bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Crossmodal inhibition and enhancement in trimodal multisensory integration

Jonas Misselhorn, Jonathan Daume, Andreas K. Engel* & Uwe Friese*

Department of Neurophysiology and Pathophysiology,

University Medical Center Hamburg-Eppendorf,

Martinistr. 52, 20246 Hamburg, Germany

* These authors contributed equally to this work.

Corresponding author: Jonas Misselhorn, Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany. E-mail: j.misselhorn@uke.de

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Any instant of conscious perception is shaped by the differential contributions of all of our sensory organs. Complementary information in distinct sensory streams can be exploited by crossmodal integration. The principles that govern these integrative processes are not yet completely understood. The traditional bimodal nature of multisensory research hampers the exploration of the complex interactions of stimulus-driven and attention-related influences on multisensory integration. For this study, a novel trimodal amplitude-matching paradigm including vision, audition and somatosensation was developed in order to investigate multisensory mechanisms in the presence of perceptual competition. We report overall differences in bimodal integration efficiency, congruence enhancement for all bimodal targets and, most crucially, an inverse relation between overall integration efficiency and the magnitude of congruence enhancement. Furthermore, congruence enhancement between unattended/attended pairs determined the impact of distraction. Our results suggest the existence of multiple distinct mechanisms of inhibition and enhancement in multisensory integration.

Keywords: Multisensory integration; trimodal; congruence enhancement; attention; perceptual competition;

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Any instant of conscious perception is shaped by the differential contributions of all of our sensory organs (Ghazanfar & Schroeder, 2006). These inputs provide us with complementary information about the environment. Constantly, multisensory interactions occur in the brain to integrate these inputs and to evaluate crossmodal matching or conflict of sensory signals. The principles that govern these integrative processes are not yet completely understood. In the following, we introduce key aspects of crossmodal integration and propose a novel approach in studying them.

Perceptual processing is fundamentally competitive. As brain resources are limited, sensory processing in one modality is carried out at the expense of processing resources in other modalities (Desimone & Duncan, 1995; Kahneman, 1973). A good example for such crossmodal competition is the Colavita visual dominance effect (Colavita, 1974). In his seminal studies, participants failed to notice an above-threshold auditory stimulus when it was paired with a concurrent visual stimulus. This effect of visual dominance has been reproduced in numerous studies (for a review, see Spence, 2009). Recently, Sinnett and colleagues (2008) suggested that visual dominance over auditory processing might be due to early inter-sensory crosstalk. This mutual influence is asymmetric in that auditory stimulation can enhance visual detection, but visual stimulation impairs auditory detection. The same asymmetry was shown between somatosensation and vision, however, not between audition and somatosensation (Hecht & Reiner, 2009).

This asymmetry of crossmodal interactions implies that, besides inhibitory mechanisms in multisensory interaction, crossmodal enhancement exists as well. Already early on it was noticed that the co-occurrence of stimuli in distinct modalities can enhance perceptual processing (Todd, 1912). Later, it was found that this crossmodal enhancement depends on the degree to which stimuli share temporal and spatial characteristics (Stein & Meredith, 1993; Calvert, Spence, & Stein, 2004). In particular, the co-occurrence of relatively

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

weak stimuli evokes larger crossmodal enhancement than the combination of strong stimuli, a phenomenon termed inverse effectiveness (Otto, Dassy, & Mamassian, 2013). It has been concluded that these fundamental mechanisms allow us to bind sensory input into a coherent percept of an event, and conversely, to process sensory information separately if it is unlikely that a common origin is emitting these signals. Indeed, the bulk of multisensory research has been devoted to what has been termed the crossmodal binding problem (Senkowski, Schneider, Foxe & Engel, 2008; Spence, 2010; Treisman, 1996).

Apart from temporal and spatial factors, semantic congruence and crossmodal correspondence have been shown to influence multisensory integration (Doehrmann & Naumer, 2008; Spence, 2011). For instance, semantic congruence between visual and auditory information enhances performance in a detection task of natural objects (Schneider, Debener, Oostenveld, & Engel, 2008; Yuval-Greenberg & Deouell, 2007). Crossmodal correspondence relates to low-level stimulus features that are consistently associated crossmodally and are, thus, described as “natural cross-modal mappings” (Evans & Treisman, 2010). A very robust correspondence, for instance, exists between auditory pitch and visual brightness (Marks, 1987). That is, detection of visual and auditory stimuli is enhanced when jointly presented in a congruent fashion (i.e. for instance high pitch tone and bright visual stimulus) compared to incongruent presentation (i.e. for instance low pitch tone and bright visual stimulus).

Another key aspect in multisensory integration is the influence of attention. To that end, sensory processing of an attended event is enhanced at the expense of other sensory input (Desimone & Duncan, 1995; Kastner & Ungerleider, 2001; Wascher & Beste, 2010). This attentional modulation constantly interacts with the mechanisms of crossmodal inhibition and enhancement described above (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). For instance, crossmodal audio-visual enhancement can guide attention as was shown in a visual search paradigm (Van der Burg, L, Bronkhorst, & Theeuwes, 2008). Temporally aligned, but

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spatially uninformative auditory cues shortened search intervals of visual target detection in a screen of distractors. Vice versa, attention can directly interfere with multisensory integration (Talsma et al., 2010). For instance, visual attention to irrelevant lip movements interferes with audio-visual speech recognition. (Senkowski, Saint-Amour, Gruber, & Foxe, 2008). Finally, integration of coincident but spatially disparate sensory input seems to imply a crossmodal spreading of attention (Talsma et al., 2010). For example, dividing attention between visual and tactile spatial dot patterns (congruent or incongruent) leads to larger congruence effects compared to when attention is directed to only one of both modalities (Göschl, Engel, & Friese, 2014).

