Audiovisual integration is affected by performing a task jointly

Basil Wahn (bwahn@uos.de)
Institute of Cognitive Science – Neurobiopsychology, University of Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany

Ashima Keshava (akeshava@uos.de)
Institute of Cognitive Science – Neurobiopsychology, University of Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany

Scott Sinnett (ssinnett@hawaii.edu)
Department of Psychology, University of Hawai'i at Mānoa, 2530 Dole Street, Honolulu, HI 96822-2294, USA

Alan Kingstone (alan.kingstone@ubc.ca)
Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, Canada V6T1Z4

Peter König (pkoenig@uos.de)
Institute of Cognitive Science – Neurobiopsychology, University of Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany
Institut für Neurophysiologie und Pathophysiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

Abstract

Humans constantly receive sensory input from several sensory modalities. Via the process of multisensory integration, this input is often integrated into a unitary percept. Researchers have investigated several factors that could affect the process of multisensory integration. However, in this field of research, social factors (i.e., whether a task is performed alone or jointly) have been widely neglected. Using an audiovisual crossmodal congruency task we investigated whether social factors affect audiovisual integration. Pairs of participants received congruent or incongruent audiovisual stimuli and were required to indicate the elevation of these stimuli. We found that the reaction time cost of responding to incongruent stimuli (relative to congruent stimuli) was reduced significantly when participants performed the task jointly compared to when they performed the task alone. These results extend earlier findings on visuotactile integration by showing that audiovisual integration is also affected by social factors.

Keywords: multisensory integration; joint action; task distribution; social cognition.

Introduction

In everyday life, humans constantly process sensory input from several sensory modalities. If sensory input from multiple sensory modalities coincides in space and/or time, it is frequently integrated into a unitary percept (Alais & Burr, 2004; Ernst & Banks, 2002; Körding et al. 2007; Rohe & Noppeney, 2015; for a review, see: Spence, 2007) – a process referred to as "multisensory integration". Multisensory integration can result in perceptual benefits as well as costs. In particular, if the multisensory inputs contain redundant information (e.g., visual and auditory stimuli originate from the same spatial location), human localization performance is faster and more accurate (e.g., Körding et al. 2007; Rohe & Noppeney, 2015; Wahn & König 2015a,b; 2016). Yet, if the sensory inputs provide conflicting information (e.g., visual and auditory stimuli originate from different spatial locations but still coincide in time), human localization performance is slowed down and less accurate (Heed, Boukje, Sebanz, & Knoblich, 2010; Plöchl et al., 2016; Rohe & Noppeney, 2015; Spence, Pavani, & Driver, 2004). In the past, researchers have explored how attentional processes influence the multisensory integration process, and more generally, how attentional processing is distributed across the sensory modalities (e.g., Alais, Morrone, & Burr, 2006; Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Helbig & Ernst, 2008; Wahn & König, 2015a,b, 2016; Wahn, Murali, Sinnett, & König, 2017; for recent reviews, see Talsma, 2015; Wahn & König, 2017). However, to date, researchers have largely neglected how social factors could affect the integration process. Thus we know relatively little about how the social presence of another person, and/or how performing a task with another person, influences the process of multisensory integration.

To date, to the best of our knowledge, only two studies (Heed et al., 2010; Teneggi, Canzoneri, di Pellegrino, & Serino, 2013) have addressed the extent to which social factors can modulate multisensory integration. In Heed et al.'s experiment participants performed a visuotactile congruency task that was either performed alone, or jointly

with another person. When the task was performed alone, participants were required to hold two foam cubes, one in each hand, and indicate with foot pedal presses the spatial elevation of a tactile stimulus that could either appear at the top of the cube (i.e., felt at the index finger) or at the bottom of the cube (i.e., felt at the thumb). The participants also simultaneously received irrelevant visual stimuli that either appeared at the same spatial location as the tactile stimulus or not (i.e., stimuli were presented either in congruent or incongruent positions). Thus, the visual stimuli provided either conflicting or redundant spatial information, resulting in costs or benefits of multisensory integration, respectively. Heed et al. (2010) replicated earlier results (Spence, Pavani, & Driver, 1998; 2004) by finding that reaction times were faster when indicating the location of the tactile target if the visual stimulus appeared in a congruent position compared to an incongruent position. This effect is referred to as the "crossmodal congruency effect" (CCE). That is, when the tactile and visual stimuli provide redundant information (i.e., are presented in the same (congruent) position), localization performance is faster compared to when conflicting information is provided (i.e., stimuli are presented in different, i.e. incongruent positions). When participants performed the task in pairs, one of them indicated the elevation of the tactile stimuli (as before) while the second participant indicated the elevation of the visual stimuli. But note, the person detecting tactile stimuli was still exposed to the congruent or incongruent visual stimuli. Heed et al. found that the magnitude of the CCE was reduced when performing the crossmodal congruency task jointly compared to performing it alone. In particular, when participants performed the task jointly, incongruent presentations had less of an effect on reaction times when compared to performing the task alone. This observation suggests that the cost of incongruent presentations on multisensory integration is reduced when the task is performed jointly.

