the processing level of multisensory stimuli in multisensory selection by means of crossmodal aftereffects. Either feature or response associations were induced in a multisensory flanker task while the amount of interference in a subsequent crossmodal flanker task was measured. The results of Experiment 1 revealed the existence of crossmodal interference after multisensory selection. Experiments 2 and 3 then demonstrated the dependence of this effect on the perceptual associations between features themselves, rather than on the associations between feature and response. Establishing response associations did not lead to a subsequent crossmodal interference effect (Experiment 2), while stimulus feature associations without response associations (obtained by changing the response effectors) did (Experiment 3). Taken together, this pattern of results suggests that associations in multisensory selection, and the interference of (crossmodal) distractors, predominantly work at the perceptual, rather than at the response, level.

Keywords: Multisensory perception, multisensory selection, distractor processing, multisensory associations, processing level

1

2

Public Significance Statement

3

4

5

6

7

8

9

10

This study demonstrates that crossmodal distractor interference induced by the learning of multisensory associations predominantly operates on a perceptual, rather than on a response, level. Crossmodal interference was elicited when features in a previous multisensory task were frequently presented together. Interference was still found when the response changed across tasks, but not when associations between stimulus features were prevented. These findings provide novel insight concerning the processing of multisensory stimuli. Additionally, such insights may have significant implications for the design of future multisensory alerts and machine interfaces.

Introduction

Each and every day, we are exposed to multiple sensory signals that help us to explore our surroundings efficiently. Understanding the underlying mechanisms of multisensory perception is fundamental as far as improving the development of everyday environments, or technical devices that humans interact with, is concerned. When confronted with inputs from different sensory modalities, we can mostly tell whether they come from the same source (i.e., the same object or event) based on various rules for what belongs together. For example, we learn that cats meow while dogs bark (Chen & Spence, 2010). Along similar lines, we also learn arbitrary associations, such as the ringing of the doorbell or the ring tone from our mobile device (see Walker-Andrews, 1994, on the notion of arbitrary crossmodal correspondences). In the laboratory, associations like this can be generated simply by presenting stimuli together frequently.

Multisensory association learning

After frequent exposure to a specific combination of sensory signals, the presentation of one signal can improve the processing of the other signal, as this can lead to the recall of the other signal. This is a very well-established idea in older theories of ‘binding’, ‘conditioning’, and ‘perceptual and associative learning’ (e.g., Hall, 1991; Hall & Honey, 2014; Hommel, 2004, 2009; Mackintosh, 1965; McLaren & Mackintosh, 2000; Pavlov, 1927; Rescorla, 1969; Rescorla & Wagner, 1972; Wagner, 1969a, b). Such ideas have recently been revived in the field of multisensory information processing (see e.g., Murray, Lewkowicz, Amedi, & Wallace, 2016, for a review). From unisensory studies, we know that the frequent co-occurrence of a pair of stimuli can affect the strength of associations as lifelong learned associations (yellow-banana) appear to be stronger than newly-established ones (e.g., Colzato, Raffone, & Hommel, 2006; see also Treisman, 1988). In the case of multisensory associations (i.e., when stimuli

1 consist of features from different modalities; see Murray et al., 2016; Shams & Seitz, 2008, for
2 a review), associations might equally affect processing. Multisensory associations were shown
3 to modulate interactions between the senses and subsequent information processing. This was
4 found for highly overlearned associations or correspondences (e.g., Fiebelkorn, Foxe, &
5 Molholm, 2010; Guo & Guo, 2005; Knoeferle, Knoeferle, Velasco, & Spence, 2016; Lehmann
6 & Murray, 2005; Molholm, Martinez, Shpaner, & Foxe, 2007; Molholm, Ritter, Javitt, & Foxe,
7 2004; Murray et al., 2004; Murray, Foxe, & Wylie, 2005; Seitz, Kim, & Shams, 2006; Von
8 Kriegstein & Giraud, 2006), but also when the different sensory signals had no obvious relation,
9 or correspondence, before the experiment (e.g., Baier, Kleinschmidt, & Müller, 2006; Ernst,
10 2007; Zhangenehpour & Zatorre, 2010).

11 Within experiments with long-term learned stimuli, the recall of a unisensory stimulus
12 is improved when it is learned in a meaningful multisensory context (Guo & Guo, 2005;
13 Lehmann & Murray, 2005; Murray et al., 2004, 2005). Similarly, the previous presentation of
14 corresponding audiovisual stimuli facilitates visual and auditory task performance (Knoeferle
15 et al., 2016; Seitz et al., 2006; Von Kriegstein & Giraud, 2006). For instance, the presentation
16 of a product or brand with its characteristic advertisement sound had a positive effect on visual
17 search latencies (Knoeferle et al., 2016). Next to reaction times (RTs), eye-tracking data were
18 considered and revealed a similar benefit of the sounds for the visual processing, thus showing
19 that visual attention is already affected at an early perceptual stage.

20 Concerning short-term learned associations, participants were better in discriminating
21 visual (or haptic) stimuli that were presented with a congruent haptic (or visual) distractor when
22 the two dimensions had been specifically correlated in a previous training phase (Ernst, 2007).
23 Accordingly, participants learned (and integrated) the correlation of specific stimulus
24 dimensions concerning visual and haptic information, which affected their subsequent
25 processing in the posttests. These effects were still present in a second posttest conducted one

day later. Moreover, and as might be expected, changes can also be seen in brain activity as a result of multisensory association learning. For instance, Zhangenehpour and Zatorre (2010) documented crossmodal activation (e.g., activation of the visual area when a sound is presented) after short-term exposure to arbitrary audiovisual stimuli (480 trials in 45 min), which could be measured up to one day after exposure.

These latter studies demonstrate that multisensory associations can affect information processing even when they are newly-learned and the stimuli have only been experienced together for a short period of time. This kind of arbitrary association typically benefits from ‘categorical congruence’ (Barenholtz, Lewkowicz, Davidson, & Mavica, 2014; see also Shams & Seitz, 2008). That is, combinations of stimuli that are statistically more likely to go together in daily life are learned with advantage in comparison to incongruent or neutral combinations. Consequently, subsequent processes are more affected by the learning of congruent stimulus pairings. Thus, these natural correlations always have to be considered. Comparing unisensory and multisensory learning, Shams and Seitz (2008) suggested an advantage for multisensory associations as a more natural form of learning than unisensory associations, being more effectively acquired as well as stronger. However, this mainly holds true for congruent multisensory learning and might be affected by several other moderators. Taken together, learning processes (short- as well as long-term) have a lasting impact on subsequent processing, multisensory as well as unisensory.

The level of interference in multisensory selection

The present study was designed to investigate the processing level at which multisensory associations induce crossmodal distractor interference in a crossmodal selection task in order to draw conclusions concerning the processing of the stimuli in multisensory selection. In studies of ‘selection’, perceptual effects are typically distinguished from response-

related effects (see e.g., Zhang, Zhang, & Kornblum, 1999). In other words, the facilitation or interference by an irrelevant stimulus might occur due to perceptual congruency (associated vs. not-associated distracting input) or due to response compatibility (same vs. different response of distracting input). When participants have to carry out a specific response to a stimulus, a response association between target and response is created (e.g., Frings et al., in press; Hommel, 2004). In the case of response compatibility, distractor features are mapped on to the same or different response as the target and interference is induced by this response overlap. Previous studies have demonstrated this kind of response interference for crossmodal stimuli without perceptual congruency (Mast, Frings, & Spence, 2014; Wesslein, Frings, & Spence, 2014). Namely, a distractor in a different sensory modality from the target has been shown to influence the RT to the target simply because the same response was learned for both stimuli previously. Meanwhile, at the perceptual level, robust crossmodal congruency effects have been demonstrated between stimuli from different sensory modalities (Chen & Spence, 2010, 2011, 2013, 2018; Frings & Spence, 2010; Spence, Pavani, Driver, 2004).

The definition of perceptual congruency in this case is by no means simple to determine as features from different modalities cannot be identical (which normally is meant with congruency). When the term crossmodal congruency is used, it always refers to associations between features. These associations might be semantic (e.g., barking sound and the image of a dog, Chen & Spence, 2010; see also Spence, 2011), common characteristics shared between different stimuli (e.g., such as their elevation in a spatial discrimination task, Spence, Pavani, & Driver, 2004; rhythm, Frings & Spence, 2010) or associations due to frequent co-presentation (for newly-acquired combinations that are not already considered as semantically-related). In the present study, frequent co-presentation from the same location/object is used to induce crossmodal congruency. In order to avoid any mix-up with congruency based on identical

features (which is referred to in the multisensory flanker task), the term perceptual interference will be used for this concept.

