

Audiovisual interactions depend on context of congruency

Beatriz R. Sarmiento · David I. Shore · Bruce Milliken · Daniel Sanabria

Published online: 10 December 2011
© Psychonomic Society, Inc. 2011

Abstract In this study, we addressed how the particular context of stimulus congruency influences audiovisual interactions. We combined an audiovisual congruency task with a proportion-of-congruency manipulation. In Experiment 1, we demonstrated that the perceived duration of a visual stimulus is modulated by the actual duration of a synchronously presented auditory stimulus. In the following experiments, we demonstrated that this crossmodal congruency effect is modulated by the proportion of congruent trials between (Exp. 2) and within (Exp. 4) blocks. In particular, the crossmodal congruency effect was reduced in the context with a high proportion of incongruent trials. This effect was attributed to changes in participants' control set as a function of the congruency context, with greater control applied in the context where the majority of the trials were incongruent. These data contribute to the ongoing debate concerning crossmodal interactions and attentional processes. In sum, context can provide a powerful cue for selective attention to modulate the interaction between stimuli from different sensory modalities.

Keywords Crossmodal interactions · Visual perception · Audition · Congruency context · Multisensory processing

B. R. Sarmiento (✉) · D. Sanabria
Departamento de Psicología Experimental,
Universidad de Granada,
Campus de Cartuja s/n,
18071 Granada, Spain
e-mail: bearsarmiento@ugr.es

D. I. Shore · B. Milliken
Department of Psychology, Neuroscience, & Behaviour,
McMaster University,
Hamilton, Ontario, Canada

Research in the field of cognitive control has been tightly linked to the study of stimulus congruency. The present research aimed to study the interplay between audiovisual interactions and attentional control processes driven by changes in context—that is, the relative proportions of congruent and incongruent trials.

In a typical congruency task (e.g., a Stroop task), two stimulus dimensions, one task-relevant and one task-irrelevant, can be congruent (triggering the same response) or incongruent (triggering incompatible responses). For example, in the Stroop task, the colour word and the colour in which the word is written can be congruent (e.g., red written in red) or incongruent (e.g., red written in blue). The difference in performance between congruent and incongruent trials (i.e., faster response times [RTs] and higher response accuracy on congruent trials) is called the *congruency effect*. It is argued (e.g., Botvinick, Cohen, & Carter, 2004; Egner & Hirsch, 2005) that the size of the congruency effect reflects the level of control that the participant exerts to avoid the interference of irrelevant information while making a fast and accurate response. According to this view, the larger the size of the congruency effect, the more interference has occurred, and thus, the less control was exerted.

A number of studies have shown that the congruency effect can vary as a function of the proportion of congruent trials. In a seminal study, Logan and Zbrodoff (1979) manipulated the proportions of congruent and incongruent trials presented in a Stroop task, finding a decreased congruency effect in those blocks of trials with a higher proportion of incongruent items. The authors (see also Lindsay & Jacoby, 1994) concluded that this proportion-congruent effect was due to changes in word-reading strategies in response to changes in the likelihood of congruency.

Crump, Gong, and Milliken (2006), using the same rationale as the above-mentioned studies, manipulated the

proportions of congruent items presented in a consecutive-trial variant of the Stroop task. A colour-word prime was presented in white at fixation and was followed by a coloured rectangle target that appeared above or below fixation. The colour of the target either matched (i.e., was congruent with) or mismatched (i.e., was incongruent with) the meaning of the preceding colour-word prime. The participants' task was to identify the colour of the rectangle. Unlike previous experiments, the coloured rectangle targets were presented in either of two different contexts (above or below a central fixation point), mixed at random within blocks of trials. One of these contexts/locations was associated with a high proportion of congruent trials (75%), and the other was associated with a low proportion of congruent trials (25%). Participants in Crump et al.'s (2006) study could not predict which proportion-congruent context the current trial belonged to until the onset of the coloured target itself. Hence, participants could not take advantage of preparatory strategies prior to the onset of the prime word that would facilitate or inhibit word reading in accord with the proportion of congruency. Nonetheless, the results of Crump et al. (2006) showed that the proportion-congruent context, even when manipulated randomly at target onset within a block of trials, modulated the Stroop interference in accordance with previous studies: a significantly reduced congruency effect on the low-proportion-congruent context relative to the high-proportion-congruent context. They proposed that these context-sensitive control effects reflected learning of associations between the context (e.g., location) and the likelihood of congruency, giving rise to context-specific proportion-congruent (CSPC) effects.

Crump et al. (2006) suggested that attentional processes were responsible for the CSPC effects obtained in their study. On the one hand, selective attention was necessary in order to associate a proportion of congruency with a particular context (see Crump, Vaquero, & Milliken, 2008, Exp. 3). On the other hand, this association implied that a particular context triggered attentional adjustments to select the suitable incoming information (Crump & Milliken, 2009).

The critical question addressed in our study was whether CSPC effects can also be obtained when congruency refers to a stimulus dimension shared between inputs from two sensory modalities. The manipulation of congruency using multisensory stimuli allows for the testing of mechanisms of crossmodal interaction (Calvert, Spence, & Stein, 2004). In particular, when inputs from two sensory modalities are incongruent, performance can reveal whether an input from one sense is able to affect the perception of inputs in the other sensory modality. In effect, both behavioural (e.g., Sanabria, Spence, & Soto-Faraco, 2007) and neuroimaging (e.g., Alink, Singer, & Muckli, 2008; Watkins, Shams, Josephs, & Rees, 2007) research has shown that crossmodal congruency effects can reflect interactions at the perceptual level of information processing prior to attentional

deployment, in contrast with unimodal congruency effects, such as the Stroop effect, that occur at higher, postperceptual levels of information processing.

Relevant here is the study by Shore and Simic (2005), who showed a modulation of the visuotactile congruency effect by the relative proportions of congruent and incongruent trials. The authors manipulated the proportions of congruent and incongruent trials between blocks, obtaining a reduced congruency effect for errors in blocks of trials with 25% congruent trials, with respect to blocks of trials with 75% congruent trials. However, this effect was only obtained when the visual distractor was presented 100 ms prior to the tactile target. The authors concluded that visuotactile integration was immune to top-down influences, and, if present, these influences could be exerted only when the two sensory inputs were presented asynchronously. Therefore, the questions remain unresolved whether CSPC effects could be obtained in the crossmodal domain when the two sensory inputs are presented in synchrony (i.e., maximising the crossmodal interaction process) and, moreover, whether CSPC effects could be shown when two different contexts of congruency are manipulated within the same block of trials, as in Crump et al.'s (2006) study.

