BRILL

Seeing and Perceiving 24 (2011) 369–389

SEEING AND PERCEIVING
brill.nl/sp

Capture of Intermodal Visual/Tactile Apparent Motion by Moving and Static Sounds *

Lihan Chen [1,2] **and Xiaolin Zhou** [1,2,**]

[1] Center for Brain and Cognitive Sciences and Department of Psychology, Peking University, Beijing 100871, China

[2] Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China

Received 13 January 2011; accepted 20 May 2011

Abstract

Apparent motion can occur within a particular modality or between modalities, in which a visual or tactile stimulus at one location is perceived as moving towards the location of the subsequent tactile or visual stimulus. Intramodal apparent motion has been shown to be affected or 'captured' by information from another, task-irrelevant modality, as in spatial or temporal ventriloquism. Here we investigate whether and how intermodal apparent motion is affected by motion direction cues or temporal interval information from a third modality. We demonstrated that both moving and asynchronous static sounds can capture intermodal (visual–tactile and tactile–visual) apparent motion; moreover, while the auditory direction cues have less impact upon the perception of intramodal visual apparent motion than upon the perception of intramodal tactile or intermodal visual/tactile apparent motion, the auditory temporal information has equivalent impacts upon both intramodal and intermodal apparent motion. These findings suggest intermodal apparent motion is susceptible to the influence of dynamic or static auditory information in similar ways as intramodal visual or tactile apparent motion.

Keywords

Ventriloquism, apparent motion, intermodal interaction, crossmodal capture

1. Introduction

Multisensory interaction has often been demonstrated through intersensory bias in which the percept is predominantly defined according to properties of information

* This article is part of the Multisensorial Perception Collection, guest edited by S. Wuerger, D. Alais and M. Gondan.

** To whom correspondence should be addressed. Department of Psychology, Peking University, Beijing 100871, China; e-mail: xz104@pku.edu.cn

 DOI:10.1163/187847511X584434

from one particular modality. Intersensory bias occurs when information available to different sensory modalities is discrepant along some critical dimensions and the perceptual system attempts to maintain normal perception in face of this discrepancy (Welch and Warren, 1980, 1986). One typical example of intersensory bias is the ventriloquism illusion, where the spatial locations or temporal occurrences of stimuli in one modality can be affected or captured by events arising in another modality (Bertelson, 1999; Bertelson and Aschersleben, 1998; Caclin *et al.*, 2002; Freeman and Driver, 2008; Keetels and Vroomen, 2008; Morein-Zamir *et al.*, 2003; Shi *et al.*, 2010; Soto-Faraco *et al.*, 2004a, b; Vroomen and de Gelder, 2004; Vroomen and Keetels, 2006; Welch and Warren, 1980). In the spatial domain, the localization of auditory or tactile stimuli can be biased by simultaneous presentation of a visual stimulus at a different spatial position; the apparent location of a sound can be biased towards the location of a tactile stimulus when they are synchronous (Bertelson, 1999; Bertelson and Aschersleben, 1998; Caclin *et al.*, 2002). In the nonspatial domain, task-irrelevant sounds can influence the temporal order perception of lights, where sounds intervening between two lights lead to a decline in performance by ‘pulling’ the lights closer and sounds outside of the lights improve performance by ‘pulling’ the lights apart (Morein-Zamir *et al.*, 2003; Shi *et al.*, 2010).

The ventriloquism effect has been found not only for static events, but also in apparent motion tasks, in which judgment of motion direction of stimuli in one modality is affected by repetition of static stimulus or stream of moving stimuli in another modality. Apparent motion was first demonstrated in visual modality with two briefly flashing lights (Exner, 1875). When presented at two different spatial locations with an appropriate temporal interval, the perceiver would see a light moving from the first to the second location. Since this very first demonstration, apparent motion has been found for other modalities and for crossmodal events (Galli, 1932; Harrar *et al.*, 2008; Zapparoli and Reatto, 1969). The phenomenon of stimuli in one modality affecting the perceived direction of apparent motion in another modality has been termed ‘crossmodal dynamic capture’ (Soto-Faraco *et al.*, 2002). Typically, stimuli presented in the distractor modality are congruent or incongruent with stimuli in the target modality in terms of motion direction or synchrony and participants are instructed to judge the motion direction of stimuli in the target modality while trying to ignore the stimuli in the distractor modality (Slutsky and Recanzone, 2001; Soto-Faraco *et al.*, 2002, 2004a, b).