In a previous study, we introduced a multisensory flanker task (Jensen, Merz, Spence, & Frings, 2019), in which the participants were frequently presented with, and hence responded to, specific combinations of visual and auditory features. In particular, a red (blue) light and high (low) pitch was mapped to a response while participants were not allowed to respond to a reversed combination (i.e., catch trials; red/blue light with low/high pitch). The target stimuli that participants responded to were presented from one specific location, while, at the same time, irrelevant multisensory stimuli, i.e., audiovisual feature combinations, were presented from another. These distractors significantly influenced information processing. Considering the constraints of associative learning, it would seem most likely that the participants learned the associations between the visual and auditory features as both dimensions were task-relevant. Moreover, when the location of the irrelevant stimulus was overtly attended (i.e., people fixated it), an interaction between the multisensory distractor features was obtained, one that can be traced back to these associations. The relevant target combinations seemed to influence processing to a greater extent than did other combinations, especially in the case of congruent features. From the literature reviewed so far, it can be assumed that these multisensory associations might strengthen over time and even be maintained after the task has finished (see e.g., Ernst, 2007; Zangenehpour & Zatorre, 2010). If so, it should be possible to induce interference between the associated features in a subsequent crossmodal flanker task, in which the stimuli are unisensory but from different modalities.

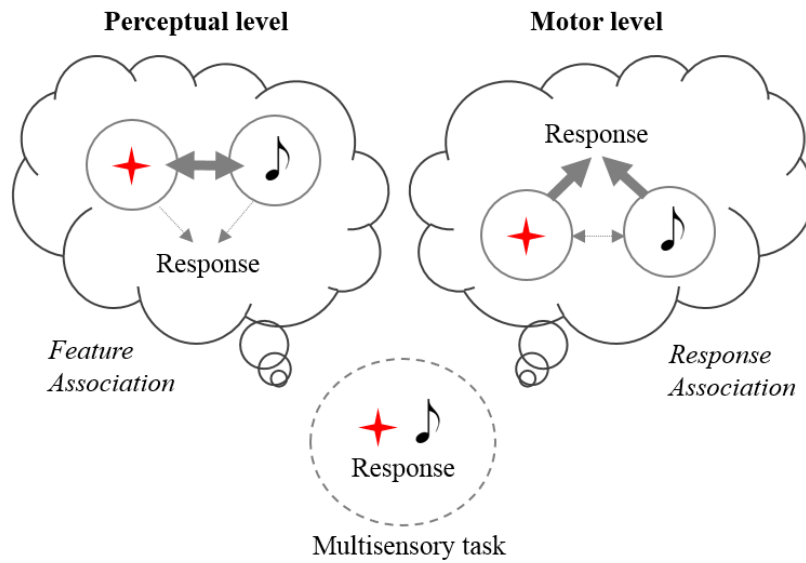


Figure 1. Schematic representation of the mental processes that might be taking place during the multisensory flanker interference task. The star represents the visual feature, the note represents the auditory feature. Dotted circles indicate common presentation/action. The thought bubbles represent the two possibilities as to how associations might be generated in the multisensory task (feature vs. response association). Note that the multisensory distractor is left out of this scheme.

Regarding the level of interference, there are, however, two possibilities (see Figure 1; see Spence & Frings, in press, for a discussion on feature integration in multisensory processing). In the multisensory flanker task (Jensen et al., 2019), the two target features (e.g., red light and high pitch) were presented together on 80% of the trials, thus promoting the learning of the associations between these features (see e.g., Murray et al., 2016). However, as participants pressed one specific key (or footpedal) when the features were presented, associations between the individual features and the response might also be created. Consequently, it remains unclear which type of association dominates processing and hence

elicits **interference between the features in a crossmodal setting**. On the one hand, the two features from different sensory modalities could each be independently associated with the response (response association, see the right thought bubble in Figure 1). The link between response and feature would be stronger than the link between the features. In this case, both features should independently be capable of recalling this associated response and crossmodal interference would predominantly be based on response compatibility. On the other hand, the association between the features might be stronger than between feature and response (feature association; see the left thought bubble in Figure 1). In the latter case, crossmodal interference would rather be evoked at the perceptual level, independent of whether the response in the crossmodal situation happens to be the same as in the previous multisensory context or not. The effect that goes back to this feature association, without considering the response effect, is referred to as a perceptual interference effect. In contrast, the effect of having the same response association is called response interference effect.

The present study

Across three experiments, we investigated the processing level at which multisensory associations induce crossmodal distractor interference by measuring the aftereffects of multisensory selection. We combined a multisensory flanker interference task (Jensen et al., 2019) with audiovisual target and distractor stimuli, and a crossmodal flanker interference task with unisensory stimuli from the visual and auditory modality.

In the multisensory flanker task, the stimulus features of the two targets are typically paired in 80% of the trials. **That is, red light and high pitch (as an example) are presented more often than the reversed combination (these catch trials accounted for the remaining 20% of the trials. Thus, feature associations were allowed.** In Experiment 1, we demonstrate an interference effect in a crossmodal interference task induced by previous multisensory processing. As the

same responses (footpedal responses) were used in the multisensory and the crossmodal task, this finding can be explained by perceptual as well as by response associations of the features. In order to disentangle these two possibilities, two further experiments were also conducted. In Experiment 2, the multisensory task was modified by using four equally often presented targets, instead of two, and switching from footpedal to manual responses. The responses in the multisensory and crossmodal task were still the same, though, as hands were used for responding in the crossmodal task as well. The target features of Experiment 1 (two visual and two auditory features) were varied orthogonally to each other, thus resulting in four combinations, either of which was assigned to a response. In other words, the previous catch trials were now target trials and accounted for 50% of all trials. With this arrangement, any effects evidenced must only be based on response compatibility as no specific feature combination holds advantage over another. In Experiment 3, we used the multisensory task from Experiment 1 with two targets again, except that now the response effector was changed from the multisensory to the crossmodal task in order to eliminate any response compatibility between the tasks. With this, we could evidence any effects that were based solely on perceptual associations.

To foreshadow our findings, multisensory flanker effects were always observed in the multisensory interference task. In the crossmodal flanker task, interference was observed in Experiments 1 and 3, but not in Experiment 2. This pattern of results suggests that the multisensory associations in the multisensory task lasting beyond the task primarily operate at a perceptual, rather than at a response, level.

Experiment 1

Method

Participants. In the literature, and based on our own previous laboratory research, flanker interference effects typically elicit medium effects sizes (irrespective of the modality in which the stimuli happen to be presented). Hence, in order to make sure that we at least observed a unisensory flanker effect, we 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 (note that power analyses were run with G-power 3.1.9.2; Erdfelder, Faul, & Buchner, 1996; Faul, Erdfelder, Lang, & Buchner, 2007). Thirty-two students from the University of Trier took part in the study in return for course credit. However, four of the participants were excluded due to extremely high error rates in the catch trials in the first part of the study, thus presumably indicating a lack of understanding of the task (these participants were outliers with respect to the errors in catch trials of the sample, Mean = 10%, SD = 9%, outlier with > 49% errors). Another participant was excluded due to extremely high error rates in both parts of the study (again, this participant was an outlier with respect to error rates of the sample, Mean = 4%, SD = 3%, outlier with 23% errors). A total of 27 participants (6 male) aged from 19 to 46 years (Mean = 22, SD = 5) were included in the final data analyses. All of the participants had normal or corrected-to-normal vision and audition 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 (70 cm) was kept constant by means of a chinrest with forehead support. The instructions were presented in grey lettering against a black background in the middle of the screen. Combinations of three different tones (low pitch presented at 440 Hz, medium pitch presented at 1046 Hz, and high pitch presented at 2793 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 experimental stimuli. These were presented over two custom-made multisensory cubes with two LEDs at the front side (one on top, the other at the bottom) and a loudspeaker on the rear side (see Merz,

1 Jensen, Spence, & Frings, 2018, for a detailed description of the cubes). The upper LED on
2 each cube was used to present the visual color stimuli. The cubes were placed laterally, one to
3 either side of the computer monitor (see Figure 2). A distance of 80 cm separated the cubes
4 (corresponding to a visual angle of 60°). The participants responded by means of foot pedals
5 using Chronos® (Version 1.0). The participants were instructed to stare at a fixation cross in
6 the middle of the computer screen throughout each block of experimental trials. The experiment
7 was conducted with the E-Prime software (Version 2.0) and the data were analyzed with IBM
8 SPSS Statistics (Version 23).