The congruency manipulation in our study would produce more accurate responses on congruent trials and more errors on incongruent trials, with respect to a baseline unimodal visual condition. A significant difference in accuracy was also predicted between congruent and incongruent trials. Crucially, in line with Crump et al.'s (2006) argument about the implication of attentional processes in CSPC effects, we expected a larger congruency effect in the high- than in the low-proportion-congruent context, supporting the idea that audiovisual interactions depend on the deployment of attentional control, in line with relevant empirical evidence from behavioural (e.g., Sanabria, Soto-Faraco, & Spence, 2007), ERP (e.g., Talsma, Doty, & Woldorff, 2007), and neuroimaging studies (e.g., Fairhall & Macaluso, 2009) that have demonstrated these top-down influences.

In Experiment 1, we developed an audiovisual task that resulted in a congruency effect whereby the perceived duration of a visual stimulus was modulated by the synchronous presentation of an auditory stimulus of the same (congruent) or different (incongruent) duration (see Klink, Montijn, & van Wezel, 2011, for a somewhat similar procedure). The data showed more accurate performance on congruent than on incongruent trials, revealing a congruency effect similar to the one reported in previous studies (e.g., Fairhall & Macaluso, 2009). Signal detection theory (SDT) analyses have revealed that visual inputs influenced auditory perception at a perceptual level of processing (see Sanabria, Spence, & Soto-Faraco, 2007; Watkins et al., 2007). In Experiments 2–4, we associated the audiovisual stimuli with two different contexts, with regard to the

likelihood of congruency. The particular context of congruency was manipulated both between blocks of trials (Exp. 2) and within blocks of trials (Exp. 3 and 4). On the basis of the findings of Crump and colleagues (Crump et al., 2006; Crump & Milliken, 2009; Crump et al., 2008), we expected to find a reduced crossmodal congruency effect in the context with more incongruent trials than in the context with a higher proportion of congruent trials.

The combination of Crump et al.'s (2006) paradigm and crossmodal congruency constitutes a novel approach, which allows for an innovative way of studying the role of attention in crossmodal interaction. Our findings would then shed new light on the interplay between attentional processes and crossmodal interactions (see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010, for a review).

Experiment 1

Method

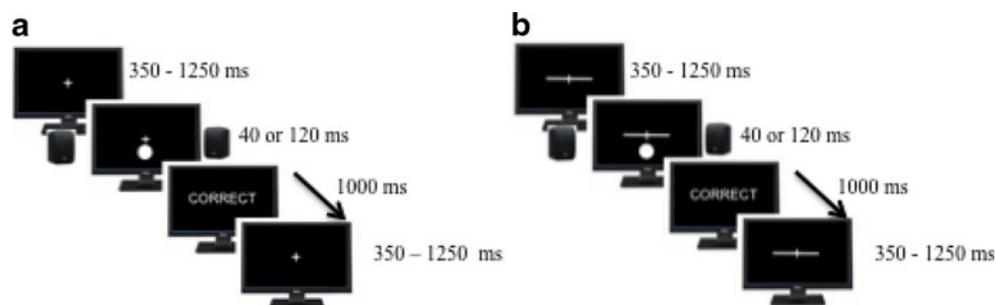
Participants The participants were 19 undergraduate students (13 females; age range 17–43 years old, mean age 19 years) who received course credits in exchange of their participation. All of the participants in this study reported normal hearing and normal or corrected-to-normal vision, and they gave their informed consent to participate in the study, which was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus and stimuli The experiment was conducted on an Intel Core 2 Duo PC with a 17-in. LCD monitor. E-Prime software was used for both stimulus presentation and response collection (Schneider, Eschman, & Zuccolotto, 2002).

The visual stimuli consisted of a white cross that served as the fixation point and a target white circle (3.01° in diameter). Two loudspeakers, positioned on each side of the computer screen, were used to present the auditory stimulus, which consisted of a white-noise burst (60 dB[A] measured at ear level).

Procedure Participants sat in a comfortable chair approximately 57 cm from the computer monitor in a dark room.

Fig. 1 Schematic illustration of the setup used in Experiments 1–3 (a) and Experiment 4 (b)



They were asked to discriminate the duration (short or long) of a central white circle displayed below (4.52° from the centre of the circle) the fixation point for either 40 or 120 ms, while ignoring the synchronous presentation of a white-noise burst that could last for 40 or 120 ms. This manipulation gave rise to congruent trials (the visual stimulus was of the same duration as the auditory stimulus) and incongruent trials (the visual stimulus was of a different duration than the auditory stimulus).

At the beginning of each trial, participants were presented with a fixation cross displayed in white against a black background for a random duration between 350 and 1,250 ms that remained present for the whole trial (see Fig. 1 for the trial sequences of all experiments). The circle target was then presented for either 40 or 120 ms, accompanied by the white-noise burst, which could be congruent or incongruent in duration. The short and long stimuli had the same probabilities of appearance. This was also true for congruent and incongruent trials. Accuracy, rather than response speed, was stressed, so participants had no response time pressure. Half of the participants were told to press the “C” key if the circle was short in duration and the “N” key if the circle was long in duration, and the reverse stimulus–response mapping was used for the remaining participants. Feedback about response accuracy was provided to the participant, and the next trial began 1,000 ms after the feedback.

The experiment began with a unimodal discrimination task (the visual target stimuli were presented in the absence of auditory distractors) in which participants completed 8 unimodal practice trials, followed by a block of 40 (20 short and 20 long) unimodal experimental trials. The unimodal block was followed by the crossmodal discrimination task, in which participants completed 8 crossmodal practice trials and four blocks of 40 crossmodal trials each.

Results

The data for 1 participant were removed from the analysis because her performance did not reach 60% accuracy on the baseline condition. The remaining 18 participants (12 females; age range 17–43 years old, mean age 19 years) were included in the complete data analysis.

The mean response accuracy¹ for each participant and condition were analysed using pair-wise two-tailed *t* tests.