To date, the modulation of the CCE by social factors as found by Heed et al. (2010) has not been investigated with other sensory modalities. In particular, it is an open question whether audiovisual integration is similarly affected by social factors. Given that the tactile sensory modality processes events in close proximity while the auditory sensory modality is also able to sense more distal events, it is not clear whether audiovisual integration would be similarly affected by social factors as visuotactile integration. Thus rather than the visuotactile congruency task as used in Heed et al. the present study required participants to perform an audiovisual congruency task, either alone or jointly. If social factors modulate the CCE for audiovisual stimuli, we predict that the CCE will be reduced when performing the audiovisual congruency task jointly as compared to performing the task alone. Conversely, if social factors do not affect the CCE, then the CCE should not be modulated regardless of whether the task is performed jointly or alone.

Methods

Participants

Twelve pairs of individuals (15 female, M = 21.92 years, SD = 3.35 years) participated in the study at the University of Osnabrück. Prior to the experiment, participants signed informed written consent. The study was approved by the ethics committee of the University of Osnabrück. After the experiment had been completed, participants were debriefed and received monetary compensation or participation hours.

Experimental setup

Participants sat in a dark room in front of a computer screen (Apple 30" LCD screen, resolution 2560 x 1600 pixels, 77.53 x 48.46 visual degrees) at a distance of 50 cm. Four USB speakers (Mini HiFi USB 2.0 mini speaker), which were connected via a USB hub (Orico HF9US-2P USB 9-Port HUB) were arranged in a 2 x 2 grid above and below the monitor (vertically and horizontally 1600 pixels, equivalent to 48.45 visual degrees, apart) in front of the participants (Figure 1). The positions of the visual flashes (80 x 80 pixels, 2.42 visual degrees wide, 100 ms) were arranged in the same 2x2 grid, such that the visual flashes were observed from approximately the same spatial locations as the auditory stimuli (sine wave tone, 4800 Hz, 100 ms) – they were vertically displaced by 2.4 cm.

Participants sat in two chairs placed in front of the computer screen (left and right of the fixation cross, respectively) with keyboards on their laps.

Figure 1: Experimental Setup.

Experimental conditions and procedure

In the experiment, participants performed an audiovisual congruency task either alone or jointly. In this task, participants received visual flashes and auditory tones, originating either from the same (i.e., congruent) or a different (i.e., incongruent) spatial elevation. In addition, stimuli could originate either from the same or opposite side. For example, either both stimuli could originate from the left side or one could originate from the left and one from the right side. The task was to indicate the elevation of one of these stimuli using the keyboard with the mapping of keys F/up & C/down for the visual stimuli; keys K/up &

M/down for the auditory stimuli. We set the time limit for responses to 2 seconds (see Figure 2, for a trial overview).

When participants performed the task jointly, they sat next to each other in front of the computer screen in close proximity (~10 cm) to ensure that they shared peripersonal space (Heed et al., 2010). In this condition, one participant would indicate the elevation of the auditory stimuli while the other participant would indicate the elevation of the visual stimuli. When participants performed the task alone, one participant was asked to wait outside the experiment room while the other participant performed the task, indicating the stimulus elevation for their assigned modality. Note, regardless of whether participants performed the task alone or jointly, the seating positions of participants remained constant in all conditions within a pair and were counterbalanced across pairs (i.e., in half of the pairs, the participant responding to the auditory stimuli was sitting on the right side).