9

Multisensory Selection

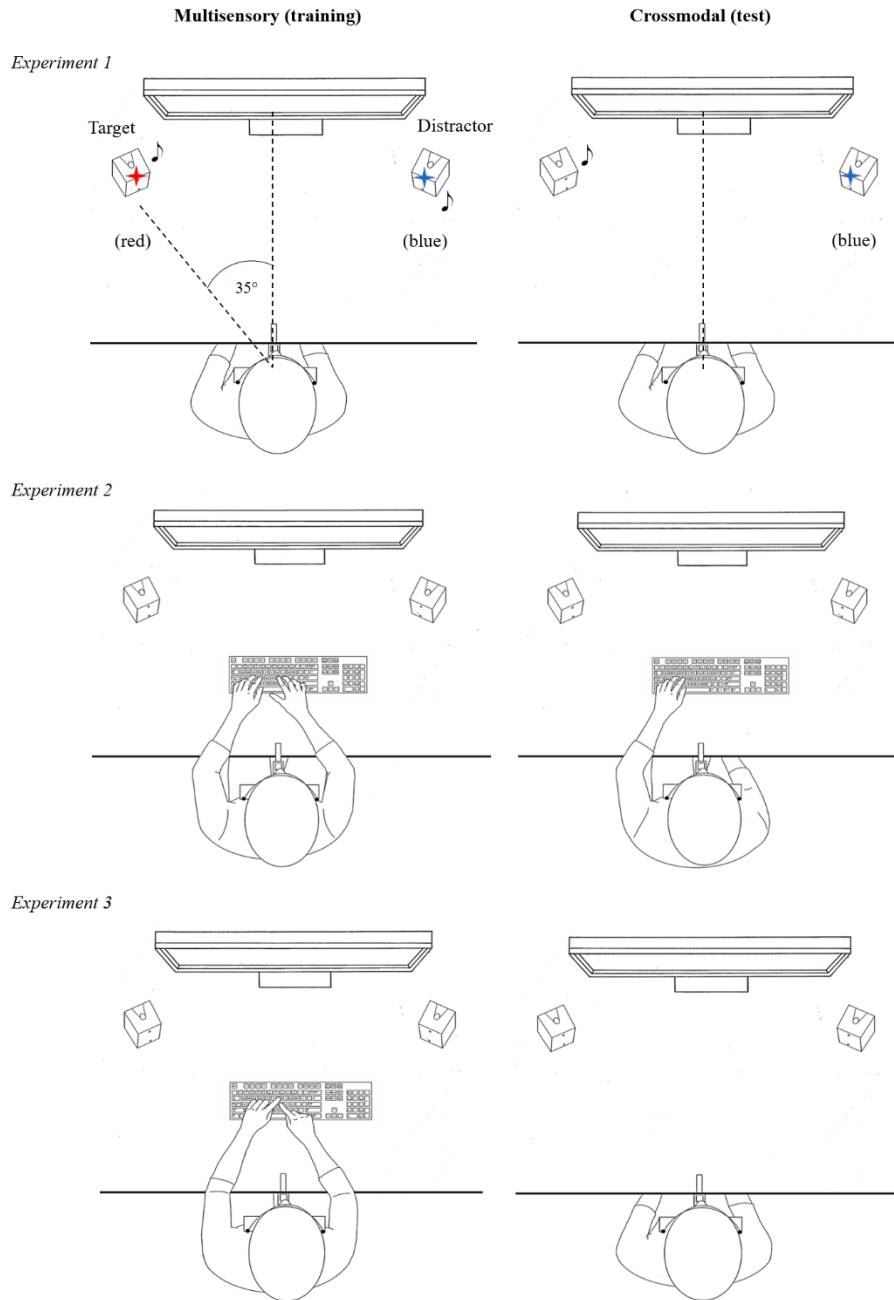


Figure 2. Experimental set-up and response effectors in Experiments 1-3. In the first part of the experiment, the cubes presented multisensory stimuli (left) whereas crossmodal stimuli were presented in the second part of the study (right). The identity of the visual feature is shown in parentheses. Note that this is an example of an incongruent trial in **both tasks as a high pitch is mapped with a red light (left target) and blue light/low pitch are incongruent in this case**. In Experiment 2, manual responses were used and in the multisensory part, four fingers were used for four targets (instead of two). In Experiment 3, any response compatibility

between the tasks was eliminated by changing the response effector from multisensory (hand) to crossmodal task (feet). Note that the feet in Experiments 1 and 3 were left out of this scheme.

Procedure. The participants were tested individually in a soundproof and dimly-lit experimental chamber. They sat 70 cm from the display with their chin resting on a chinrest. They were instructed to fixate the fixation cross in the center of the computer screen throughout each experimental block. The participant's feet were placed on footpedals, and their task was to identify the target via foot-pedal responses. As we intended to measure the aftereffects of multisensory processing, participants worked through two subsequent tasks, the multisensory and crossmodal flanker task.

In the first part of the study, the participants performed the multisensory flanker task (Jensen et al., 2019). Specifically, they had to respond to two different multisensory target stimuli, a red light combined with a high-pitched tone and a blue light combined with a low-pitched tone. The stimulus-response-mapping was balanced across participants (right vs. left footpedal response). The participants were instructed to respond to the target stimuli that were presented from one cube (i.e., the target cube) as rapidly and accurately as possible while at the same time ignoring the distractor stimuli presented from the other cube. In order to ensure that the participants processed both target features, and did not just respond to one of the sensory features, 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 cube. The participants were instructed not to respond to these stimuli, thus turning their task into a go-/no-go-task. These catch trials accounted for 20% of all trials. In the first part of the study, the distractor stimuli consisted of a combination of a visual and an auditory feature, selected from three possible features in either modality (two target features and one response-neutral feature each). A green light and a medium pitched tone served as neutral features (not mapped on to any response).

These neutral features were implemented in order to help emphasize the irrelevance of the distractor cube to participants. The visual and auditory 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 (see Figure 3; see also Jensen et al., 2019). The nine different distractor conditions (e.g., congruent-congruent) were presented with the same amount of trials (32 trials) throughout the experimental phase, respectively. The target and distractor side (left, right) was manipulated within-participants in two blocks of 180 trials each while the order (left vs. right first) was counterbalanced across participants. An additional break was offered in the middle of each block of trials.

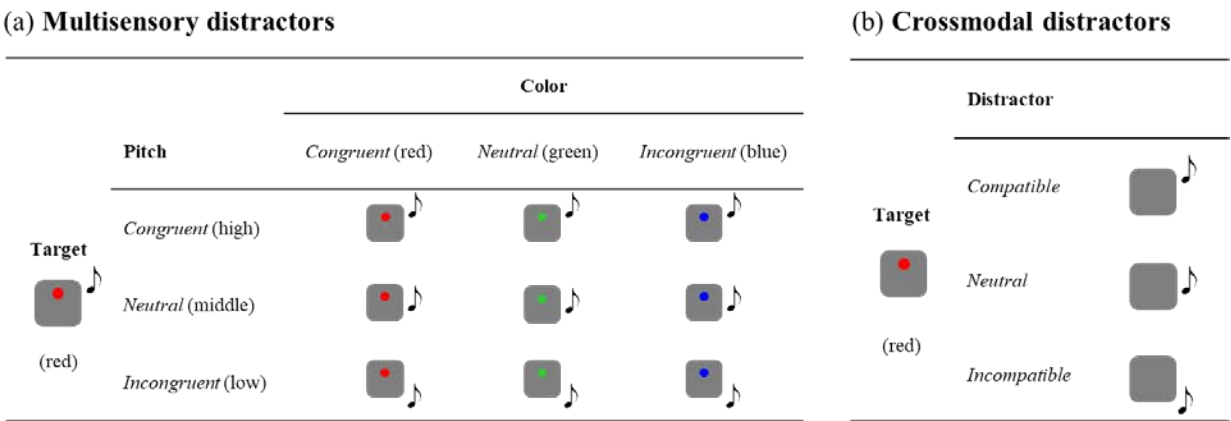


Figure 3. Distractor conditions in the multisensory (a) and crossmodal (b) flanker task for a red-light-high-pitch (red light) target as an example. Note that the multisensory distractors vary as a function of (visual and auditory) congruency while the crossmodal distractors vary as a function of compatibility.