A major hindrance for understanding the complex interactions between multisensory mechanisms and attention is the fact that the majority of multisensory studies focus on bimodal integration. In order to investigate multiple aspects of multisensory integration in one approach, we developed a novel trimodal paradigm including vision, audition and somatosensation. Participants see a centrally presented drifting grating, hear a noise stimulus and receive vibro-tactile stimulation to the fingertips of both hands. In an amplitude-matching task, two modalities are attended while the third modality serves as task-irrelevant perceptual distractor. The stimuli in all modalities undergo a brief change in amplitude simultaneously, and participants have to indicate whether the changes in the attended pair of modalities were congruent or incongruent. This procedure allows comparing integration between different bimodal combinations, analyse differences in crossmodal enhancement and observe how perceptual competition influences mechanisms of bimodal integration.

Based on the literature reviewed above, we hypothesize that integration efficiency is highest in bimodal combinations including vision and that congruence between target stimuli will enhance integration in all bimodal conditions. Furthermore, for incongruent targets, congruence between attended stimuli and distractors is expected to impair performance due to a spread of attention to the unattended modality.

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Participants

Forty-nine participants were recruited for the study and received monetary compensation for their participation. Fifteen candidates were not able to complete the training successfully and were dismissed from further participation (performance was below an average of 70% correct answers). The remaining 34 participants were on average 24 years old (±4 years) and 20 of them were male (14 female). All had normal or corrected to normal vision and had no history of neurological or psychiatric disorders. After an explanation of the experimental procedure, participants gave written consent. The ethics committee of the Medical Center Hamburg-Eppendorf approved the study.

Stimulation

Participants were seated comfortably in a soundproofed and dimly lit chamber. In all trials, three sensory modalities were stimulated by signals of 2 s that underwent a transient change in amplitude (see Fig. 1.B). Visual stimulation consisted of a drifting, continuously expanding circular grating (see Fig. 1.A) presented centrally on a CRT screen (distance: 70 cm, visual angle: 5°) against a grey background . Phase onsets were randomized and contrast of the grating was experimentally increased or decreased. Auditory stimulation was delivered via ear-enclosing headphones (Ultrasone HIFI 680, transmission range 15-25.000 Hz). A complex sinusoidal sound was created by modulating high-frequency carrier signals (13 sine waves: 64 Hz and its first 6 harmonics as well as 91 Hz and its first 5 harmonics) with a low-frequency modulator (0.8 Hz) and was presented to the participants binaurally at 70 dB SPL. Changes of the complex tones were amplitude modulations with cosine-tapered transitions. Vibro-tactile stimulation was administered to both index fingers with C2 tactors (diameter: 2.97 cm, height: 0.76 cm, optimal stimulation frequency: 200-300 Hz, see http://www.atactech.com/PR_tactors.html). Two identical instances of these C2 tactors were

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

embedded into custom-made foam retainers to prevent repulsion related modulation of the vibration. They were located on a tray attached to the chair ensuring a comfortable posture throughout the whole experiment. Vibro-tactile stimulation was set to a frequency of 250 Hz and the amplitude of the perpendicular contactor displacement was experimentally increased or decreased. Phase onset of the complex sound and vibro-tactile stimulus were locked to the phase onset of the expanding grating.

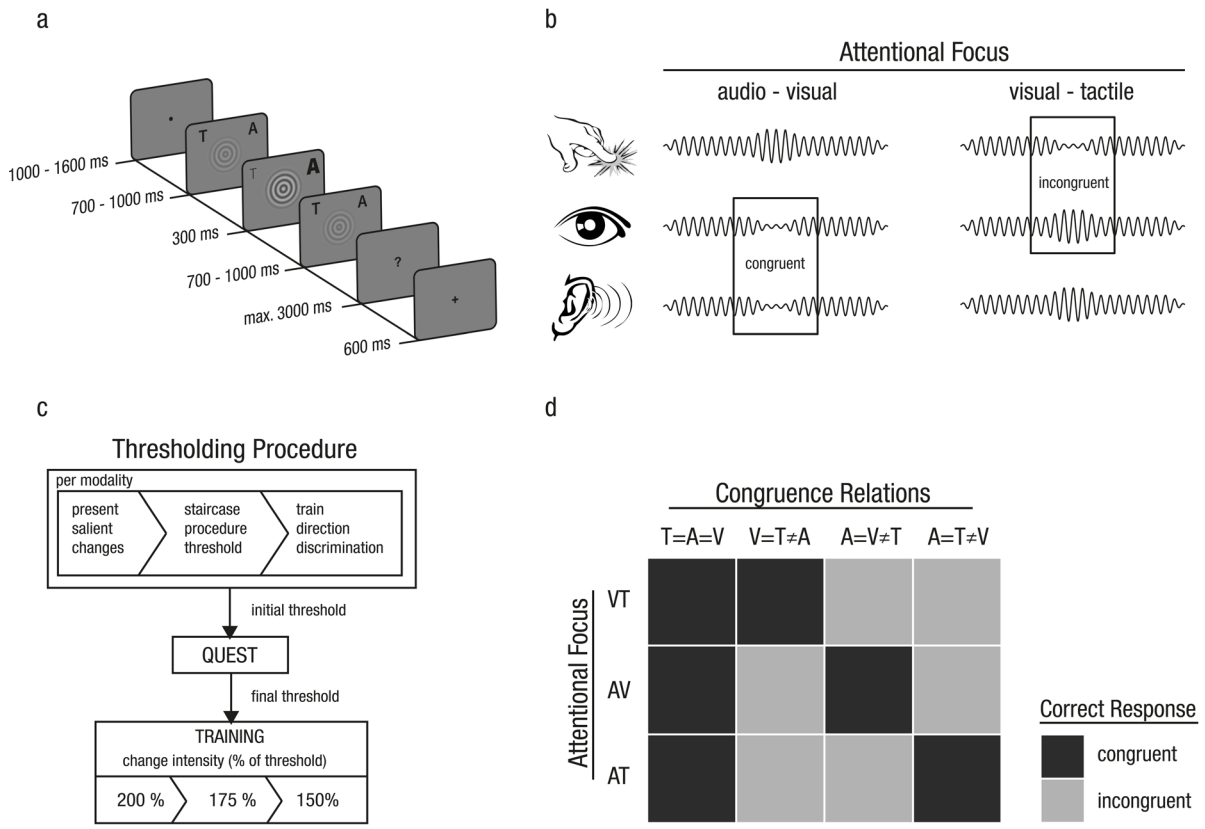

Fig. 1.