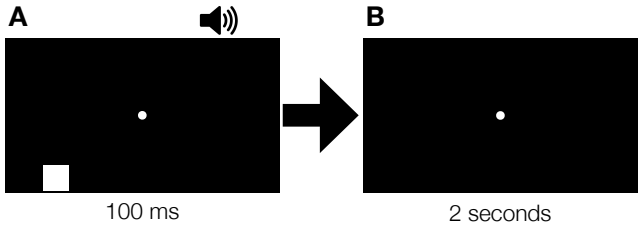

Figure 2: Trial overview. (A) Participants simultaneously received a visual and auditory stimulus. (B) Participants were required to indicate the elevation of one of the stimuli using the keyboard. In this example trial, the auditory stimulus would be in the upper location on the right side, the visual stimulus in the bottom location on the left side (i.e., an incongruent opposite side trial). After two seconds passed, the next trial started automatically.

In sum, the experiment consisted of a 2x2x2 factorial design with Congruency (Congruent, Incongruent), Side (Same, Opposite), and Condition (Individual, Joint) as factors.

The experiment consisted of six blocks, each composed of 144 trials. In these trials, each combination of the factor levels for the factors Congruency and Side occurred equally often in a randomized order. The factor Condition was varied across blocks. That is, there were three types of blocks: 1) The participant responding to the visual stimuli performing the task alone, 2) the participant responding to the auditory stimuli performing the task alone, 3) both participants performing the task jointly. Participants performed a pseudorandomized sequence of these three types of blocks twice. We avoided repetitions of the same block type in consecutive blocks.

The experiment took approximately 40 minutes. It was programmed in Python 2.7.3.

Data preparation and analysis

In line with Heed et al. (2010), we restricted our analysis to the participant in a pair responding to the auditory stimuli. That is, given that visual stimuli are considerably easier to localize than auditory stimuli, CCE effects are only observed for the participants responding to the auditory stimuli. Prior to performing inferential statistical tests, we tested whether the normality assumption was given with a Shapiro-Wilk test. In the case of a violation, we transformed the data using a log transformation.

Results

On a descriptive level (see Figure 3A & B), when examining the reaction times of correctly localized auditory cues, participants were slower to localize the cues in the incongruent condition compared to the congruent condition. This observation establishes the well-known CCE effect. Furthermore, in line with earlier studies (Heed et al., 2010), the CCE was more pronounced for stimuli that were shown on the same side compared to the opposite side. Importantly, for same side stimuli, the CCE was reduced profoundly in the joint condition relative to the individual condition.

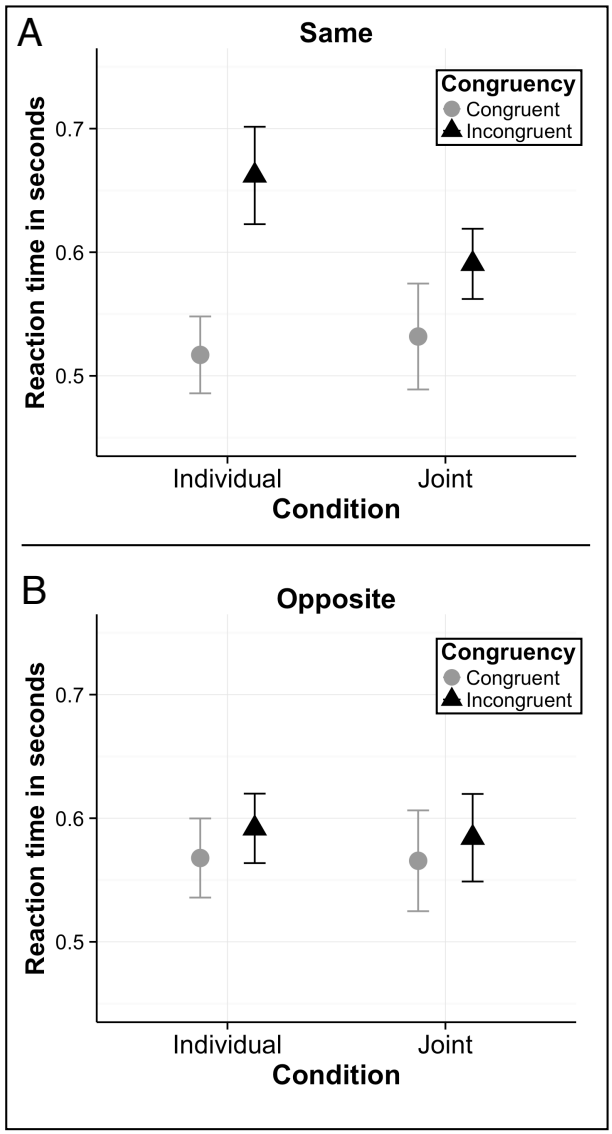

Figure 3: Mean reaction time (in seconds) as a function of the factors Condition (Individual, Joint) and Congruency (congruent, incongruent), separately for same side (A) and opposite side stimuli (B). Error bars in both panels are standard error of the mean.