There was a significant difference between congruent and incongruent trials, $t(17) = 10.98$, $p < .001$ (see Fig. 2). Responses on congruent trials (83% correct) were more accurate than responses on incongruent trials (45% correct).

The comparison between each type of trial and the baseline (unimodal condition) reached statistical significance: $t(17) = 3.17$, $p = .005$, and $t(17) = 8.78$, $p < .001$, for congruent and incongruent trials, respectively.

Sensitivity (d') and criterion (c) indexes were assessed in the baseline, short auditory, and long auditory conditions for each participant on the basis of the proportions of hits (correct short visual target detections in short-target trials) and false alarms (incorrect short visual responses in long-target trials).

Hypothesis-driven pair-wise two-tailed *t* tests revealed differences in d' between the baseline unimodal (1.51) condition and both crossmodal conditions (short [0.89] and long [0.89] auditory)— $t(17) = 3.55$, $p = .002$, and $t(17) = 3.19$, $p = .005$, respectively—revealing a higher sensitivity in the unimodal condition. The difference in d' between the two crossmodal conditions was not significant, $t < 1$. The differences in c between the baseline condition (0.32) and the two crossmodal conditions (0.69 for short and 0.52 for long), along with the difference between the two crossmodal conditions themselves, were significant, all $ps < .003$. The data revealed a bias towards responding “short duration” for the circle when it was presented with a short auditory stimulus, and towards responding “long duration” when the duration of the auditory stimulus was long.

Discussion

Experiment 1 showed that participants committed more errors on incongruent than on congruent trials in the visual temporal discrimination task. This result, in accordance with previous studies (e.g., Klink et al., 2011; Shams, Kamitani,

¹ In contrast to the majority of studies using congruency tasks, which have focused on RTs, accuracy was our main dependent variable. We believe that accuracy was the more suitable dependent variable for the following reasons: (1) Temporal pressure in our study would have resulted in an even greater number of errors, which would have prevented us from obtaining enough correct responses to analyze the RT data. (2) The focus of interest here was response precision, not response speed. (3) Since participants in the task used in our study committed errors even without temporal pressure, accuracy was considered to be a very informative variable of audiovisual interactions, given that those errors were not committed due to response pressure.

Note also that other studies using tasks similar to ours have also used accuracy as their main dependent variable (e.g., Klink et al., 2011; Shams, Kamitani, & Shimojo, 2002). In any case, the analyses on the RT data from Experiments 1–4 did not reveal any statistically significant effect.

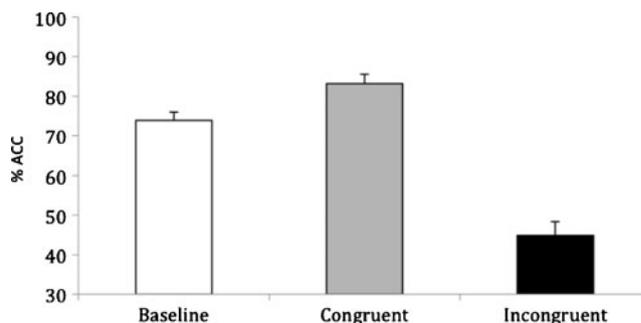


Fig. 2 Percentages of correct responses for unimodal versus crossmodal (congruent and incongruent) trials in Experiment 1. Error bars represent the standard errors of the means across participants

& Shimojo, 2000, 2002), suggests that the auditory stimulus altered the perception of the visual stimulus when the two synchronously presented stimuli were incongruent in duration.

SDT analyses suggest that this modulation occurs at both perceptual (d') and decisional (c) levels of information processing (see Sanabria, Spence, & Soto-Faraco, 2007). Crucially, the d' effect on crossmodal trials (as compared to the baseline, unimodal condition) did not depend on whether the auditory stimulus was short or long in duration.

In Experiment 2, we investigated whether the congruency effect found in Experiment 1 could be modulated by the likelihood of congruent trials. To this end, the multisensory stimuli were presented in two contexts with different frequencies of congruent trials. SDT analyses were conducted for each congruency context in order to study the level of processing at which the modulation by context occurred.

Experiment 2

Method

Participants The participants were 32 undergraduate students (20 women; age range 17–39 years old, mean age 19 years) who volunteered in exchange for course credit.

Apparatus, stimuli, and procedure These were the same as in Experiment 1, except in the following respects: In Experiment 2, we included two different congruency contexts. For half of the experimental blocks (high-congruency context), 80% of the trials were congruent and the remaining 20% of trials were incongruent. The reverse was true for the remaining half of the blocks (low-congruency context).

Participants performed eight blocks of 40 trials, four in each of the high- and low-congruency-context conditions. Half of the participants performed four blocks of trials in the high-congruency-context condition followed by the four blocks of trials in the low-congruency-context condition. The reverse order of presentation of the congruency context

conditions was used for the remaining half of the participants. After the first four blocks of trials, participants were informed that they had completed half of the crossmodal discrimination task. At the end of the experiment, participants completed a questionnaire to evaluate whether they had been aware of the congruency manipulation. They were asked whether they perceived any difference between the two halves of the crossmodal discrimination task and what that difference might be. They were also asked regarding the reliability of the feedback information (i.e., whether they believed that the information provided by the feedback was in accordance with their response).

Results

Accuracy measures The data from 8 of the participants were removed from the analysis because their performance did not reach 60% accuracy in the baseline condition. We set this criterion to make sure that participants understood the task at hand. The remaining 24 participants (14 females; age range 17–39 years old, mean age 19 years) were included in the complete data analysis.

The mean accuracy for each participant and condition were submitted to a 2 (Congruency: congruent, incongruent) × 2 (Proportion congruent: high, low) ANOVA, which revealed a significant main effect of congruency, $F(1, 23) = 64.88, MSE = 2.59, p < .001, \eta_p^2 = .74$: Responses on congruent trials (80% correct) were more accurate than responses on incongruent trials (47% correct). The main effect of proportion congruent was not significant, $F < 1$.