Overview of study design and procedure. In this example of a trial (a), visual contrast increases (central grating), vibration strength (T) decreases and auditory amplitude (A) increases. Visualization of crossmodal congruence (b) for an audio-visual (left column) and an incongruent visual-tactile trial (right column). Thresholding and training procedure (c) preceding the experiment. Correct answers (c) for all possible congruence relations ('=' is congruent, '≠' is incongruent) per attentional focus (VT = visual-tactile, AV = audio-visual, AT = audio-tactile).

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Experimental Paradigm

A central fixation dot was presented for 1000-1600 ms (see Fig. 1.A). Concurrent visual, auditory and tactile stimulation lasted for 700-1000 ms until a change in stimulus intensity occurred for 300 ms simultaneously in all modalities. After change offset, stimulation continued for another 700-1000 ms depending on the pre-change interval length (total duration of stimulation was always 2 s). After presentation, participants responded verbally (response interval max. 3000 ms) and received feedback for 600 ms (ISI = [3600, 4200] ms). The participants' task was to judge the congruence between changes of an attended pair of modalities irrespective of the change in the third modality. The attentional focus was cued block-wise as audio-visual (AV), visual-tactile (VT) or audio-tactile (AT). A change was defined as congruent if the changes in the two attended modalities had the same direction, e.g. a decrease in contrast of the grating and a coincidental decrease in loudness of the sound while AV was cued for this block (Fig. 1.B, first column). The direction of change in the ignored modality was irrelevant for the choice. A change was defined as incongruent if the directions of change in the attended modalities differed, e.g. a decrease in vibration intensity and a coincidental increase in contrast of the grating while VT had been cued for this block (Fig. 1.B, second column). After each trial, the participant was asked to report verbally if the change in the attended modalities was congruent ("gleich", German for "equal") or incongruent ("verschieden", German for "different"). To exclude possible response time (RT) differences on the basis of differences in the vocalization of the two response options, participants initially said "Yes" as soon as they felt ready to respond. RTs were measured from change onset until verbal response onset ("Yes"). Subsequently, the response ("gleich"/"verschieden") was given and evaluated online using custom-made speech recognition software. Feedback was given after each trial and at the end of each block. Each block consisted of 64 trials in random order and each stimulus configuration occurred 8 times. The order of attentional foci was randomized for each participant (e.g. AT, AV, VT) and

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

repeated 5 times. Thus, the complete experiment comprised a total of 15 blocks with each 64 trials.

Thresholding Procedure

The magnitude of change for a given modality and direction (increase or decrease) was individually adapted to ensure comparable perceptual salience and detection performance across modalities. The following procedure was implemented after extensive piloting (Fig. 1.C.). Initially, salient unimodal sample trials of both changes per modality were presented to get the participants acquainted with the stimulus material. These trials were repeated until the participants reported to have an intuitive feeling for how these stimuli and, most importantly, their changes look, sound or feel. Subsequently, a threshold was estimated with a reversed staircase procedure. Initial change magnitude was close to zero and was incremented from trial to trial. This was repeated until participants reported to be absolutely sure to have perceived the change. Subsequently, the direction of changes with the previously estimated magnitudes had to be judged by the participants. If they were not able to correctly categorize 80 % or more of the presented changes, estimated thresholds were increased and the task was repeated. After that, the actual threshold estimation was carried out using an implementation of the Bayesian adaptive psychometric method QUEST (Watson & Pelli, 1983). In this, all three stimuli were presented concurrently with only one stimulus changing per block that had to be attended. Increases or decrease of intensity were randomized and the magnitude of change was iteratively varied over 30 trials per change direction (see Fig. 2). The estimate from the preceding staircase procedure served as the initial prior. Subsequently, the last trial from the respective change direction in combination with information on success or failure to detect the change served as input for the Bayesian model to compute presentation intensity for the next trial. After 30 trials, stable estimations of detection thresholds were reached. These ranged between 25 % and 40 % of baseline amplitude and were largely symmetric with

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

respect to change direction within modalities (Fig. 2). The subsequent training comprised three levels. At the first level, change magnitudes were set to 200 % of the estimated threshold to facilitate learning the task. When all conditions were mastered at a performance level of 70 % correct answers or more, the second level with changes at a magnitude of 175 % of the estimated threshold followed. Finally, changes were presented at 150 % of the estimated threshold. This magnitude of changes was also used in the experimental blocks.