We tested whether these observations were statistically reliable by performing a 2x2x2 repeated measures ANOVA with the factors Congruency (Congruent, Incongruent), Side (Same, Opposite), and Condition (Individual, Joint). As the assumption of normality was violated, we applied a log transformation to the reaction times prior to entering them to the ANOVA.

We found a significant main effect for the factor Congruency ($F(1,11) = 19.73$, $p < .001$). We found significant two-way interactions between the factors Side and Congruency ($F(1,11) = 38.42$, $p < .001$) and the factors Condition and Congruency ($F(1,11) = 6.00$, $p = .032$). The former interaction effect suggests that the magnitude of the CCE is reduced for opposite side stimuli compared to same side stimuli. Importantly, the latter interaction effect suggests that the CCE is reduced for the joint condition compared to the individual condition. In addition, we also observed a three-way interaction ($F(1,11) = 6.61$, $p = .026$), suggesting that the reduced CCE for the joint condition compared to the individual condition depends on whether stimuli appear on the same side or opposite sides. To further investigate the three-way interaction effect, we performed two 2x2 repeated measures ANOVAs (Condition x Congruency), restricting the data either to only same side or opposite side stimuli. For same side stimuli, we found a significant main effect of Congruency ($F(1,11) = 27.29$, $p < .001$) and a significant interaction between the factors Condition and Congruency ($F(1,11) = 9.62$, $p = .01$). This demonstrates that for same side stimuli, performing a task jointly indeed reduced the CCE. However, for opposite side stimuli, we only found a significant main effect of Congruency ($F(1,11) = 5.58$, $p = .038$) but no interaction effect between the factors Condition and Congruency ($F(1,11) = 0.07$, $p = .801$). Both of these results are in line with the findings by Heed et al. (2010). That is, when investigating visuotactile integration, Heed et al. (2010) similarly found that the CCE effect was reduced in the joint condition relative to the individual condition for same side stimuli but not for opposite side stimuli.

We also tested an alternative explanation of these results by a speed-accuracy tradeoff. That is, in the joint condition, participants potentially could have localized the incongruent cues faster at the expense of being less accurate in their responses. To investigate this, we repeated the 2x2x2 repeated measures ANOVA with the dependent variable fraction correct (for a descriptive overview, see Figure 4A & B). We found significant main effects for the factors Side ($F(1,11) = 73.32$, $p < .001$) and Congruency ($F(1,11) = 29.87$, $p < .001$) and a significant interaction effect between these two factors ($F(1,11) = 66.14$, $p < .001$). Importantly, we did not find a significant main effect or interaction involving the factor Condition (Condition: $F(1,11) = 0.31$, $p = .588$; Condition x Congruency: $F(1,11) = 0.02$, $p = .881$; Condition x Congruency x Side: $F(1,11) = 0.003$, $p = .954$). These results indicate that a speed-accuracy tradeoff does not explain the reduced CCE for the joint condition relative to the individual condition reported above because the accuracy did not vary as a function of whether the task was performed in pairs or alone. Thus the latency benefit of the joint condition relative to the alone condition was not acquired at the expense of committing more errors.

In sum, the results for same side stimuli indicate that the CCE is reduced significantly when participants perform an audiovisual crossmodal congruency task jointly compared to when they perform it alone.