Crucially, the interaction between congruency and proportion congruent was significant, $F(1, 23) = 42.96, MSE = 0.57, p < .001, \eta_p^2 = .65$. The size of the congruency effect (percent correct on congruent trials – percent correct on incongruent trials) was smaller for the low-proportion-congruent context (17%) than for the high-proportion-congruent context (48%). A priori comparisons revealed that this difference was due to lower accuracy on congruent trials (15.3%), $F(1, 23) = 23.72, MSE = 0.28, p < .001$, and higher accuracy on incongruent trials (15.5%), $F(1, 23) =$

12.35, $MSE = 0.29, p < .001$, in the low-proportion-congruent context with respect to the high-proportion-congruent context (see Table 1). Participants performed more accurately on congruent than on incongruent trials in both proportion-congruent conditions: $F(1, 23) = 95.53, MSE = 2.8, p < .001$, and $F(1, 23) = 15.21, MSE = 0.36, p < .001$, for the high- and low-proportion-of-congruency contexts, respectively.

To investigate whether the CSPC effect depends on the learned association between the context and the likelihood of congruency, as suggested by Crump et al. (2006), we compared participants’ performance in Experiment 1 between the first two blocks and the last two blocks of each context. The mean accuracy for each participant and condition was submitted to a 2 (Congruency: congruent, incongruent) × 2 (Proportion congruent: high, low) × 2 (Block: first half, second half) ANOVA. The first 10 trials of the first block were considered part of the training phase and were excluded from the analysis.

Critically, the three-way interaction between congruency, proportion congruent, and block was significant, $F(1, 23) = 10.38, MSE = 0.099, p < .005, \eta_p^2 = .31$ (see Fig. 3). While accuracy on incongruent trials was significantly higher in the low-proportion-congruent context with respect to the high-proportion-congruent context in the second half of each context, $F(1, 23) = 19.47, MSE = 0.57, p < .001$, that difference did not reached statistical significance in the first half of each context, $F(1, 23) = 1.68, MSE = 0.58, p = .21$.

The short and long auditory conditions were pooled for each congruency context (high and low) to obtain sensitivity (d') and criterion (c) indexes. Pair-wise two-tailed t tests showed nonsignificant differences on d' and c between the high- and low-proportion-congruent contexts [$t < 1$ and $t(23) = 1.24, p = .22$, respectively].

Questionnaire measures Regarding the congruency manipulation, only half of the participants (46%) noticed a difference between the two proportion-congruent conditions, although none of them appeared to be aware of the proportion manipulation (see Table 2). Instead, those who noted a difference between the two halves of the experimental session offered

Table 1 Mean correct responses (in percentages), and standard errors, for Experiments 2–4

Experiment	Proportion Congruent	Congruent		Incongruent	
		M	SE	M	SE
2	High	87.60	1.83	39.32	4.08
	Low	72.27	3.54	54.85	3.46
3	High	82.94	2.24	53.65	3.40
	Low	81.59	2.78	52.65	3.41
4	High	77.57	2.21	53.82	2.64
	Low	77.03	2.38	57.18	2.64

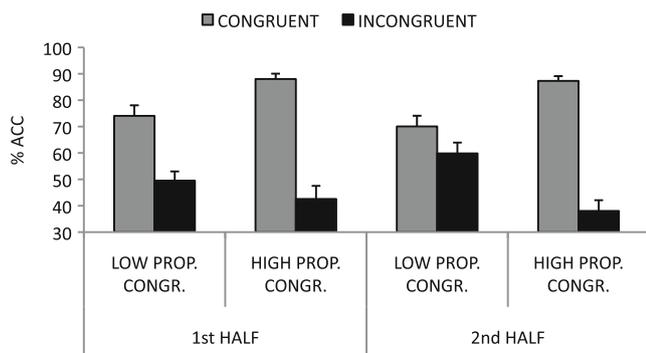


Fig. 3 Percentages of correct responses as a function of congruency, proportion congruent, and block in Experiment 2. Error bars represent the standard errors of the means across participants

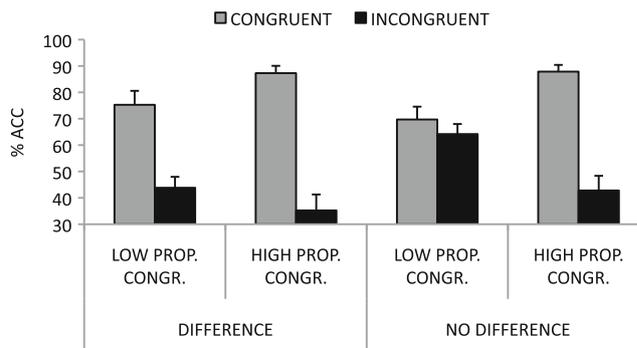


Fig. 4 Percentages of correct responses as a function of congruency, proportion congruent, and context difference in Experiment 2. Error bars represent the standard errors of the means across participants

comments like “the first part was easier due to tiredness” or “I made fewer mistakes on the second part of the experiment, presumably due to practice.”

To determine whether the performance of participants who noticed a difference between the two congruency contexts differed significantly from the performance of participants who did not, a 2 (Congruency: congruent, incongruent) × 2 (Proportion congruent: high, low) mixed design ANOVA with Context Difference (difference, no difference) as a between-participants factor was conducted.

The three-way interaction of congruency, proportion congruent, and context difference was significant, $F(1, 22) = 4.66, MSE = .053, p < .05, \eta_p^2 = .17$ (see Fig. 4). Although both groups of participants showed a significant interaction between congruency and proportion congruent [$F(1, 10) = 9.74, MSE = 0.117, p < .002, \eta_p^2 = .49$, and $F(1, 12) = 45.92, MSE = 0.508, p < .001, \eta_p^2 = .79$, for the difference and no-difference groups, respectively], the difference in the sizes of the congruency effect between contexts was larger for participants who did not notice any difference between the two contexts (39.54%) than for those who did notice a difference (20.60%). In both cases, the CSPC effects were driven by changes in accuracy on incongruent trials, which were more pronounced in the group of participants who did not notice a difference between the two contexts (8% and 21% for participants who noticed a difference and those who did not, respectively).

The majority of the participants (83%) considered the feedback information unreliable (see Table 3). At the end

of the experiment, they would report that the “incorrect response” feedback display appearing at the end of some trials (presumably incongruent trials) was not in accordance with their performance on those trials. For instance, a participant in a visual short (40 ms)–auditory long (120 ms) trial would respond “long,” getting an “incorrect response” feedback display. However, he or she would believe that his/her response was correct. Therefore, the information provided by the feedback in that case was considered unreliable by the participant. This points to the idea that the crossmodal effect reported here can be considered an audiovisual illusion, whereby the perceived duration of the visual event was driven by the actual duration of the auditory stimulus (cf. Soto-Faraco, Spence, & Kingstone, 2004).