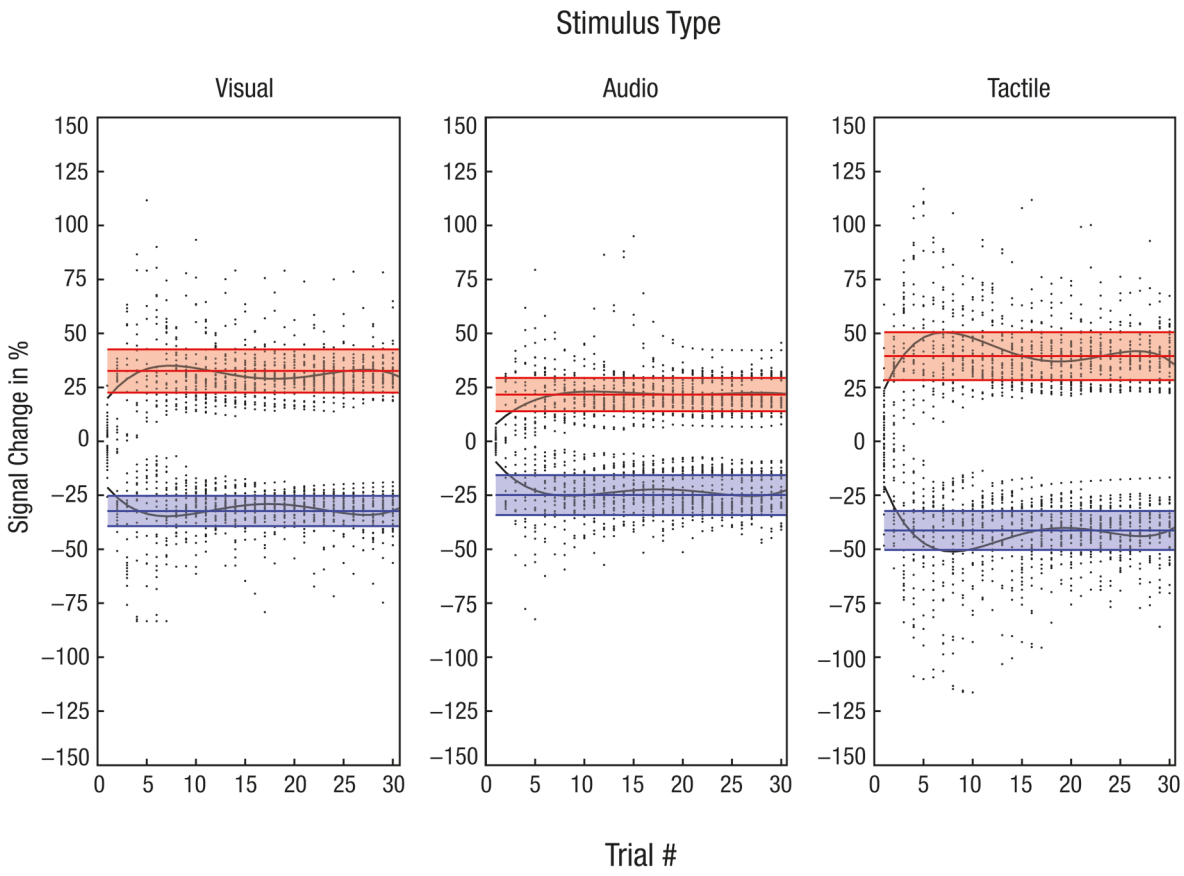

Fig. 2.

Summary of the thresholding procedure for the detection of visual, auditory and tactile stimulus intensity changes (increases/decreases). Scatter plots depict stimulation intensity change (y-axis) per trial (x-axis) from all participants. Black lines depict average progression of the detection threshold estimation (line of best fit for the scatter plot). Colored lines are drawn at the average final threshold (increases in red, decreases in blue; shaded areas represent standard error of the mean).

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Analysis

Response times (RTs) of correct responses were subject to an outlier analysis that was performed separately for each stimulus configuration, attentional focus and participant using MATLAB (Release 2012b, The MathWorks, Inc., Natick. Massachusetts, United States). Responses below or above an absolute z-transformed RT of 3 were excluded from further analysis, which was on average 0.87 % of all trials. The statistical analysis was performed using SPSS (IBM Corp. Released 2011. Version 20.0. Armonk, NY: IBM Corp.) and was focused on congruence between attended stimuli (*congruence*), the impact of perceptual competition (*distractor*) within the different attentional conditions (*attention*) and respective interactions. The direction of changes per se was ignored. Mean RTs and accuracies (ACCs) were calculated for each level of factors *attention* (VT, AV, AT), *congruence* (incongruent versus congruent attended) and *distractor* (congruence relation of unattended stimulus to attended stimuli). All averages can be found in Table 1 where conditions are ordered according to Figure 1.D.

It must be noted that the factor *distractor* is ambiguous with respect to the levels of *congruence*. For congruent attended conditions, distractor differentiates between fully congruent stimuli and a diverging unattended stimulus. For incongruent attended, distractor contrasts the two possible congruent pairs of the unattended stimulus with one of the attended stimuli. Therefore, an initial 3 (*attention*) x 2 (*congruence*) repeated-measures ANOVA was followed by a separate 3 (*attention*) x 2 (*distractor*) repeated-measures ANOVA for each level of the factor *congruence*. Simple effects of *attention* were computed using paired-sample t-tests with an alpha correction according to Holm-Bonferroni (Holm, 1979). The factor *distractor* was analysed further based on normalized difference scores. The net difference between *distractor* levels within a given attentional focus was divided by the mean of both values and scaled by 100. Simple effects of *distractor* were computed as one-sample t-tests against 0, interactions were evaluated with paired-sample t-tests between these

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

differences (again, alpha was adjusted according to Holm-Bonferroni). If sphericity was not given, reported values are corrected according to Greenhouse-Geisser. Effect sizes are given as partial eta squared (η^2). A priori estimation of required sample size was carried out for the 2 x 3 ANOVA reported above using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Alpha error probability was set to 5 % and power to 95 %. For detecting a medium sized interaction effect (f = 0.3, cf. Cohen, 1988), calculations suggested a sample size of N = 31. No stopping rule was used.

Table 1.

Mean response times (RT) and accuracies (ACC)

	T=A=V	V=T=V	A=V=V	A=T=V
VT				
ACC	88.97	82.98	85.48	84.19
	[85.89, 92.05]	[78.66, 87.30]	[81.59, 89.36]	[80.43, 87.95]
RT	1181.03	1208.03	1222.01	1222.97
	[1091.04, 1271.02]	[1118.86, 1297.20]	[1126.07, 1317.96]	[1133.80, 1312.14]
AV				
ACC	86.25	69.93	76.14	79.04
	[83.55, 88.95]	[64.33, 75.52]	[72.09, 80.19]	[74.68, 83.41]
RT	1224.27	1281.76	1246.06	1262.87
	[1126.19, 1322.34]	[1173.24, 1390.28]	[1151.14, 1340.98]	[1159.87, 1365.88]
AT				
ACC	82.72	65.88	69.56	79.38
	[97.58, 85.86]	[61.71, 70.05]	[65.20, 73.92]	[75.68, 83.07]
RT	1299.64	1393.49	1395.02	1324.24
	[1210.47, 1388.81]	[1298.35, 1488.63]	[1296.39, 1493.66]	[1230.54, 1417.94]

Note: VT = visual-tactile, AV = audio-visual, AT = audio-tactile; '=' denotes 'is congruent to', '≠' denotes 'is incongruent to'. Accordingly, 'V=T≠A' specifies trials in which visual and tactile stimuli changed congruently but the tactile stimulus was incongruent to both other; 95 % confidence intervals are given in brackets below the mean values.