In the second, crossmodal, part of the study, the participants were now presented with crossmodal target and distractor stimuli. The target modality (visual or auditory) was balanced across-participants. The crossmodal distractor could either be compatible, neutral, or

incompatible with respect to the current target stimulus based on the composition of the target feature in the multisensory task (see Figure 3). For instance, a red light was compatible with a high-pitched tone as this particular stimulus combination had previously been a multisensory target. Each compatibility condition was conducted with an equal number of trials (24 trials throughout the experimental phase). Once again, the target and distractor sides (left, right) were manipulated within-participants in two blocks of 72 trials while the order (left vs. right first) was counterbalanced across participants and an additional break was offered in the middle of each block. The targets started at the same side as in the beginning of the multisensory task, meaning that there was another switch of target side between the two parts of the experiment.

The exact sequence of stimuli on each trial was the same for both parts of the study: A cross was presented in the center of the screen serving as the fixation point and initiating 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 distractor¹. Both of the stimuli were then presented for 1000 ms and the participants had to respond within 1500 ms of the onset of the distractor. Feedback was only provided 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 was presented from both cubes. The stimulus lasted for 500 ms and was followed by a Response-Stimulus-Interval (RSI) of 500 ms.

Instructions were provided on the screen prior to the experiment. The participants were instructed to always respond as fast and as accurate as possible. Before the start of the experiment, they were encouraged to ask the experimenter (who was in the control room next

¹ The onset of the distractor occurred 100 ms before the onset of the target in order to maximize any congruency effects that were observed. Note that congruency effects, unisensory as well as crossmodal, have been shown to increase when the distractor precedes the target, often peaking at an SOA of about 100 ms (e.g., Chen & Spence, 2013, 2018; Flowers & Wilcox, 1982; Shore, Barnes, & Spence, 2006).

to the experimental room during the whole experiment), if anything was unclear **during the experiment**. The multisensory part started with a learning phase for the stimulus-response mapping with eight trials and a first multisensory practice phase with 16 trials. In this practice phase, the target and catch trials both accounted for 50% of all trials and were presented on the target cube while no stimuli were presented from the distractor cube. The second multisensory practice phase with 54 trials was exactly the same as the multisensory experimental phase (20% catch trials) except for the fact that feedback was provided on correct trials as well. The crossmodal part of the study started with a practice phase consisting of 18 trials, which was exactly the same as the crossmodal experimental phase except (once again) for the feedback on correct trials. In total, the experiment lasted for about 50 minutes.

Design. The crossmodal part of the experiment measuring aftereffects consisted of a repeated measures experimental design with the variable Compatibility (compatible, neutral, and incompatible). The modality of the target was manipulated between-participants but was left out of the analyses as there were no effects. The multisensory part of the study consisted of a 3×3 repeated measures 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.

Results

For the analysis of the RT data, only those trials in which the participants responded correctly to the target were considered. All of the trials in which the RT fell 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, 5.7% (4.7% in the crossmodal part; 6.2% in the multisensory part) of the trials

were excluded from the analysis due to these restrictions. As we were mainly interested in the results of the crossmodal task (the test phase), the result section starts with the analyses of the crossmodal trials. Table 1 depicts the mean RT and error data for the crossmodal trials.

Table 1. Mean RT (in milliseconds; error rates in % in parentheses) as a function of compatibility in Experiments 1-3.

Experiment	Crossmodal distractor feature		
	<i>Compatible</i>	<i>Neutral</i>	<i>Incompatible</i>
1	592 (2.9)	603 (2.7)	604 (4.3)
2	572 (3.5)	581 (2.7)	577 (3.1)
3	602 (6.0)	612 (6.2)	618 (5.5)

Crossmodal task. To measure aftereffects, a multivariate analysis of variance (MANOVA) with Compatibility (compatible vs. neutral vs. incompatible) as the independent variable and Pillai's trace as the criterion was conducted on the crossmodal trials, with the mean RT as the dependent variable. The MANOVA revealed a significant main effect of Compatibility, $F(2, 25) = 3.91$, $p = .033$, $\eta_p^2 = .24$; that is, the participants responded significantly faster when the crossmodal distractor feature was congruent than when it was incongruent.² Further analyses revealed that there was a significant difference between

² An additional analysis with Target Modality as the between-participants variable was conducted in order to check for modality differences. In this 2 (Target Modality: visual vs. auditory) \times 3 (Compatibility: compatible vs. neutral vs. incompatible) MANOVA with Pillai's trace as the criterion, an effect of Compatibility was still obtained, $F(2, 24) = 3.74$, $p = .039$, $\eta_p^2 = .24$, but no other effects were significant, all $ps > .107$. Thus, the compatibility effect was independent of modality. In the following experiments, modality was left out of all analyses as there was never any effect, all $ps > .197$, and the size of the compatibility effects was not influenced by modality.

congruent and incongruent trials, $t(26) = 2.41, p = .023$ (see Figure 4), as well as congruent and neutral trials, $t(26) = 2.69, p = .015$, but not between neutral and incongruent trials, $t(26) = 0.11, p = .915$.

For the error rates, the same MANOVA was conducted on the crossmodal trials as for the RT data. There was no significant effect of Compatibility, $F(2, 25) = 1.50, p = .243, \eta_p^2 = .11$.

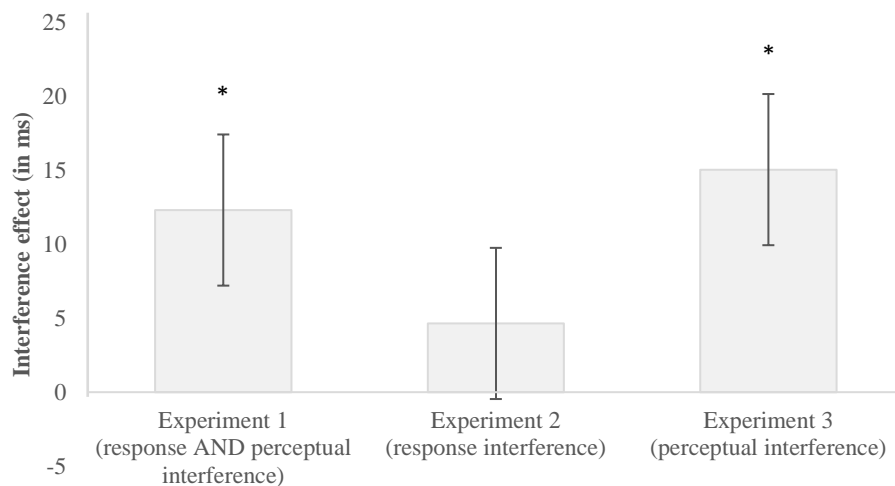


Figure 4. Interference effects (mean difference of incompatible trials minus compatible trials) in the RT data in crossmodal flanker tasks of Experiments 1-3. Asterisks indicate significant effects ($p < .05$). In Experiment 2, the effect refers to response interference, whereas in Experiment 3, the effect refers to perceptual interference instead. Please note that although the response modality differed between experiments, baseline RTs were comparable (see Table 1).

Multisensory task. Additionally, we analyzed the multisensory trials as in our previous experiments in order to determine whether multisensory congruency effects occurred in the first place (and to replicate our previous findings). We first checked whether participants completed

the multisensory task properly by examining performance on the catch trials. On average, participants responded correctly (meaning that they did not react) to 90% of these trials (range from 62-100%) showing that they processed both of the features on the target cube. For the target trials, a 3×3 (Visual Distractor Feature Congruency: congruent vs. neutral vs. incongruent) MANOVA was conducted with Pillai's trace as the criterion and mean RT as the dependent variable (see Table A1 in the Appendix for mean RT and error data). The MANOVA revealed significant main effects of Visual Distractor Feature Congruency, $F(2, 25) = 13.71, p < .001, \eta_p^2 = .52$, and Auditory Distractor Feature Congruency $F(2, 25) = 40.59, p < .001, \eta_p^2 = .77$; that is, the participants responded most rapidly when the target and distractor feature were congruent and slowest when they were incongruent. Furthermore, there was a significant interaction between these two variables, $F(4, 23) = 3.62, p = .020, \eta_p^2 = .39$, indicating that the distractor features influenced each other.

For the error rates, the same 3×3 MANOVA was conducted as for the RT data and no significant effects were found, all $ps > .090$.

Discussion

Experiment 1 was designed to investigate whether the association of two features, in different sensory modalities, that people learned and responded to in a multisensory task could elicit a lasting interference effect that could be measured in a following crossmodal flanker task. In the first part of the experiment, participants responded to audiovisual combinations of features that were flanked on one side by audiovisual distractors. The two distractor features were either congruent, neutral, or incongruent with respect to the corresponding target feature. In the second part of the experiment, the same features were used as crossmodal stimuli in order

to assess the aftereffects of multisensory associative learning. The two target features of one modality were presented on the target cube while the other cube presented a compatible, neutral, or incompatible feature in the other modality. Compatibility was defined by the target feature combination that had been introduced in the preceding multisensory task. The results revealed an interference effect of the crossmodal feature. After matching a visual and an auditory feature in a multisensory selection task, one feature facilitated the response to the other feature and hence interfered with the processing of the complementary crossmodal feature. This result indicates that the associations between the crossmodal features that were established during the multisensory part of experiment lasted well beyond the participant's completion of the task.