Discussion

Three primary results were obtained in Experiment 2. First, we replicated the congruency effect described in Experiment 1. Second, and more importantly, this congruency effect was modulated by the congruency context. Specifically, the congruency effect was reduced in the low-proportion-congruent context relative to the high-proportion-congruent context. This contrasts with the results reported by Shore and Simic (2005), who failed to show a modulation of the visuotactile congruency effect by the proportion-of-congruency manipulation when the two inputs were presented in synchrony. Third, the modulation of the congruency effect was emphasized on the

Table 2 Percentages of participants who noticed or did not notice any difference between the two congruency contexts in Experiments 2–4

Experiment	Difference	No Difference
2	45.83	54.17
3	16.67	83.33
4	16.67	83.33

Table 3 Percentages of participants who considered the feedback information as reliable or unreliable in Experiments 2–4

Experiment	Unreliable	Reliable
2	83.33	16.67
3	79.17	20.83
4	66.67	33.33

second half of each context, presumably due to the learning of the association between the context of congruency and the proportion of congruent trials. SDT results did not show any effect on either the d' or the c index.

The results of the questionnaire used to assess whether participants were aware of the proportion-congruent manipulation appeared to indicate no evidence of participants' explicit knowledge of the congruency manipulation. What is more, participants who reported having noticed a difference between the two congruency contexts (whatever that difference was) showed a reduced CSPC effects with respect to those participants who did not notice a difference. This suggests that the different performance for the two contexts was not due to voluntary control, but rather to some form of cognitive control that was involuntarily cued by the context.

Crump and colleagues (Crump et al., 2006; Crump & Milliken, 2009; Crump et al., 2008) obtained their results using a within-blocks congruency manipulation, while the data from Experiment 2 came from a between-blocks congruency manipulation. We conducted Experiment 3 to test whether the same results could be obtained when the proportion of congruency was manipulated within blocks of trials.

Experiment 3

Method

Participants The participants were 36 undergraduate students (25 women; age range 18–30 years old, mean age 20 years) who received course credit in exchange for their participation.

Apparatus, stimuli, and procedure These were the same as in Experiment 2, with the exception that the two congruency contexts were presented within the same block of trials. The two halves of the computer screen (top/bottom) defined the two congruency contexts. As such, the white circle (3.01° in diameter) was presented below or above the fixation point (4.52° from the centre of the circle). The fixation point was presented in the centre of the screen, 15% lower with respect to Experiment 2, in order to define both contexts symmetrically. For half of the participants, the top context was associated with a high proportion of congruent trials (75%) and the bottom context was associated with a low proportion of congruent trials (25%). The reverse was true for the remaining half of the participants.

Results

Accuracy measures The data from 12 of the participants were removed from the analysis because their performance did not

reach 60% accuracy in the baseline condition. The remaining 24 participants (16 women, age range 18–30 years old, mean age 20 years) were included in the complete data analysis.

The mean accuracy for each participant and condition was submitted to a 2 (Congruency: congruent, incongruent) \times 2 (Proportion congruent: high, low) ANOVA, which revealed a significant main effect of congruency, $F(1, 23) = 99.22$, $MSE = 2.04$, $p < .001$, $\eta_p^2 = .81$. Responses on congruent trials (82% correct) were more accurate than responses on incongruent trials (53% correct). The main effect of proportion congruent was not significant, $F < 1$. In contrast with Experiment 2, the Congruency \times Proportion Congruent interaction did not reach statistical significance, $F < 1$. The sizes of the congruency effects were almost identical in the two contexts (28.9% and 29.3% in the low- and high-proportion-congruent conditions, respectively; see Table 1).

To examine any learning effect of the association between context and likelihood of congruency, the data in Experiment 2 were divided into two groups: the first four blocks and the last four blocks. The first 10 trials of the first block were again considered as part of the practice and were removed from the analysis. A 2 (Congruency: congruent, incongruent) \times 2 (Proportion congruent: high, low) \times 2 (Block: first half, second half) ANOVA showed that the three-way interaction of congruency, proportion congruent, and block did not reach statistical significance, $F(1, 23) = 1.27$, $MSE = .006$, $p = .27$, $\eta_p^2 = .05$ (see Fig. 5).

Neither the d' [$t(23) = 1.46$, $p = .16$] nor c [$t(23) = 1.27$, $p = .22$] indexes differed between the high- and low-proportion-congruent conditions.

Questionnaire measures Only 4 (17%) of the participants noticed that one of the contexts differed from the other, but no participants were specifically aware of the context manipulation (see Table 2). Instead, the difference between the two contexts was attributed to chance or to the fact that one of the locations was simply easier to respond to than the

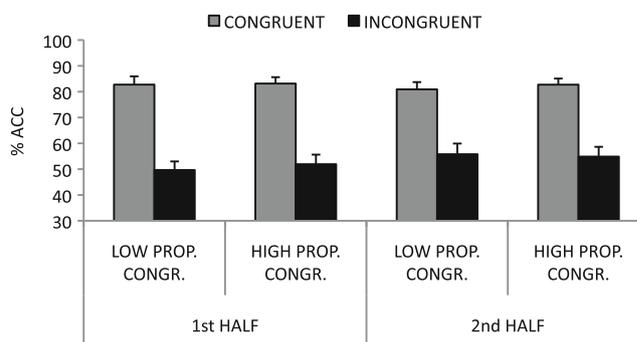


Fig. 5 Percentages of correct responses as a function of congruency, proportion congruent, and block in Experiment 3. Error bars represent the standard errors of the means across participants

other. As in Experiment 2, the majority of the participants (79%) considered the information provided by the feedback unreliable (see Table 3).

Discussion

In Experiment 3, the congruency effect was not modulated by the congruency context manipulation, which contrasts with the results obtained in Experiment 2. One possibility is that the lack of a significant interaction between congruency and proportion congruent in Experiment 3 owes to a weak context manipulation, which prevented the association between the spatial location and the likelihood of congruency. Note that Crump et al. (2006) suggested that a learned association between the context in which the stimuli are presented and the likelihood of congruency is crucial to obtaining a context-specific proportion-congruent effect. To test the hypothesis that the context manipulation was simply too weak to be learned in Experiment 3, the distinction between the two congruency contexts was made more salient in Experiment 4.