Crossmodal dynamic capture has also been found in studies manipulating the temporal relation between stimuli from two different modalities. Judgment of motion direction or categorization of the type of apparent motion in a target modality are altered by the presence of temporal asynchronous static stimuli from another modality (Getzmann, 2007; Freeman and Driver, 2008; Shi *et al.*, 2010). Freeman and Driver (2008), for example, found that when two flashing bars are presented in alternation in two hemifields with equal intervals between two consecutive flashes, auditory beeps slightly lagging or leading the flashes strongly influence the perceived visual motion direction, although the biaurally presented beeps provide no spatial information relevant to the flashes. In this study, the timing of visual events is altered by auditory stimuli with higher temporal acuity, and this ‘temporal ventriloquism’ has been observed in a number of other studies (Bertelson, 1999; Bertelson and Aschersleben, 1998; Getzmann, 2007; Keetels and Vroomen, 2008; Morein-Zamir *et al.*, 2003; Vroomen and de Gelder, 2004; Vroomen and Keetels, 2006).

An interesting finding in the previous studies is that the crossmodal dynamic capture effect is asymmetric between the auditory, visual and tactile modalities. The direction of visual stimuli can capture the direction of auditory apparent motion but it has not yet been shown that the direction cues in the auditory stimuli can capture the direction of visual apparent motion (Soto-Faraco *et al.*, 2002; Soto-Faraco *et al.*, 2004b; Strybel and Vatakis, 2004). Likewise, the tactile motion distractors have a stronger influence upon the perception of auditory motion direction than auditory motion distractors do on tactile perception (Soto-Faraco *et al.*, 2004a), and the visual motion distractors have a stronger influence on the perception of tactile motion direction than tactile motion distractors do on visual perception (Bensmaïa *et al.*, 2006; Craig, 2006; Lyons *et al.*, 2006). These asymmetries in both spatial and temporal manipulations have been attributed to differences in functional appropriateness and precision between different modalities. Vision is described as the most accurate in spatial tasks (Welch and Warren, 1980, 1986) and, thus, in general visual direction cues can impose influences upon the perception of the direction of apparent motion in other modalities, as summarized by Soto-Faraco *et al.* (2003). Conversely, audition takes priority in temporal precision, with auditory temporal information calibrating the perception of temporal intervals between visual events (Freeman and Driver, 2008; Kafaligonul and Stoner, 2010; Shi *et al.*, 2010). However, the functional superiority of vision in spatial localization as well as the dominance of temporal precision in audition could be task-dependent. In a spatial localization task, when visual stimuli are severely blurred, sound captures vision in spatial localization rather than *vice versa* (Alais and Burr, 2004).

Apparent motion takes place not only within a single modality, but also between modalities (Harrar and Harris, 2007; Harrar *et al.*, 2008). For example, with an appropriate time interval between a visual stimulus at one location and a tactile stimulus at another location, the participants would perceive some kind of motion stream from the first to the second location. In this kind of intermodal apparent motion, the motion stream is composed of stimuli from two different modalities, with different combinations having different perceptual saliencies or perceived strengths (Harrar and Harris, 2007; Harrar *et al.*, 2008). This differential saliency can also be found in comparisons between intermodal and intramodal apparent motion. For example, Harrar *et al.* (2008) varied the distance and stimulus onset asynchrony (SOA) between two stimuli, which could be light and/or touch, and asked the participants to assess how good the perceived apparent motion was on a 5-point scale. The perceived quality of apparent motion varied, with the preferred SOA increasing as the distance between the stimuli increased (i.e., consistent with Korte's third law of apparent motion, Korte, 1915) for visual–visual apparent motion, but with no such functional relationship for visual–tactile apparent motion.