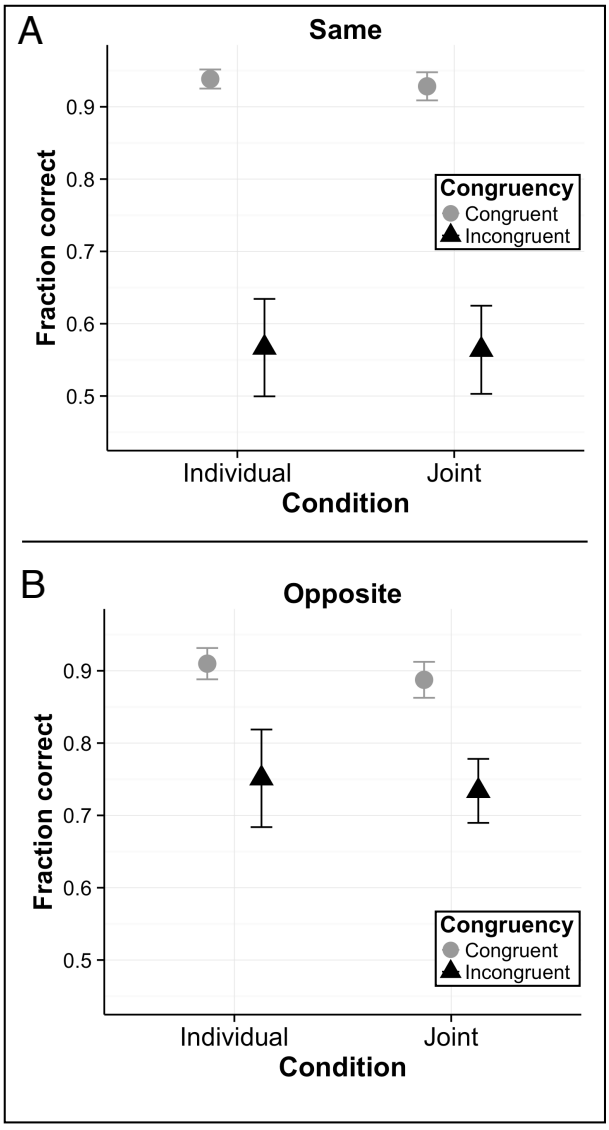

Figure 4: Mean fraction correct as a function of the factors Condition (individual, joint) and Congruency (Congruent, Incongruent), separately for same side (A) and opposite side stimuli (B). Error bars in both panels are standard error of the mean.

Discussion

The present study investigated whether the modulation of the CCE by social factors found in earlier studies investigating visuotactile integration (Heed et al. 2010) can also be observed for audiovisual presentations. In line with Heed et al., we found that the CCE is indeed reduced for same side stimuli when participants perform an audiovisual crossmodal congruency task jointly compared to performing it alone. Furthermore, we found that the data are not explained by a speed-accuracy tradeoff. Collectively, the present results extend Heed et al.'s earlier findings of a modulation of the CCE for a visuotactile crossmodal congruency task, and indicate that this social effect generalizes to audiovisual integration.

A possible "mechanism" for our present social effect could be a co-representation process (Sebanz, Knoblich, & Prinz, 2003; for reviews see: Sebanz, Bekkering, & Knoblich, 2006; Vesper et al., 2017). That is, when participants perform the task jointly, participants co-represent the task of their partner (e.g., that the partner responds to the visual stimuli) which could lead to a reduced processing of the stimuli relevant for the partner but irrelevant for the own task. As a consequence, the irrelevant stimuli could be perceived as less distracting for incongruent stimulus presentation but still sufficiently processed for congruent presentations, yielding faster reaction times. Alternatively, the effects in the present study could be explained by a dynamic modulation of the co-actor's peripersonal space as found in an earlier study (Teneggi et al., 2013) or by a general withdrawal of attention to the stimuli to which the co-actor responds (Szpak et al., 2015).

Future studies could discern further how social factors contribute to the modulation of the CCE. In the present study, pairs of participants performed the crossmodal congruency task in the same peripersonal space and both participants performed the task. Earlier findings (Heed et al., 2010) showed that the CCE for visuotactile stimuli is only affected by social factors if both participants perform the task *and* are located in their respective peripersonal spaces. It is an open question whether a reduction of the CCE for audiovisual stimuli would be observed when only one of these factors is manipulated. For instance, when participants are in the same peripersonal space but only one of them performs the task, *or* when both of them perform the task but from separate peripersonal spaces. In contrast to the tactile modality, both the visual and the auditory modality investigated here sample distant events. Thus, it is quite conceivable that visuotactile integration is dependent on jointly executing the task in peripersonal space while this might not be the case for audiovisual integration.

As another point of note, our finding that performing the crossmodal congruency task jointly affects the CCE for same side stimuli but not for opposite side stimuli could be explained by the observation that for opposite side stimuli the CCE was already greatly reduced in the individual condition. That is, an already lower CCE may not allow for any additional modulations by social factors.

Future studies could also test whether the social effects found in this study can alternatively be explained by other factors (Stenzel & Liepelt, 2016). For instance, it could be investigated whether a non-human co-actor (e.g., a robot) responding to the distractors is sufficient to find the effects in the present study (Stenzel et al., 2012).