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Attention and Congruence

Response times. The timing of responses was affected by *attention*, *congruence* and the interaction of both (see left panel of Fig. 3; *attention*: $F_{2,66} = 27.76$, $p < .001$, $\eta^2 = .45$; *congruence*: $F_{1,33} = 50.06$, $p < .001$, $\eta^2 = .60$; interaction: $F_{2,66} = 25.30$, $p < .001$; $\eta^2 = .43$). Participants gave fastest responses in condition VT and were slowest in AT (VT-AV: $t_{33} = -3.31$, $p = .002$; AV-AT: $t_{33} = -4.43$, $p < .001$; VT-AT: $t_{33} = -6.24$ $p < .001$). Congruence of attended targets speeded participants' responses significantly (VT: $t_{33} = -3.55$, $p = .001$; AV: $t_{33} = -3.97$, $p < .001$; AT: $t_{33} = -10.48$, $p < .001$). The magnitude of this congruence enhancement differed with respect to the levels of *attention*. While the effect of *congruence* was comparable between VT and AV (AV-VT: $t_{33} = -1.54$, $p = .134$), AT showed a considerably larger effect of congruence (AV-AT: $t_{33} = -4.73$, $p < .001$; VT-AT: $t_{33} = -6.33$, $p < .001$).

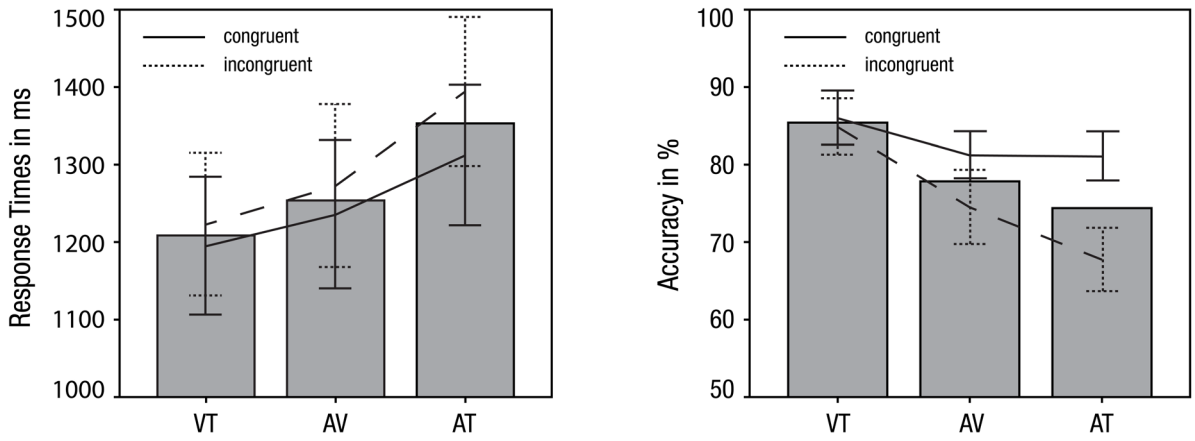

Fig. 3.

Response times (left panel) and accuracy (right panel) by factors *attention* and *congruence*. Bars depict mean values per levels of *attention* (VT = visual-tactile, AV = audio-visual, AT = audio-tactile). Lines depict differences within attentional conditions with respect to target congruence (error bars represent 95 % confidence intervals).

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accuracy. With respect to the accuracy of responding, the same overall picture of results emerged as for RTs (see right panel of Fig. 3; *attention*: $F_{2,66} = 27.25$, $p < .001$, $\eta^2 = .45$; *congruence*: $F_{1,33} = 44.06$, $p < .001$, $\eta^2 = .57$; interaction: $F_{2,66} = 21.84$, $p < .001$; $\eta^2 = .40$). Participants were most accurate in VT and made most errors in AT (VT-AV: $t_{33} = -5.14$, $p < .001$; AV-AT: $t_{33} = -2.45$, $p = .020$; VT-AT: $t_{33} = -6.55$ $p < .001$). Again, congruence of attended stimuli improved participants' overall performance significantly. In contrast to RTs, participants' accuracy in VT was not improved by congruence (VT: $t_{33} = 1.19$, $p = .245$). Significant congruence enhancements were found for AT and AV (AV: $t_{33} = 4.27$, $p < .001$; AT: $t_{33} = 7.23$, $p < .001$). Still, congruence enhancement was largest in AT (AV-AT: $t_{33} = 3.30$, $p = .002$).

Attention and Distraction

Response times. In the next two paragraphs, results from ANOVA of response times will be reported separately for congruent and incongruent targets.

Congruent targets. Participants' response timing was significantly affected by *attention* and *distractor* but not their interaction (upper left panel of Fig. 4; *attention*: $F_{2,66} = 18.53$, $p < .001$, $\eta^2 = .36$; *distractor*: $F_{1,33} = 29.42$, $p < .001$, $\eta^2 = .47$; interaction: $F_{2,66} = .10$, $p > .250$). The *attention* effect was identical to the overall effect of *attention* (VT-AV: $t_{33} = -2.99$, $p = .005$; AV-AT: $t_{33} = -3.67$, $p = .001$; VT-AT: $t_{33} = -5.10$ $p < .001$). Normalized differences across distractor levels indicated that the presence of a diverging distractor led to a significant slowing of participants' responses (VT: $t_{33} = -3.97$, $p < .001$; AV: $t_{33} = -3.29$, $p = .002$; AT: $t_{33} = -3.05$, $p = .004$).