However, the crossmodal interference effect documented in Experiment 1 can be explained in terms of both response and perceptual associations (see Figure 1). In the multisensory task, the target features were presented together in 80% of the trials (as catch trials accounted for the remaining 20% of the trials). Accordingly, the features could be predominantly associated with each other and interference would be induced at the perceptual level. Alternatively, both features might be rather independently linked to the same motor response and therefore induce interference. In the case of perceptual interference, the effect should also be seen when the response given to the multisensory combination in the first part of the study is different from the response given to the unisensory stimulus in the crossmodal part of the study. Furthermore, crossmodal interference should be eliminated when the features in the multisensory part of the study are no longer associated with each other anymore (i.e., one feature is not more frequently presented with another). However, when response association is indeed the main factor underlying the effect, there should still be an effect when there are no feature associations, but not when the response changes between tasks. Thus, we conducted two further experiments in which either of these constraints were implemented in order to

investigate the processing level at which the crossmodal interference effects induced by multisensory selection take place.

Experiment 2

In Experiment 2, the aim was to investigate whether a crossmodal effect could be elicited at the response level without any **specific** perceptual associations between the features. We changed the multisensory task of the experiment by varying the target features orthogonally to each other, so that each of the target features in one modality was presented equally often with the two features in the other modality (see Figure 5). Namely, the former catch stimuli now served as targets as well and accounted for 50% of all trials. Accordingly, all targets were presented equally often and participants had to respond to four multisensory stimuli instead of just two. Critically, response compatibility in the crossmodal part of the study was created by choosing two out of the four possible response keys used in the multisensory task. The crossmodal feature was compatible when it has been part of the feature combination that was assigned to the response key that is required for the current target. By introducing this manipulation, **the specific** perceptual associations between the features, **which were induced in Experiment 1**, were removed **as feature associations were counterbalanced**. Any crossmodal interference effects that were documented after this modified multisensory task can now only be explained in terms of response associations.

Method

Participants. Thirty-two students from the University of Trier took part in the study in return of course credit. Two students were excluded due to extremely high error rates (these participants were outliers with respect to error rates of the sample, Mean = 15%, SD = 6%,

outlier with $> 37\%$ errors). A total of 30 participants (6 male) aged from 18 to 31 years (Mean = 22 years, $SD = 3$) were used for analyses. All of the participants had normal or corrected-to-normal vision and audition.

Apparatus, materials, design and procedure. The apparatus and materials used were exactly the same as in Experiment 1 except for the fact that keys on the keyboard were used for responding instead of foot pedals throughout the whole experiment. Design, procedure, trial number and trial sequence were exactly the same as in Experiment 1 while the task was changed slightly. In the multisensory part, there were no more catch trials with reversed feature assignment as participants now had to respond to these trials with a key-press instead of withholding their response. Thus, participants responded to four different targets with four different keys. We used the keys “d”, “f”, “j”, and “k” as keys 1 to 4, so that the keys were arranged in one line on the keyboard. Participants used the middle and index fingers of both hands in order to press the keys. Stimulus response mapping was balanced across-participants in the way that two complementary targets (e.g., the two original targets from Experiment 1) were assigned to one hand (see Figure 5 for an example of one possible mapping). In the crossmodal part, two out of these four keys (either the left two or the right two) were used for responding. Thus, the participants only used one hand for responding. The choice of these two keys accordingly defined the response stimulus compatibility (see Figure 5). For instance, when the two original targets (red light high pitch, blue light low pitch) in the multisensory part were assigned to key 1 and 2 (and therefore to the left hand) and participants used these two keys in the crossmodal part, then red light and high pitch (and blue light and low pitch) were response compatible to each other as both features had been assigned to the same key in part one. Thus, note that the variable Compatibility refers to response compatibility in this case.

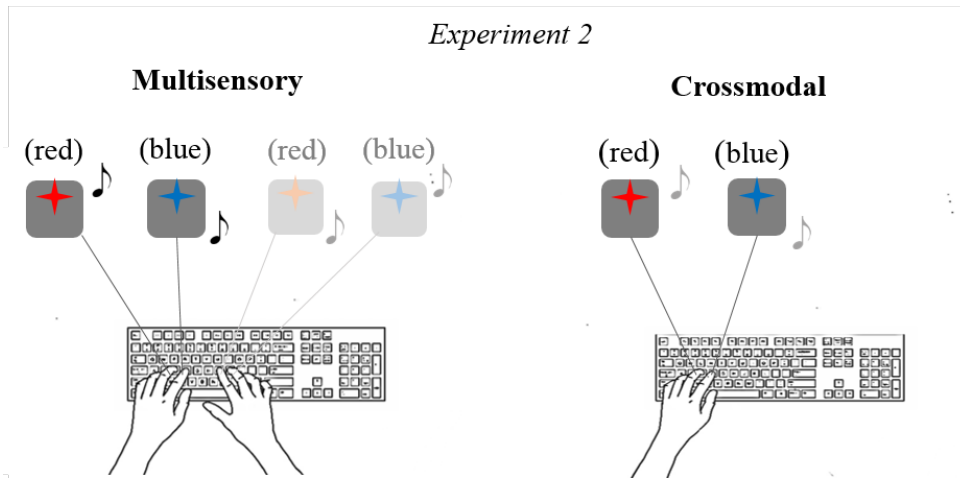


Figure 5. Example of the stimulus response mapping in Experiment 2 for the multisensory and visual crossmodal task. The response compatibility between the features is determined by the hand used in the crossmodal task (left hand in this example). The shaded features in the crossmodal task (right) represent the compatible crossmodal distractor feature. The identity of the visual feature is shown in parentheses.

Results

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, 15% of the trials were excluded from the analysis due to these restrictions. Table 1 depicts the mean RT and error data of the crossmodal trials.

Crossmodal task. A MANOVA with the independent variable Compatibility (compatible vs. neutral vs. incompatible) and Pillai's trace as the criterion was conducted on the crossmodal trials, with mean RT as the dependent variable. In contrast to Experiment 1, the MANOVA revealed no significant main effect of Compatibility, $F(2, 28) = 1.167, p = .326, \eta_p^2$

= .08, suggesting that the crossmodal distractor feature did not influence target processing (see Figure 4). Additional Bayesian analysis comparing the congruent and incongruent condition within a t-test revealed a *Bayes Factor* (BF_{01}) = 2.83, indicating that this null-result is about three times more likely than the presence of an effect. As the effect is thought to be induced by the association that was learnt during the multisensory part of the study, it should be strongest right at the start of the crossmodal experimental phase. We therefore additionally looked at the compatibility effect in the first block of the crossmodal part only. However, this was still not significant, $F(2, 28) = .873$, $p = .429$, $\eta_p^2 = .06$, with a $BF_{01} = 1.78$ for the comparison of the congruent and incongruent conditions.

For the error rates, we conducted the same MANOVA on the crossmodal trials as for the RT data. There was no significant effect of Compatibility, neither in total, $F(2, 28) = 1.08$, $p = .353$, $\eta_p^2 = .07$, nor for the first block, $F(2, 28) = .412$, $p = .666$, $\eta_p^2 = .03$.

Multisensory task. Once again, an additional 3 (Visual Distractor Feature Congruency: congruent vs. neutral vs. incongruent) \times 3 (Auditory Distractor Feature Congruency: congruent vs. neutral vs. incongruent) MANOVA was conducted on the multisensory trials, with mean RT as the dependent variable (see Table A1 in the Appendix for mean RTs and error data). There was a significant main effect of Auditory Distractor Feature Congruency, $F(2, 28) = 12.28$, $p < .001$, $\eta_p^2 = .47$; this is, participants responded most rapidly when the auditory distractor feature was congruent and slowest when it was incongruent. However, there was no effect of the visual feature, nor any interaction, all $ps > .153$. As it seemed a little odd that there was no effect of the visual feature in the RT data (as this is typically pretty strong; see Jensen et al., 2019), we thought it prudent to test whether the visual feature did not produce any effect at all. As the neutral condition is rather unpredictable (somewhere between congruent and incongruent) and therefore less reliable (see Jensen et al., 2019; Jonides & Mack, 1984), we ran an additional 2 (Visual Distractor Feature Congruency: congruent vs. incongruent) \times 2

(Auditory Distractor Feature Congruency: congruent vs. incongruent) MANOVA without the neutral condition. The analysis revealed a significant main effect of the visual, $F(1, 29) = 5.79$, $p = .023$, $\eta_p^2 = .17$, and auditory features, $F(1, 29) = 22.78$, $p < .001$, $\eta_p^2 = .44$, in the expected directions. Once again, there was no interaction, $F(1, 29) = 1.103$, $p = .302$, $\eta_p^2 = .04$.