More generally, the present findings are relevant to, and may benefit, real-world situations in which humans perform tasks jointly while processing multisensory information. That is, our data and the earlier findings of Heed et al., (2010) suggest that the benefits of multisensory integration are preserved when performing a task jointly (i.e., participants respond faster to congruent multisensory stimuli) while the costs of multisensory integration are reduced (i.e., participants are slowed down less by incongruent stimuli). Future studies could investigate further how the benefits of multisensory processing (e.g., due to multisensory integration (Alais & Burr, 2004; Ernst & Banks, 2002; Körding et al. 2007; Rohe & Noppeney, 2015), sensory augmentation (König et al., 2016; Goeke, Planera, Finger, & König, 2016), or circumventing limited attentional resources (Alais & Burr, 2004; Arrighi, Lunardi, & Burr, 2011; Wahn, et al. 2016; for a review, see: Wahn & König, 2017)) may facilitate human performance in other joint settings.

Acknowledgments

This research was supported by H2020—H2020-FETPROACT-2014641321—socSMCs (for BW & PK).

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257-262.

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*, 1339-1345.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839-843.

Arrighi, R., Lunardi, R., and Burr, D. (2011). Vision and audition do not share attentional resources in sustained tasks. *Frontiers in Psychology*. 2:56. doi:10.3389/fpsyg.2011.00056

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Goeke, C. M., Planera, S., Finger, H., & König, P. (2016). Bayesian alternation during tactile augmentation. *Frontiers in Behavioral Neuroscience*, *10*.

Heed, T., Habets, B., Sebanz, N., Knoblich, G. (2010). Others' actions reduce crossmodal integration in peripersonal space. *Current Biology, 20*, 1345–1349.

Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision*, *8*, 1-16.

König, S. U., Schumann, F., Keyser, J., Goeke, C., Krause, C., Wache, S., ... & König, P. (2016). Learning new sensorimotor contingencies: Effects of long-term use of sensory augmentation on the brain and conscious perception. *PLoS ONE*, *11*(12), e0166647. doi: 10.1371/journal.pone.0166647

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9), e943.

Plöchl, M., Gaston, J., Mermagen, T., König, P., & Hairston, W. D. (2016). Oscillatory activity in auditory cortex reflects the perceptual level of audio-tactile integration. *Scientific Reports*, *6*:33693.

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, *88*, 11-21.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70-76.

Spence, C., Pavani, F., & Driver, J. (1998). What crossing the hands can reveal about crossmodal links in spatial attention. *Abstracts of the Psychonomic Society*, 3, 13.

Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 148-169.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61-70.

Stenzel, A., & Liepelt, R. (2016). Joint Simon effects for non-human co-actors. *Attention, Perception, & Psychophysics*, *78*(1), 143-158.

Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1073.

Szpak, A., Loetscher, T., Churches, O., Thomas, N. A., Spence, C. J., & Nicholls, M. E. (2015). Keeping your distance: Attentional withdrawal in individuals who show physiological signs of social discomfort. *Neuropsychologia*, *70*, 462-467.

Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*:19. doi:10.3389/fnint.2015.00019

Teneggi, C., Canzoneri, E., di Pellegrino, G., & Serino, A. (2013). Social modulation of peripersonal space boundaries. *Current Biology*, *23*(5), 406-411.

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., ... , & Wahn, B. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, *7*, 2039.

Wahn B., & König P. (2015a) Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration. *Frontiers in Psychology,* 6:1084. doi:10.3389/fpsyg.2015.01084

Wahn B., & König P. (2015b) Vision and haptics share spatial attentional resources and visuotactile integration is not affected by high attentional load. *Multisensory Research 28,* 371-392. doi:10.1163/22134808-00002482

Wahn, B., Schwandt, J., Krüger, M., Crafa, D., Nunnendorf, V., & König, P. (2016). Multisensory teamwork: using a tactile or an auditory display to exchange gaze information improves performance in joint visual search. *Ergonomics*, 59, 781-795. doi: 10.1080/00140139.2015.1099742

Wahn B., & König P. (2016) Attentional resource allocation in visuotactile processing depends on the task, but optimal visuotactile integration does not depend on attentional resources. *Frontiers in Integrative Neuroscience,* 10:13.

Wahn B., Murali, S., Sinnett, S., & König P. (2017) Auditory stimulus detection partially depends on visuospatial attentional resources. *i-Perception*, 1-17. doi: 10.1177/2041669516688026

Wahn B., & König P. (2017) Is attentional resource allocation across sensory modalities task-dependent? *Advances in Cognitive Psychology*, *13*(1), 83-96. doi: 10.5709/acp-0209-2