Incongruent targets. The timing of responses differed significantly with respect to *attention* (upper right panel of Fig. 4; $F_{2,66} = 33.41$, $p < .001$, $\eta^2 = .36$) but neither *distractor* nor the interaction significantly affected RTs (*distractor*: $F_{2,66} = 2.24$, $p = .144$; interaction:

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$F_{1,33} = 2.95$, $p = .069$). Again, participants were fastest in VT and slowest in AT (VT-AV: $t_{33} = -3.46$, $p = .002$; AV-AT: $t_{33} = -4.92$, $p < .001$; VT-AT: $t_{33} = -7.12$, $p < .001$).

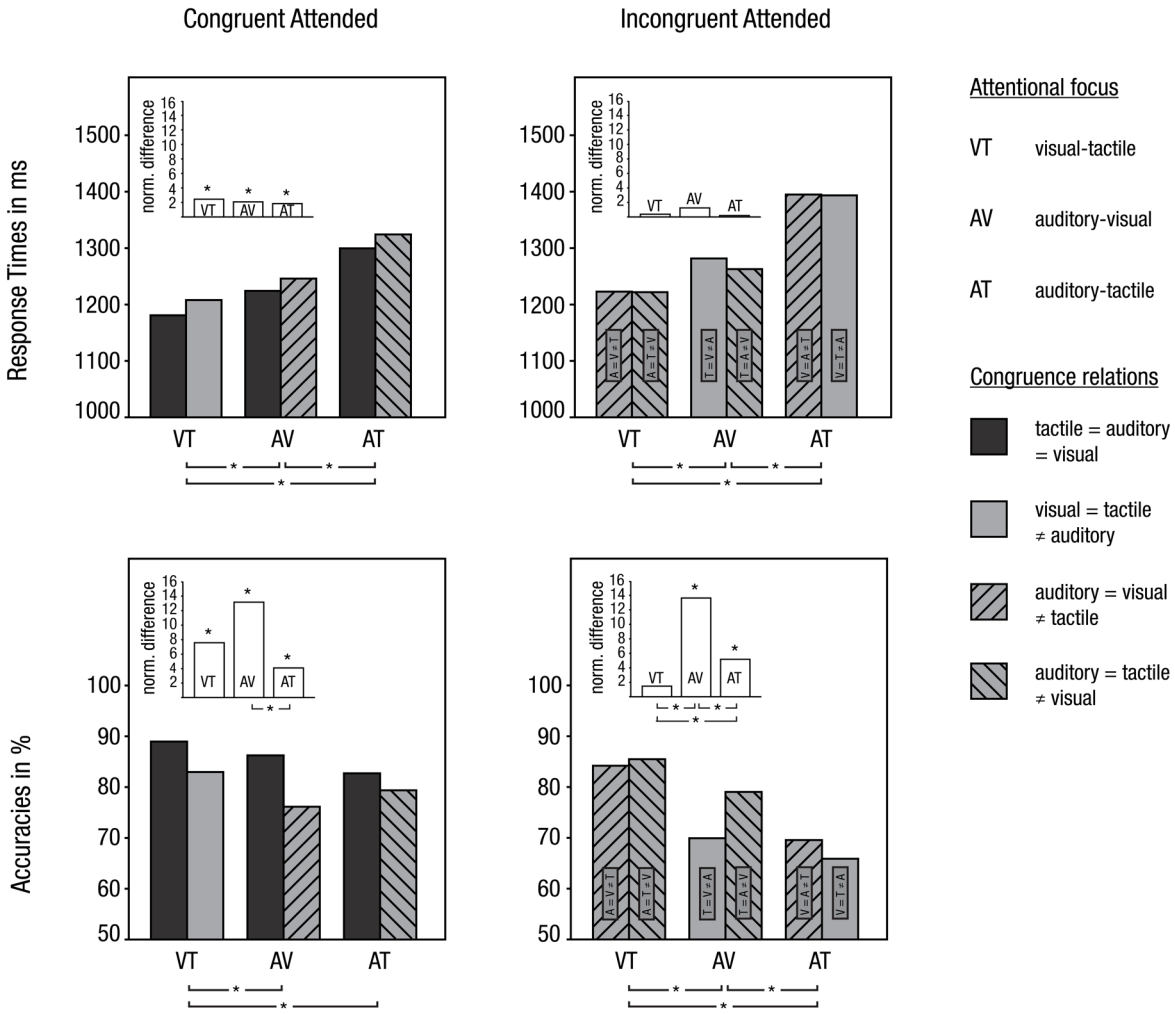

Fig. 4.

Response times and accuracies by factors *attention* and *distraction*. Bars depict averages of response times (upper row) and accuracies (lower row) for the congruent (left column) or incongruent (right column) attended conditions. Color-coding indicates the congruence pattern of the trimodal stimulus (legend on right margin). Significant mean differences are indicated by * below each plot. The subplots depict absolute values of normalized differences (see Method section for details) between *distractor* levels (neighboring bars per attentional focus). Again, * indicates significant differences tested against 0 (above bars) and significant differences of these values across attentional conditions (below chart). Alpha levels were adjusted according to Holm-Bonferroni. Note for the legend: '=' denotes 'is congruent to', '≠' denotes 'is incongruent to'.

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accuracy. In the next two paragraphs, results from ANOVA of accuracies will be reported separately for congruent and incongruent targets.

Congruent targets. Accuracy of responses was most profoundly affected by *distractor* (lower left panel of Fig. 4; $F_{1,33} = 63.99$, $p < .001$, $\eta^2 = .66$). Participants made significantly more errors if the distractor diverged from the congruent targets (VT: $t_{33} = 4.07$, $p < .001$; AV: $t_{33} = 6.08$, $p < .001$; AT: $t_{33} = 2.63$, $p = .013$). Also, the overall performance in the attentional conditions and the magnitude of distraction per attentional conditions differed significantly (*attention*: $F_{2,66} = 6.38$, $p = .004$, $\eta^2 = .16$; interaction: $F_{2,66} = 5.41$, $p = .007$, $\eta^2 = .14$). Participants were significantly more accurate in VT than in both AV and AT (VT-AV: $t_{33} = 3.51$, $p = .001$; VT-AT: $t_{33} = 2.72$, $p = .010$), but performance in AV and AT was comparable (AV-AT: $t_{33} = 0.10$, $p > .250$). While participants were equally strong distracted in VT and AV as well as VT and AT (VT-AV: $p = .059$, VT-AT: $p = .219$), tactile divergence in AV distracted significantly stronger than visual divergence in AT ($t_{33} = 3.09$, $p = .004$).