For the error rates, the 3×3 MANOVA revealed a main effect of Auditory Distractor Feature Congruency, $F(2, 28) = 9.20$, $p = .001$, $\eta_p^2 = .40$, but a non-significant effect of Visual Distractor Feature Congruency, $F(2, 28) = 2.89$, $p = .07$, $\eta_p^2 = .17$. There was no interaction between these two variables, $F(4, 26) = 1.876$, $p = .145$, $\eta_p^2 = .22$. The 2×2 MANOVA on the multisensory trials revealed a significant main effect of the auditory, $F(1, 29) = 17.26$, $p < .001$, $\eta_p^2 = .37$, as well as the visual feature, $F(1, 29) = 8.07$, $p = .008$, $\eta_p^2 = .22$, and, once again, no interaction, $F(1, 29) = .335$, $p = .576$, $\eta_p^2 = .01$.

Discussion

In Experiment 2, we investigated whether a crossmodal interference effect can be induced by a multisensory flanker task at the response level without **specific** feature associations. We changed the design of our multisensory flanker task by including the former catch trials in the original task as target trials and presenting all targets equally often. Accordingly, every feature from the one modality was presented equally often with the features of the other modality. Thus, there should not be any preferences for specific feature combinations. Response compatibility was established by choosing two of the four possible keys of the multisensory task for responding in the crossmodal task. The crossmodal feature that belonged to the combination that was assigned to one specific key in the multisensory part of the study was compatible in the crossmodal part when the other feature of that combination was the current target. With this modulation, we did not find a crossmodal flanker effect after

a multisensory selection task, not even if we just looked at the first block of trials in the crossmodal congruency task. Although congruency effects emerged in the modified version of the multisensory flanker task, there was no subsequent interference based on response compatibility in the crossmodal task. These results therefore are a first hint to the assumption that the interference at the perceptual level (see Figure 1) continues to have an effect after the multisensory selection situation, whereas interference at response level is not persistent.

Regarding the multisensory task, we did not find a significant interaction between the two distractor features, as had been observed in Experiment 1. The interaction is seen as an indicator that features are not processed independently from one another. Further, this interaction is known to be influenced by the availability of attentional resources (Jensen et al., 2019). A reason for this diverging finding might be a greater difficulty of the modified multisensory task as compared to the original one. Target processing might need more attentional resources and the amount available for the distractor decreases, thus leading to feature-based independent rather than interactive processing.

Experiment 3

To further corroborate the finding that perceptual feature associations rather than response associations are crucial for the crossmodal distractor interference induced by multisensory selection, a third experiment was conducted, in which we again used the original multisensory task as in Experiment 1 but manipulated response compatibility. We changed the response from the multisensory to the crossmodal part of the study. In the crossmodal part, participants continued to react with left and right footpedal responses whereas, in the multisensory part of the study, the reaction was now an up-down manual response instead. With this, we test whether a crossmodal interference effect after a multisensory flanker task is, in fact, elicited on the perceptual level.

Method

Participants. Twenty-six students³ from the University of Trier took part in the study in return of course credit or were paid 6€ (~ 7.5\$). Two students were excluded due to their extremely high error rates (these participants were outliers with respect to the error rates of the sample, Mean = 15%, SD = 6%, outlier with > 53% errors). A total of 24 participants (6 male) aged from 20 to 36 years (Mean = 25, SD = 4) were used for analyses. All of the participants had normal or corrected-to-normal vision and audition.

Apparatus, materials, design, and procedure. The apparatus and materials used were exactly the same as in Experiment 1 except for one major change. In the multisensory part of the experiment, participants used the keyboard instead of foot pedals to give their response. The key with the letter ‘u’ was marked as key 1 while the key with the letter ‘j’ was marked as key 2, resulting in an up-down-elevation of the keys. The design, procedure, task, and trial sequence were exactly the same as in Experiment 1. Note that the variable Compatibility refers to perceptual interference in this case. Response-mapping was not balanced in this part of the experiment as the auditory feature of the multisensory target (high vs. low pitch) is matching this elevation (see e.g., Spence, 2011) and a reversed assignment would be confusing.

Results

For the analysis of the RT data, only those trials in which the participants responded correctly to the target were considered. All of the trials in which the RT was shorter than 200

³ Note that the number of participants was reduced in this sample. However, with the effect size of the (non-significant) crossmodal compatibility effect in Experiment 2 of $\eta_p^2 = .08$ in mind, and given an alpha-level of .05 and a desired power of at least $1 - \beta > .80$ in a repeated measure MANOVA with three measures, we calculated a new minimum of 23 participants, which is fulfilled (power analyses were run with G-power 3.1.9.2; Erdfelder et al., 1996, 2007).

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, 7.2% of the trials were excluded from the analysis due to these restrictions. Table 1 depicts the mean RT and error data for the crossmodal trials.

Crossmodal task. A MANOVA with the independent variable Compatibility (compatible vs. neutral vs. incompatible) and Pillai's trace as the criterion was conducted on the crossmodal trials, with mean RT as the dependent variable. The main effect of Compatibility approached significance, $F(2, 22) = 3.19, p = .055, \eta_p^2 = .23$, with participants responding more rapidly on compatible than on incompatible trials. As in Experiment 2, we looked at the effect in the first block of the crossmodal part only, which turned out to be significant, $F(2, 22) = 5.49, p = .012, \eta_p^2 = .33$. Additionally, the crossmodal interference effect was computed as the difference between incompatible and compatible trials and tested against zero (see Figure 4). In this case, the overall effect was significant, $t(23) = 2.47, p = .021$, with a $BF_{01} = 0.20$. We also conducted pair-wise comparison with the neutral condition, which revealed a marginal significant difference between congruent and neutral trials, $t(23) = 2.03, p = .05$, which was significant in the first block, $t(23) = 2.53, p = .019$, but no significant difference between neutral and incongruent trials, neither overall, $t(23) = 0.97, p = .343$, nor in the first block, $t(23) = 0.31, p = .759$.

When the same MANOVA was conducted on the error rates for the crossmodal trials, as for the RT data, there was no significant effect of Compatibility, neither overall, $F(2, 22) = .385, p = .685, \eta_p^2 = .03$, nor when considering just the first block, $F(2, 22) = .309, p = .738, \eta_p^2 = .03$.

Multisensory task. As for Experiment 1, we first checked whether the participants worked through the multisensory task properly by examining the catch trials. On average, the

participants responded correctly in 80% of these trials (range from 46-99%) showing that by-and-large they had processed both features on the relevant cube in Experiment 3 as well. For the target trials, 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 (see Table A1 in the Appendix for mean RT and error data). The effects of Visual, $F(2, 22) = 29.09, p < .001, \eta_p^2 = .73$, as well as Auditory Distractor Feature Congruency, $F(2, 22) = 19.85, p < .001, \eta_p^2 = .64$, were significant, with participants responding most rapidly on congruent trials and slowest on incongruent trials. Further, there was a significant interaction between the two features, $F(4, 20) = 3.76, p .019, \eta_p^2 = .43$. For the error rates, the 3×3 MANOVA showed no significant effects, all $ps > .596$.

Discussion

In Experiment 3, we investigated whether a crossmodal interference effect would be induced after a multisensory selection task in which there was no response compatibility between the two tasks. We modified the multisensory task by using manual instead of footpedal responses as well as switching from a left-right to an up-down response to maximize the difference between multisensory and crossmodal task. With this, any possibility of response interference is prevented. Overall, the effect of compatibility (neutral trials included) just failed significance. As interference might be largest directly after the multisensory task, the first block of the crossmodal trials was analyzed separately. Here, a significant effect was found. Additionally, the difference between incompatible and compatible trials was calculated as a crossmodal interference effect and reached significance over all trials. Taken together, these findings prove that crossmodal interference was successfully induced. As any response compatibility between the multisensory and the crossmodal task was eliminated, this effect was

Multisensory Selection

- 1 solely based on the associations between the features in the multisensory task. This result
- 2 supports the finding from Experiment 2 that, in multisensory selection, the perceptual
- 3 associations between the features are learned more strongly than are the associations with
- 4 responses. Subsequent interference in the crossmodal task thus works predominantly at the
- 5 perceptual level.