Experiment 4

Method

Participants The participants were 52 undergraduate students (47 women; age range 18–48 years old, mean age 22 years) who received course credit in exchange for their participation.

Apparatus, stimuli, and procedure These were the same as in Experiment 3, but in this experiment the white circle (3.01° in diameter) was presented 6.02° (from the centre of the circle) either below or above a central horizontal line, rather than 4.52° below or above a central fixation point.

Results

Accuracy measures The data from 16 of the participants² were removed from the analysis because their performance

² A 2 (Congruency: congruent, incongruent) × 2 (Proportion congruent: high, low) ANOVA was conducted on the data from all of the participants in Experiments 2–4. In Experiment 2, the interaction between congruency and proportion congruent was significant [$F(1, 31) = 46.14$, $MSE = 0.67$, $p < .001$, $\eta_p^2 = .60$]. In Experiment 3, this interaction was not significant [$F(1, 35) = 2.37$, $MSE = 0.007$, $p = .133$, $\eta_p^2 = .06$]. In Experiment 4, the interaction was marginally significant [$F(1, 51) = 3.70$, $MSE = 0.01$, $p = .06$, $\eta_p^2 = .06$]. Therefore, the results from Experiments 2–4 did not change substantially when all of the participants were included. In any case, we maintained the exclusion criterion to ensure that our participants understood the task and were responding above chance in the baseline condition.

did not reach the 60% accuracy criterion in the baseline condition. The remaining 36 participants (32 women; age range 18–48 years old, mean age 22 years) were included in the complete data analysis.

The mean accuracy for the remaining participants in each experimental condition was submitted to a 2 (Congruency: congruent, incongruent) × 2 (Proportion congruent: high, low) ANOVA. This analysis revealed a significant main effect of congruency, $F(1, 35) = 46.23$, $MSE = 1.71$, $p < .001$, $\eta_p^2 = .57$. Responses on congruent trials (77% correct) were more accurate than responses on incongruent trials (56% correct). The main effect of proportion congruent was significant, $F(1, 35) = 5.53$, $MSE = 0.007$, $p = .024$, $\eta_p^2 = .14$. Participants performed better for the low-proportion-congruent context (67.1%) than for the high-proportion-congruent context (65.7%). Crucially, the interaction between congruency and proportion congruent was significant, $F(1, 35) = 4.75$, $MSE = 0.01$, $p = .036$, $\eta_p^2 = .12$. Participants performed more accurately on congruent than on incongruent trials for both proportion-congruent conditions, $F(1, 35) = 33.42$, $MSE = 0.71$, $p < .001$, and $F(1, 35) = 54.39$, $MSE = 1.01$, $p < .001$, for high and low congruency, respectively. However, in line with the results of Experiment 2, the size of the congruency effect was smaller for the low-proportion-congruent context (20%) than for the high-proportion-congruent context (24%), $t(36) = 2.18$, $p < .04$ (see Table 1). A priori comparisons revealed that this difference was due to increased accuracy on incongruent trials (3.4%), $F(1, 35) = 7.89$, $MSE = 0.02$, $p = .008$, in the low-proportion-congruent context relative to the high-proportion-congruent context. There was no statistical difference between the two contexts on congruent trials, $F < 1$.

A 2 (Congruency: congruent, incongruent) × 2 (Proportion congruent: high, low) × 2 (Block: first half, second half) ANOVA was also conducted for Experiment 4 (see Fig. 6). The first 10 trials were excluded from the analysis. The three-way interaction between congruency, proportion

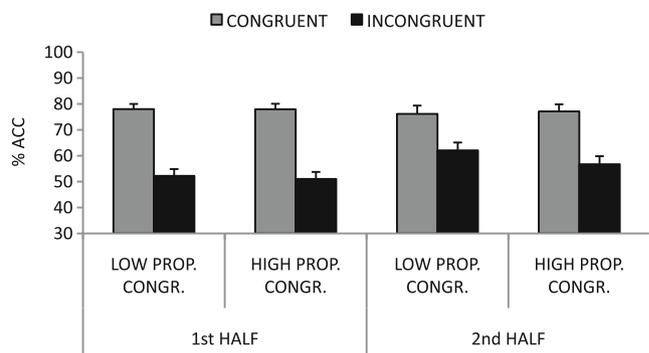


Fig. 6 Percentages of correct responses as a function of congruency, proportion congruent, and block in Experiment 4. Error bars represent the standard errors of the means across participants

congruent, and block did not reach statistical significance, $F(1, 35) = 1.79$, $MSE = 0.011$, $p = .19$, $\eta_p^2 = .048$. However, post hoc analyses showed a significant interaction between congruency and proportion congruent for the second half of the experiment, $F(1, 35) = 5.16$, $MSE = 0.036$, $p < .03$, $\eta_p^2 = .13$, while this interaction was not significant for the first half of the experiment, $F < 1$. A priori comparisons revealed, in line with the results of Experiment 2, a significant difference between accuracy on incongruent trials in the second half of the experiment, $F(1, 35) = 12.02$, $MSE = 0.052$, $p < .002$, with better performance (5.36%) for the low-proportion-congruent context than for the high-proportion-congruent context. In the first half of the experiment, this difference was not significant, $F < 1$.

In contrast to the results of Experiment 2, pair-wise two-tailed t tests conducted between high and low proportions congruent for the d' and c indexes showed that d' was significantly higher in the low-proportion-congruent condition (2.24) than in the high-proportion-congruent condition (1.96), $t(35) = 2.35$, $p = .02$, while no differences were obtained for c (0.28 for high, 0.32 for low, $t < 1$).

Questionnaire measures Only 6 participants (17%) noticed a difference between the two contexts, although none of them was aware of the specific proportion-congruent manipulation (see Table 2). Instead, the perceived difference was explained with comments related to tiredness and body position.