Incongruent targets. *Attention* affected participants' accuracy most strongly (lower right panel of Fig. 4; $F_{2,66} = 38.05$, $p < .001$, $\eta^2 = .54$). Participants showed highest accuracy in VT and made most errors in AT (VT-AV: $t_{33} = 5.48$, $p < .001$; AV-AT: $t_{33} = 3.50$, $p = .001$; VT-AT: $t_{33} = 8.15$ $p < .001$). Further, *distractor* and the interaction affected the accuracy of responding (*distractor*: $F_{1,33} = 8.75$, $p = .006$, $\eta^2 = .21$; interaction: $F_{2,66} = 23.71$, $p < .001$, $\eta^2 = .42$). Participants were significantly less accurate in AV when the tactile distractor was congruent to the visual target compared to when congruent to the auditory target (AV: $t_{33} = 5.83$, $p < .001$). Likewise, accuracy was lower in AT when the visual distractor was congruent to the tactile target compared to when congruent to the auditory target (AT: $t_{33} = 2.67$, $p = .012$). In VT, participants responded equally accurate in both conditions ($p > .250$). Thus, the differences in AV and AT were significantly larger than in VT (AV-VT: $t_{33} = 3.99$, $p < .001$; AT-VT: $t_{33} = 2.41$ $p = .022$). In addition, the difference in AV was significantly larger than in AT (AV-AT: $t_{33} = 7.40$, $p < .001$).

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

With this novel trimodal paradigm, we investigated how the interaction between multisensory congruence enhancement and attention is influenced by perceptual competition. To that end, a bimodal amplitude-matching paradigm was expanded by a third distracting modality that was task irrelevant. As we expected from theories on crossmodal correspondence (Spence, 2011), we found that congruence between attended stimuli facilitated responses in both timing and accuracy. We further demonstrated that this enhancement is even increased by an irrelevant but congruent stimulus to a third modality. This finding is in line with the idea of redundant signals which has also been shown for trimodal redundant targets (Diederich & Colonius, 2004). Although the overall congruence enhancement was present for all attended pairs of modalities, the magnitude of facilitation was inversely related to the overall efficiency of integration. That is, the highest congruence-related improvement was found for audio-tactile integration which overall was least efficient. Conversely, visual-tactile integration was overall most efficient but gained little from crossmodal congruence.

Global differences between attended modality pairs were more pronounced than expected. Not only vision-including target combinations (AV and VT) were integrated more efficiently than non-visual targets (AT), but also VT compared to AV. A plain explanation would make reference to visual dominance declaring that if a visual stimulus was target, integration was more efficient than if not. Even though this phenomenon is far from being fully understood, most authors agree that the dominance of visual input may largely be explained by the naturally higher salience of visual input compared with auditory or somatosensory input (Stubblefield, Jacobs, Kim, & Goolkasian, 2013). Our results, however, cannot be fully explained by salience because individual thresholds for each stimulus were estimated elaborately and balanced across modalities. Global differences are therefore not likely due to differences in stimulus salience per se, but are rather related to fundamental

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

differences in integration efficiency.

Alternatively, these differences in crossmodal processing might be related to a recent finding by Wang et al. (2012). In a target detection task, participants were presented with visual, auditory or tactile targets in the presence or absence of perceptual competition by the respective other two modalities. They observed that perceptual processing of auditory targets was inhibited by uninformative perceptual competition while vision was unaffected and tactile detection facilitated. Translating this finding to the current study, visual-tactile targets might be integrated most efficiently because the tactile component of the multisensory stimulus was enhanced and vision remained uninhibited. Conversely, audio-tactile integration might be difficult because the tactile component was enhanced while the auditory component was inhibited. This pattern would suggest that multisensory integration is most efficient if the modalities to be integrated have comparable levels of perceptual gain.

Yet, the inverse relation of overall integration efficiency and magnitude of congruence enhancement remains unexplained. It might be that multisensory inhibition/facilitation as described by Wang et al. (2012) and multisensory congruence enhancement represent distinct aspects of multisensory integration realized by distinct physiological mechanisms. In a subsequent study, Wang et al. (2013) correlated the replicated behavioural effects of multisensory facilitation and found a strong correlation to the P3 component of event-related potentials. Source analysis suggested that this effect involved the precuneus in the parietal lobe. In contrast, electrophysiological data from a cross-modal priming paradigm localized audio-visual congruence effects reflected in the P3 component as well as in gamma band power in temporal regions (Schneider et al., 2008). Further elucidation of possible neurophysiological mechanisms could be achieved in electrophysiological studies employing the current paradigm.

Furthermore, we demonstrated that congruence enhancement influences the impact of perceptual competition. Visual-tactile congruence had significantly more distracting impact

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on the integration of incongruent audio-visual targets compared to audio-tactile congruence. Likewise, integration of audio-tactile incongruent targets was impaired more severely by visual-tactile congruence than by than audio-visual congruence. The difference, however, was smaller in the latter compared to the former case. The resulting pattern of congruence-related competition enhancement (i.e. VT > AV ≥ AT) mirrors the overall pattern of integration efficiency noted earlier. This constitutes additional evidence for a superiority of visual-tactile integration and suggests that these differences in integration efficiency exist to some extent independently of attention. This contrasts with other studies that have observed attenuated multisensory integration between attended and unattended stimuli (Göschl, Engel, & Friese, 2014; Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008; Talsma, Doty, & Woldorff, 2007). In these studies, bimodal stimuli were presented but either uni- or bimodal targets were cued. Congruence enhancements were only evident or significantly stronger when attention was divided between modalities, leading to the conclusion that divided attention may be a necessary prerequisite for multisensory integration. Our data suggest that an exclusive bimodal focus in trimodal settings is difficult to achieve. Unimodal attention seems to represent a special case in which spread of attention is most successfully prevented.