General Discussion

The general aim of the series of experiments reported here was to investigate the processing level at which multisensory associations in a multisensory flanker task induce subsequent crossmodal distractor interference. A two-part experimental design was implemented in which a multisensory interference task was followed by a crossmodal interference task. By means of such an arrangement, the effects measured in the crossmodal task provide information concerning the processing of the stimuli in the previous multisensory task. As we were particularly interested in the processing level of multisensory associations and the subsequent crossmodal interference, response compatibility between the two tasks and the relation between the features in the multisensory task, were manipulated within three experiments. Experiment 1 examined the existence of crossmodal interference after multisensory selection without distinguishing the processing level. The crossmodal interference effect that was obtained can be explained in terms of both response and feature associations. On the one hand, there was a contingency between the features in the multisensory task – two target features from different modalities were frequently presented together (in 80% of the trials) – allowing for feature associations to be established. On the other hand, responses were the same in both tasks, promoting response associations as well. In Experiment 2, the feature associations were eliminated from the multisensory task in order to look for any crossmodal interference that might be based solely on response associations. However, no significant effect was obtained, thus suggesting that response associations are not the driving force behind the crossmodal distractor interference. In Experiment 3, response associations were eliminated while the feature associations remained. Here, a significant crossmodal interference effect was present. This indicates that the crossmodal interference depends to a greater extent on feature associations learned in the foregoing multisensory task than on response associations.

Experiments 1 and 3 proved the existence of crossmodal interference following multisensory selection with arbitrary stimulus combinations. This result is in line with previous studies that have demonstrated an influence of multisensory associations on subsequent information processing, both, with pre-existing semantic associations (Fiebelkorn et al., 2010; Lehmann & Murray, 2005; Molholm et al., 2004, 2007; Murray et al., 2004, 2005) as well as with those associations that have been established during the course of the experiment (Baier et al., 2006; Ernst, 2007; Zangenehpour & Zatorre, 2010). Similar to these latter studies, we were able to induce a crossmodal **interference** effect by combining previously unrelated stimuli (or as in Ernst, 2007, stimulus dimensions) in a learning period (here, the multisensory flanker task). **Thus, the relatively short presentation of newly learned multisensory stimulus combinations in one training episode led to the emergence of associations that persist (at least shortly) and affect subsequent processing.**

While we rather focus on the effect of the associations that are learned and responded to, one might also if there was an ‘negative’ effect of the combinations that are not responded to. In the crossmodal task, incongruent combinations, e.g., high pitch and blue light, reflect the former catch trials of the multisensory task (or a combination that is mapped to a different response in Experiment 2). Thus, it could be possible that these combinations are somehow inhibited during the multisensory task and accordingly lead to greater interference. The results, however, show that this is not very likely, as the compatibility effects found in Experiments 1 and 3 primarily go back to facilitation. There was a significant difference between congruent and neutral trials, but not between neutral and incongruent trials. Therefore, ‘negative associations’ do not seem to produce a considerable effect in the present study.

In Experiments 2 and 3, we examined the level of association in the multisensory flanker task that is necessary to induce subsequent crossmodal interference. In Experiment 2, there was no crossmodal interference effect when the direct associations between the features were

removed, and interference was only possible at the response level. In the multisensory flanker task of Experiment 2, the participants had to respond to four targets constituting orthogonal combinations of the four original target features (two visual, two auditory) from Experiment 1. Accordingly, there were no prominent crossmodal combinations that would have enabled feature association without a response link. In the crossmodal task, two out of four keys were used and defined the response compatibility of the features that were associated with these keys. The results revealed that response compatibility alone was insufficient to establish measurable interference. In Experiment 3, the original multisensory flanker task of Experiment 1 was used with two features being presented together in 80% of the trials. Critically, the response modality had been changed and was no longer compatible between the tasks. Hence, only with perceptual feature association, crossmodal interference was induced successfully. Taken together, these experiments suggest that only those associations between features, rather than between feature and response, extend, or last, beyond the multisensory task, enabling mutual impact of the features across modalities.

These results provide some important new insight into the processing of multisensory stimuli in selection situations. Looking at our sketch of mental processes (see Figure 1), it can be assumed that the link between the two individual features of the multisensory target (feature association) is stronger than the link between one feature and the response (response association). Our system does not seem to link the individual parts of an event independently of each other but takes task-instructions (as an example of a top-down process, see Wenke, Gaschle, & Nattkemper, 2007) and the statistical relation of feature combinations (as an example of a bottom-up process, see Wang & Theeuwes, 2018) into account (see also Awh, Belopolsky, & Theeuwes, 2012). When it comes to multisensory processing, the associations between the features, which are mainly indicated by frequent co-occurrence in this case, seem to be more important than response associations. Previous research also demonstrated that the

constant pairing of stimuli is a key factor for persistent effects. Logie, Brockmole, and Vandenbroucke (2009) demonstrated in a memory task that only whole object repetitions (repetition of specific combination of shape and color at a specific location) throughout the course of the experiment rather than single feature repetitions (repetition of one feature with a varying second feature) led to benefits in long-time learning. **Note, however, that in the present study, short-term learning is in the focus.** The results of the present study complement this finding by showing that a pairing with less than 100% contingency can also lead to reliable aftereffects.

Another interesting question indicated by the present results concerns the imperative need of an active response in the multisensory task for the emergence of perceptual associations. Even in Experiment 3, when response associations between the tasks were eliminated, there was a response, and probably response associations, within the multisensory task. While action and perception are closely connected as suggested by the Common Coding approach (Prinz, 1990) and the Theory of Event Coding (Hommel, Müsseler, Aschersleben, & Prinz, 2001; see Hommel, 2019, for an update), the key factor often is the imagined, seen or executed action effect. Thus, it would be interesting to test whether perceptual associations within multisensory stimuli can also be established, and produce measurable crossmodal interference, with passively perceived multisensory stimulus presentation.

The crossmodal interference effect that was documented in Experiment 3, thus, is only observed in the first block of trials when considering all compatibility levels (compatible, neutral, and incompatible). With all trials, a significant effect is found when considering the difference between incompatible and compatible trials (the crossmodal interference effect). In Experiment 1, an effect was found in all cases. This might indicate that the effect that was documented in Experiment 1 is somewhat more persistent than the effect documented in Experiment 3. In comparison to Experiment 1, the number of participants in Experiment 3 was

1 reduced, which might have led to the lack of significance when considering the effect with all
2 three compatibility levels. Concerning the arrangement, the only difference between
3 Experiments 1 and 3 was the response compatibility between the two experimental parts.
4 Consequently, it cannot be excluded that response associations have an effect after all. Yet, that
5 said, it has not been sufficiently pronounced to produce persistent interference between the
6 crossmodal features without feature associations (Experiment 2). This resembles previous
7 research that found larger effects of perceptual congruency than of pure response compatibility
8 in crossmodal settings (Wesslein, Spence, & Frings, 2014).

9

10 *Conclusions*

11 The present study investigated the influence of multisensory association learning on
12 subsequent interference processes. Participants performed a multisensory flanker task, in which
13 audiovisual targets and distractors were presented. Afterwards, interference effects of auditory
14 on visual features (and the other way around) were measured in a crossmodal flanker task.
15 Within three experiments, it was shown that crossmodal aftereffects were induced by the
16 associations between the features rather than response associations between features and
17 response created in the multisensory task. Crossmodal interference took place when the
18 response effectors changed from one task to the other, but not when the associations between
19 two specific features were removed in the multisensory task. Consequently, the integration of
20 perceptual features would appear to be what is driving aftereffects of multisensory processing.

1 *Open Practices Statement*

2 The data and codes for all experiments are available at PsychArchives
3 (<http://dx.doi.org/10.23668/psycharchives.2465>).

References

- Barenholtz, E., Lewkowicz, D. J., Davidson, M., & Kogelschatz, L. (2014). Categorical congruence facilitates multisensory associative learning. *Psychonomic Bulletin & Review*, 21, 1346-1352.
- Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, 114, 389-404.
- Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1554-1568.
- Chen, Y.-C., & Spence, C. (2013). The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. *Multisensory Research*, 26, 371-386.
- Chen, Y.-C., & Spence, C. (2018). Dissociating the time courses of the cross-modal semantic priming effects elicited by naturalistic sounds and spoken words. *Psychonomic Bulletin & Review*, 25, 1138-1146.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7:7, 1-14.