To examine differences in performance between the participants who reported any difference between the two contexts and participants who did not, a 2 (Congruency: congruent, incongruent) \times 2 (Proportion congruent: high, low) ANOVA was conducted for each group separately, taking into consideration the large disparity in the numbers of participants who belonged to each group (6 difference and 30 no difference; see Fig. 7). In line with Experiment 2, the analyses revealed a significant interaction between congruency and proportion congruent only for the group who did not report any

difference between the two contexts, $F(1, 29) = 5.87$, $MSE = 0.018$, $p < .03$, $\eta_p^2 = .168$, while the interaction did not reach statistical significant for the group of participants who did notice differences between the contexts, $F < 1$. In line with Experiment 2, a priori comparisons revealed a statistically significant difference between performance on incongruent trials in the low-proportion-congruent context (59%) relative to the high-proportion-congruent context (55%) for participants who did not report any difference between the two contexts, $F(1, 29) = 11.70$, $MSE = 0.027$, $p < .002$. Performance on incongruent trials between the two contexts did not reach a statistically significant difference in the group of participants who did report a difference between the two contexts, $F < 1$.

Once again, participants' reports at the end of the experiment revealed that the feedback information was considered unreliable by 67% of the participants (see Table 3).

Discussion

Experiment 4 revealed that increasing the spatial separation between the two congruency contexts relative to Experiment 3 produced the expected result: a larger congruency effect in the high-proportion-congruent context than in the low-proportion-congruent context. As in Experiment 2, the data from the questionnaire suggested that the effect was not due to voluntary, conscious control. Furthermore, given that observers could not predict whether the oncoming trial would belong to the high- or the low-proportion-congruent context, they could not prepare themselves in advance to exert more or less attentional control. We argue that the particular location in which the stimuli were presented became associated with a particular level of congruency, which in turn led to different control sets being cued in the two contexts.

In Experiment 4, a significant increment in sensitivity (d') in the low-proportion-congruent context was obtained. This result suggests that the congruency manipulation modulated audiovisual interactions at a perceptual level of information processing when different proportions of congruent and incongruent trials were presented mixed up in the same block of trials. This perceptual sensitivity enhancement could be related to the better performance on incongruent trials in the low-proportion-congruent condition. However, given that the same modulation of the d' index was not obtained in Experiment 2, this result should be taken with caution. A potential, and speculative, explanation refers to the trial-by-trial shifts in attentional control set that would be required in Experiment 4 in contrast to Experiment 2, in which a constant congruency proportion was presented on each block of trials. This could explain the differential influences on the d' index. Importantly, the criterion index (c) in Experiment 4 was not modulated by the context-of-congruency manipulation.

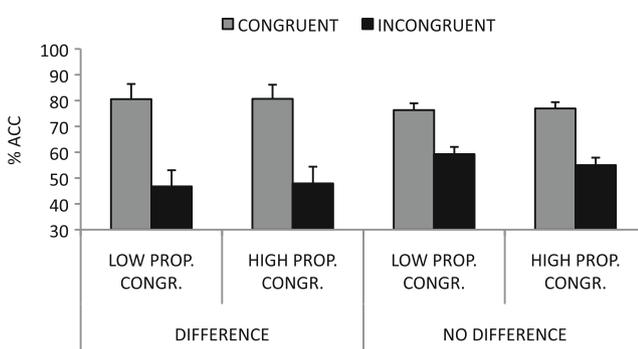


Fig. 7 Percentages of correct responses as a function of congruency, proportion congruent, and context difference in Experiment 4. Error bars represent the standard errors of the means across participants

Feedback in Experiments 1–4 was presented on the assumption that it would give information about the proportion of congruency, facilitating the association between the context and the likelihood of congruency. However, it is clear that participants were not aware of the proportion-of-congruency manipulation in any of the experiments reported in this article. What is more, once again the majority of participants in Experiment 4 considered the feedback information to be unreliable. Therefore, feedback does not seem to have played a key role in our findings.

General discussion

The purpose of this study was to examine the effect of the congruency context on the interaction of auditory and visual stimuli. The audiovisual interaction was measured by means of a task in which participants had to discriminate the duration of a visual stimulus while ignoring the duration (congruent or incongruent) of a synchronously presented auditory stimulus. The congruency context was manipulated by varying the relative proportions of congruent and incongruent trials—between blocks in Experiment 2, within blocks in Experiments 2 and 4.

The results of Experiment 1 showed that our stimuli were suitable for investigating crossmodal congruency effects similar to those that have been reported in previous studies (e.g., Andersen, Tiippana, & Sams, 2004; McGurk & MacDonald, 1976). Note that the results of Experiment 1 are consistent with the outcome of a recent study reported by Klink et al. (2011), who showed that the perceived duration of a visual stimulus depended on the duration of an auditory stimulus presented synchronously. Accordingly, given that audition dominates vision in the processing of temporal information, we argue that perceived visual duration in our study was biased towards the perceived duration of the auditory input (e.g., Romei, De Haas, Mok, & Driver, 2011; Shams et al., 2000, 2002; Walker & Scott, 1981). Our SDT analyses pointed out that the congruency effect measured here reflects crossmodal interactions occurring at perceptual and postperceptual levels of stimulus processing, in accordance with previous accounts (e.g., Sanabria, Spence, & Soto-Faraco, 2007)

The results of Experiment 2 revealed that the proportion of congruent trials, manipulated in separate blocks of trials, influenced the size of the congruency effect. Crucially, Experiment 4 demonstrated that the proportion of congruency could influence the congruency effect even when different proportions were associated with two contexts presented within the same block of trials. Taken together, these results suggest that the audiovisual interactions measured by the congruency effect are prone to top-down attentional modulations related to the context of congruency, in

contrast to the findings of previous accounts (see Shore & Simic, 2005).

The results reported here are in accord with those of Crump et al. (2006); that is, a larger crossmodal congruency effect was observed for the context in which most of the trials were congruent. We suggest that, in terms of cognitive control, this effect hinges on context-sensitive adjustments as a function of audiovisual congruency. In our study, for each location (either above or below the fixation point), the particular congruency proportion defined whether the auditory input would share its duration with the visual target input. As such, we argue that the low-proportion-congruent context triggered an attentional set for filtering the visual input from the temporally incongruent auditory input. In contrast, in the high-proportion-congruent context, the auditory input shared the same temporal parameters as the visual input on most occasions, so that attentional filtering was not generally required—or at least, not to the same extent as in the low-proportion-congruent condition. It appears, then, that context can drive the attentional set that modulates the way in which audiovisual inputs interact. The outcome of Experiment 4 demonstrated that these shifts in attentional set are highly flexible and can occur on a trial-by-trial basis.