Conclusion

In sum, we present evidence supporting the existence of highly dynamic principles of multisensory integration. We showed that multisensory congruence affects the processing of attended stimuli and contributes to perceptual competition of unattended distractors. While it enhances processing in the former case, unattended stimuli can gain access to processing capacities by means of attentional capture triggered by congruence-related enhancement. A global difference in bimodal integration efficiency as well as an inverse relation of these differences to the magnitude of congruence enhancement suggests the existence of multiple distinct mechanisms of multisensory inhibition and enhancement. One might speculate that

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

these processes reflect to some degree statistical properties of the natural environment. Vision and touch often occur congruently, in many cases even predictable if, e.g., self-initiated movements of the hands trigger visual-tactile percepts. In contrast, audio-visual as well as audio-tactile stimulus combinations predominantly have multiple external sources. As a consequence, congruence-related mechanisms are highly important in resolving the binding problem in multisensory integration. Using the paradigm studied here in electrophysiological or imaging studies will help clarify these speculations. Possibly, differences in functional and/or structural connectivity among sensory areas or between early sensory areas and multisensory regions might give rise to the behavioural results described here. Additionally, our results have implications for the development of complex human machine interfaces. Future research in this field should take different integration profiles of bimodal stimulus combinations into account in order to optimize the application of multisensory systems.

Author Contributions

A. K. Engel developed the study concept. All authors contributed to the study design. J. Misselhorn performed testing and data collection. J. Misselhorn conducted data analysis and interpretation with advice from U. Friese under supervision of A. K. Engel. J. Misselhorn drafted the manuscript, and all other authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements

This research was supported by grants from the German Research Foundation (SFB 936/A3) and the European Union (ERC-2010-AdG-269716).

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Calvert, G., Spence, C., & Stein, B. E. (2004). *The Handbook of Multisensory Processes*. MIT Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY, USA: Routledge Academic.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*, 406–412.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. doi:10.1146/annurev.ne.18.030195.001205

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–1404.

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150. doi:10.1016/j.brainres.2008.03.071

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*(1), 16–25. doi:10.1016/S1364-6613(00)01568-0

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6. doi:10.1167/10.1.6

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

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in*

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cognitive Sciences, *10*(6), 278–285. doi:10.1016/j.tics.2006.04.008

Göschl, F., Engel, A. K., & Friese, U. (2014). Attention modulates visual-tactile interaction in spatial pattern matching. *PloS One*, *9*(9), e106896. doi:10.1371/journal.pone.0106896

Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, *193*(2), 307–314. doi:10.1007/s00221-008-1626-z

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Statist*, *6*, 65–70.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia*, *39*(12), 1263–1276. doi:10.1016/S0028-3932(01)00116-6

Marks, L. E. (1987). On cross-modal similarity: perceiving temporal patterns by hearing, touch, and vision. *Perception & Psychophysics*, *42*(3), 250–256.

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*(1), 39–52. doi:10.1007/s00221-007-1080-3

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(17), 7463–7474. doi:10.1523/JNEUROSCI.4678-12.2013

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *NeuroImage*, *42*(3), 1244–1254. doi:10.1016/j.neuroimage.2008.05.033

Senkowski, D., Saint-Amour, D., Gruber, T., & Foxe, J. J. (2008). Look who's talking: The

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *NeuroImage*, *43*(2), 379–387. doi:10.1016/j.neuroimage.2008.06.046

Sinnett, S., Soto-Faraco, S., & Spence, C. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychologica*, *128*(1), 153–161. doi:10.1016/j.actpsy.2007.12.002

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research*, *176*, 245–258. doi:10.1016/S0079-6123(09)17615-X

Spence, C. (2010). Multisenory integration – Solving the crossmodal binding problem: Comment on "Crossmodal influences on visual perception" by Shams & Kim. *Physics of Life Reviews*, *7*(3), 285–286. doi:10.1016/j.plrev.2010.06.004

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. doi:10.3758/s13414-010-0073-7

Stubblefield, A., Jacobs, L. A., Kim, Y., & Goolkasian, P. (2013). Colavita dominance effect revisited: the effect of semantic congruity. *Attention, Perception & Psychophysics*, *75*(8), 1827–1839. doi:10.3758/s13414-013-0530-1

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(3), 679–690. doi:10.1093/cercor/bhk016

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. doi:10.1016/j.tics.2010.06.008

Todd, J. W. (1912). *Reaction to multiple stimuli*. Science Press.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178.

bioRxiv preprint doi: https://doi.org/10.1101/015958; this version posted March 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

doi:10.1016/S0959-4388(96)80070-5

Van der Burg, E., L, N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053–1065. doi:10.1037/0096-1523.34.5.1053

Wang, W., Hu, L., Cui, H., Xie, X., & Hu, Y. (2013). Spatio-temporal measures of electrophysiological correlates for behavioral multisensory enhancement during visual, auditory and somatosensory stimulation: A behavioral and ERP study. *Neuroscience Bulletin*, *29*(6), 715–724. doi:10.1007/s12264-013-1386-z

Wang, W. Y., Hu, L., Valentini, E., Xie, X. B., Cui, H. Y., & Hu, Y. (2012). Dynamic characteristics of multisensory facilitation and inhibition. *Cognitive Neurodynamics*, *6*(5), 409–419. doi:10.1007/s11571-012-9197-x

Wascher, E., & Beste, C. (2010). Tuning Perceptual Competition. *Journal of Neurophysiology*, *103*(2), 1057–1065. doi:10.1152/jn.00376.2009

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(5), 1090–1096. doi:10.1523/JNEUROSCI.4828-06.2007