- 1 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
2 power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research*
3 *Methods*, 39, 175-191.
- 4 Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory
5 spread of attention: How much do learned associations matter? *Cerebral Cortex*, 20, 109-120.
- 6 Flowers, J. H., & Wilcox, N. (1982). The effect of flanking context on visual classification: The
7 joint contribution of interactions at different processing levels. *Perception & Psychophysics*,
8 32, 581-591.
- 9 Frings, C., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Hommel, B., Kiesel, A., Kunde,
10 W., Mayr, S., Moeller, B., Möller, M., Pfister, R. & Philipp, A. (in press). Merkmalsintegration
11 und Abruf als wichtige Prozesse der Handlungssteuerung – eine Paradigmen-übergreifende
12 Perspektive [Feature integration and retrieval as important processes of action control – a
13 general perspective across paradigms]. *Psychologische Rundschau*.
- 14 Frings, C., & Spence, C. (2010). Crossmodal congruency effects based on stimulus identity.
15 *Brain Research*, 1354, 113-122.
- 16 Guo, J., & Guo, A. (2005). Crossmodal interactions between olfactory and visual learning in
17 *Drosophila*. *Science*, 309, 307.
- 18 Hall, G. (1991). *Oxford psychology series, No. 18. Perceptual and associative learning*. New
19 York, NY: Oxford University Press.
- 20 Hall, G., & Honey, R. (2014). Perceptual and associative learning. In S. B. Klein & R. R.
21 Mowrer (Eds.), *Contemporary learning theories: Volume II: Instrumental conditioning theory*
22 *and the impact of biological constraints on learning* (pp. 117-148). New York, NY: Psychology
23 Press.

- 1 Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in*
2 *Cognitive Sciences*, 8, 494-500.
- 3 Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological*
4 *Research*, 73, 512-526.
- 5 Hommel, B. (2019). Theory of Event Coding (TEC) V2. 0: Representing and controlling
6 perception and action. *Attention, Perception, & Psychophysics*, 1-16.
- 7 Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding
8 (TEC): A framework for perception and action planning. *Behavioral and brain sciences*, 24(5),
9 849-878.
- 10 Jensen, A., Merz, S., Spence, C., & Frings, C. (2019). Overt spatial attention modulates
11 multisensory selection. *Journal of Experimental Psychology: Human Perception and*
12 *Performance*, 45, 174-188.
- 13 Jensen, A., Merz, S., Spence, C., & Frings, C. (submitted). *Interference of irrelevant*
14 *information in multisensory selection depends on attentional set*. Manuscript submitted for
15 publication.
- 16 Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological*
17 *Bulletin*, 96, 29-44.
- 18 Knoeferle, K., Knoeferle, P., Velasco, C., & Spence, C. (2016). Multisensory brand search:
19 How the meaning of sounds guides consumers' visual attention. *Journal of Experimental*
20 *Psychology: Applied*, 22, 196-210.
- 21 Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object
22 discrimination. *Cognitive Brain Research*, 24, 326-334.

- 1 Mackintosh, N. J. (1965). Selective attention in animal discrimination learning. *Psychological*
- 2 *Bulletin*, 64, 124-150.
- 3 McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I.
- 4 Latent inhibition and perceptual learning. *Animal Learning & Behavior*, 28, 211-246.
- 5 Merz, S., Jensen, A., Spence, C., & Frings, C. (2019). *Multisensory distractor processing is*
- 6 *modulated by spatial attention*. Manuscript submitted for publication.
- 7 Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is
- 8 multisensory: Co-activation of an object's representations in ignored sensory modalities.
- 9 *European Journal of Neuroscience*, 26, 449-509.
- 10 Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object
- 11 recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, 14, 452-465.
- 12 Murray, M. M., Foxe, J. J., & Wylie, G. R. (2005). The brain uses single-trial multisensory
- 13 memories to discriminate without awareness. *NeuroImage*, 27, 473-478.
- 14 Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory
- 15 processes: A balancing act across the lifespan. *Trends in Neurosciences*, 39, 567-579.
- 16 Murray, M. M., Michel, C., M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino,
- 17 S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed
- 18 by electrical neuroimaging. *NeuroImage*, 21, 125-135.
- 19 Neumann, O., van der Heijden, A. H. C., & Allport, D. A. (1986). Visual selective attention:
- 20 Introductory remarks. *Psychological Research*, 48, 185-188.
- 21 Pavlov, I. P. (1927). *Conditioned reflexes*. London, UK: Oxford University Press.
- 22 Prinz, W. (1990). A common coding approach to perception and action. In *Relationships*
- 23 *between perception and action* (pp. 167-201). Springer, Berlin, Heidelberg.

- 1 Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the
2 effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.),
3 *Classical conditioning II: Current research and theory* (pp. 64-99). New York, NY: Appleton-
4 Century-Crofts.
- 5 Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*,
6 *16*, 1422-1427.
- 7 Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive*
8 *Sciences*, *12*, 411-417.
- 9 Shore, D. I., Barnes, M. E., & Spence, C. (2006). Temporal aspects of the visuotactile
10 congruency effect. *Neuroscience Letters*, *392*, 96-100.
- 11 Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science & Technology*, *28*,
12 61-70.
- 13 Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, &*
14 *Psychophysics*, *73*, 971-995.
- 15 Spence, C., & Frings, C. (in press). Multisensory feature integration in (and out) of the focus of
16 spatial attention. *Attention, Perception & Psychophysics*.
- 17 Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal
18 distractor congruency effects. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 148-169.
- 19 Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The*
20 *Quarterly Journal of Experimental Psychology*, *40*, 201-237.
- 21 Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison Wesley.
- 22 Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice
23 recognition. *PLoS Biology*, *4*, 1809-1820.

- 1 Wagner, A. R. (1969a). Stimulus validity and stimulus selection. In W. K. Honig & N. J.
2 Mackintosh (Eds.), *Fundamental issues in associative learning* (pp. 90-122). Halifax, NS:
3 Dalhousie University Press.
- 4 Wagner, A. R. (1969b). Stimulus selection and a “modified continuity theory”. In G. H. Bower
5 & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3, pp. 1-41). New York,
6 NY: Academic Press.
- 7 Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz & R.
8 Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp.
9 39-56). Hillsdale, NJ: Lawrence Erlbaum.
- 10 Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory
11 discrepancy. *Psychological Bulletin*, 88, 638-667.
- 12 Wesslein. A. K., Spence, C., & Frings, C. (2014). When vision influences the invisible
13 distractor: Tactile response compatibility effects require vision. *Journal of Experimental*
14 *Psychology: Human Perception and Performance*, 40, 763-774.
- 15 Zangenehpour, S., & Zatorre, R. J. (2010). Crossmodal recruitment of primary visual cortex
16 following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, 48, 591-600.
- 17 Zhang, H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of
18 stimulus-stimulus and stimulus-response compatibility. *Cognitive Psychology*, 38, 386-432.
- 19 Zmigrod, S., & Hommel, B. (2011). The relationship between feature binding and
20 consciousness: Evidence from asynchronous multi-modal stimuli. *Consciousness and*
21 *Cognition*, 20, 586-593.

1

Appendix

2

Table A1. Mean RT (in milliseconds; error rates in %, in parentheses) as a function of visual

3

and auditory distractor feature congruency in Experiments 1-3.

	Visual distractor feature			
Auditory distractor feature	Congruent	Neutral	Incongruent	Mean
Experiment 1				
Congruent	655 (2.8)	682 (3.1)	696 (3.1)	692 (3.0)
Neutral	698 (4.6)	705 (3.0)	727 (3.7)	702 (3.8)
Incongruent	724 (4.3)	719 (3.5)	727 (4.4)	717 (4.1)
Mean	678 (3.9)	710 (3.2)	723 (3.7)	
Experiment 2				
Congruent	872 (14.9)	880 (17.1)	893 (18.6)	882 (16.9)
Neutral	899 (24.2)	898 (20.8)	902 (21.8)	900 (22.3)
Incongruent	911 (22.4)	905 (22.0)	918 (24.8)	911 (23.1)
Mean	894 (20.5)	894 (20.0)	904 (21.7)	
Experiment 3				
Congruent	635 (2.9)	664 (3.3)	690 (3.0)	663 (3.0)
Neutral	696 (3.5)	706 (3.3)	709 (3.3)	703 (3.3)
Incongruent	696 (3.9)	706 (3.0)	715 (3.6)	706 (3.5)
Mean	676 (3.4)	692 (3.2)	704 (3.3)	

4