The results of Experiments 3 and 4 suggest that distinct location contexts are necessary for learning the association between those contexts and different congruency proportions. In Experiment 3, it seems likely that the two locations were not sufficiently distinct to be treated as separate contexts, and instead were processed as one. The greater distinctiveness of the two contexts in Experiment 4 increased the likelihood that they would be selectively attended to as separate contexts. Thus, participants were able to learn the association between location and proportion congruency, which resulted in the context-specific proportion-congruent effect observed in Experiment 4 but not in Experiment 3.

The results in Experiments 3 and 4 are consistent with the idea that implicit learning can depend on attention to the task-relevant dimensions (e.g., Jiménez & Méndez, 1999). Jiménez and Méndez showed the difficulty of associating a shape with a location if the shape was task-irrelevant, given that attention was not focused on the shape dimension. Crucially, the distinct location contexts used in Experiment 4 were more likely to require shifts of attention than were those used in Experiment 3, and thus the association between location and proportion congruent was also more likely to be learned in Experiment 4 than in Experiment 3 (see also Crump et al., 2008). The improvement of participants' performance on incongruent trials in the low-proportion-congruent condition (with respect to the high-proportion-congruent condition) in the second, as compared to the first, half of Experiments 2 and 4 strengthens the argument that the association between context and proportion congruency has to be learned. The fact that this improvement was absent

in Experiment 3 bolsters the idea that attentional shifts are necessary to set this association, not simply experience within a particular context of congruency. Given that participants were not aware of the association between the location of the stimuli and audiovisual congruency, we argue that this association occurred implicitly.

The subjective measures collected in our study suggest that the attentional mechanism responsible for the reduction of the congruency effect in the low-proportion-of-congruency context was involuntarily triggered by the stimulus characteristics, which entailed rapid shifts after stimulus onset. In fact, the large majority of participants in Experiments 2 and 4 did not notice the context manipulation. Moreover, when the group of participants was split into those who noticed some difference between the two contexts of congruency and those who did not, the former showed decreased modulations of the congruency effect. Noticing a difference between the two contexts (whatever that difference was) might have resulted in participants using different task strategies for each congruency context. This would have interfered with the control mechanism automatically triggered by the onset of the stimulus in each particular context.

Previous studies have shown that contextual cues can control selective attention processes during online performance in a fast and stimulus-driven manner (see Egner, 2008). In line with this idea, each location in Experiments 2 and 4 might have constituted a contextual cue that triggered the attentional control needed to select relevant information. Given that the location contexts were mixed at random across trials in Experiment 4, this form of control could not be implemented prior to stimulus onset, but instead selection of a particular control set would be cued by the onset of the stimulus in either one context or the other. By this view, context can cue control shifts involuntarily during online performance.

In sum, our study shows, for the first time, that the context of congruency, resulting in shifts in attentional control set, can modulate the outcomes of audiovisual interactions. We propose that this modulation was caused by selective attentional processes involved in the association between a context and the likelihood of congruency, presumably in an “automatic”, context-driven manner. Our study therefore contributes to the ongoing debate regarding the role of attention in multisensory perception, suggesting a crucial role for attention in crossmodal processing.

Author note This study was supported by a Discovery research grant from the Natural Sciences and Engineering Research Council of Canada to D.I.S.; a second Discovery research grant from the Natural Sciences and Engineering Research Council of Canada, to B.M.; a student scholarship (AP2008-03662) from the Ministerio de Educación y Ciencia to B.R.S.; and Grants SEJ2007-63645 and PSI2010-19655, from the Ministerio de Educación y Ciencia, and Grant SEJ-6414, from the Junta de Andalucía, to D.S. We thank Sarah Lade for her help with data collection.

References

- Alink, A., Singer, W., & Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *Journal of Neuroscience*, *28*, 2690–2697. doi:10.1523/JNEUROSCI.2980-07.2008
- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*, 301–308. doi:10.1016/j.cogbrainres.2004.06.004
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546. doi:10.1016/j.tics.2004.10.003
- Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Crump, M. J. C., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review*, *13*, 316–321. doi:10.3758/BF03193850
- Crump, M. J. C., & Milliken, B. (2009). The flexibility of context-specific control: Evidence for context-driven generalization of item-specific control settings. *Quarterly Journal of Experimental Psychology*, *62*, 1523–1532. doi:10.1080/17470210902752096
- Crump, M. J. C., Vaquero, J. M. M., & Milliken, B. (2008). Context-specific learning and control: The role of awareness, task relevance, and relative salience. *Consciousness and Cognition*, *17*, 22–36. doi:10.1016/j.concog.2007.01.004
- Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*, 374–380. doi:10.1016/j.tics.2008.07.001
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*, 1784–1790. doi:10.1038/nn1594
- Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, *29*, 1247–1257. doi:10.1111/j.1460-9568.2009.06688.x
- Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 236–259. doi:10.1037/0278-7393.25.1.236
- Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception, & Psychophysics*, *73*, 219–236. doi:10.3758/s13414-010-0010-9
- Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 219–234. doi:10.1037/0096-1523.20.2.219
- Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*, 166–174. doi:10.3758/BF03197535
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Frontiers in Perception Science*, *2*(215), 1–8. doi:10.3389/fpsyg.2011.00215
- Sanabria, D., Soto-Faraco, S., & Spence, C. (2007). Spatial attention and audiovisual interactions in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 927–937. doi:10.1037/0096-1523.33.4.927
- Sanabria, D., Spence, C., & Soto-Faraco, S. (2007). Perceptual and decisional contributions to audiovisual interactions in the perception of

- apparent motion: A signal detection study. *Cognition*, *102*, 299–310. doi:10.1016/j.cognition.2006.01.003
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788. doi:10.1038/35048669
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). A visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147–152. doi:10.1016/S0926-6410(02)00069-1
- Shore, D. I., & Simic, N. (2005). Integration of visual and tactile stimuli: Top-down influences require time. *Experimental Brain Research*, *166*, 509–517. doi:10.1007/s00221-005-2391-x
- Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 330–345. doi:10.1037/0096-1523.30.2.330
- 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*, *17*, 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*, 400–410. doi:10.1016/j.tics.2010.06.008
- Walker, J. T., & Scott, K. J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1327–1339. doi:10.1037/0096-1523.7.6.1327
- Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, *37*, 572–578.