Is intermodal apparent motion affected by spatial or temporal information from a task-irrelevant modality in the same way as intramodal apparent motion? This issue is important because answers to this question would help us to understand the mechanisms underlying crossmodal dynamic capture. Here we adapted the paradigm developed by Soto-Faraco and colleagues (Soto-Faraco *et al.*, 2002; Soto-Faraco *et al.*, 2004a, b) and presented participants with visual and tactile stimuli at two different locations (Fig. 1A) to create intermodal (visual–tactile, tactile–visual) and intramodal (visual–visual, tactile–tactile) apparent motion. Crucially, we presented participants with task-irrelevant auditory stimuli, which encoded direction information congruent or incongruent with the direction of intermodal or intramodal apparent motion (Experiment 1) or temporal order information synchronized or asynchronized with the SOA for intermodal or intramodal stimuli (Experiment 2). The empirical question was whether and how intermodal apparent motion could be affected by motion direction cues or temporal order information from the third, auditory modality. The potential intermodal capture effects were compared with the typical crossmodal capture effects between two individual modalities.

2. Experiment 1

The stimulus consecutively presented at the first or the second location (Fig. 1A) could be either visual (light-emitting diode or LED flash) or tactile (indentation tap onto the finger tip), creating four combinations: visual–visual (VV), visual–tactile (VT), tactile–visual (TV) and tactile–tactile (TT). With an inter-stimulus interval (ISI) of 100 ms between the two stimuli in this experiment, participants perceived either the intramodal (VV and TT) or the intermodal (VT and TV) rightward or leftward apparent motion (Harrar and Harris, 2007; Harrar *et al.*, 2008). The task-irrelevant auditory stimuli, presented consecutively from two speakers located at spatial positions aligned with the visual and tactile stimulations, formed another stream of apparent motion whose direction was either congruent or incongruent with the direction of intermodal or intramodal apparent motion. Participants were instructed to judge the direction of the intermodal or intramodal apparent motion while ignoring the auditory input.

The auditory stream could appear at the same time as the intermodal or intramodal apparent motion or could be delayed by 500 ms. This manipulation of delay was to provide a baseline condition in which the auditory information was outside of the normal temporal window in which multisensory integration could take place (Bertelson and Aschersleben, 1998; Soto-Faraco *et al.*, 2004a). Given the different motion saliencies for different modality combinations (Harrar *et al.*, 2008; see also Discussion for this experiment) and different functional dominances in pro-

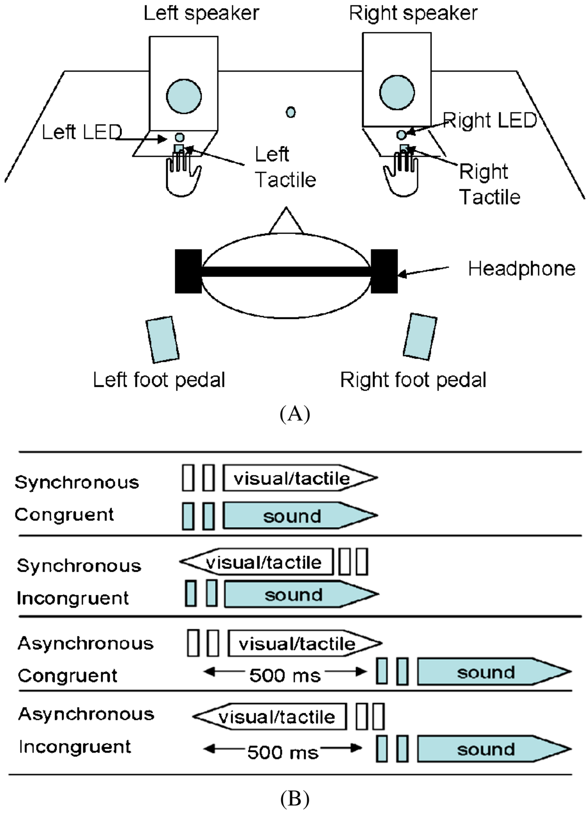

Figure 1. Experimental setup and temporal correspondence of motion streams used in Experiment 1. (A) The participant placed two middle fingers on the tactile actuators which were embedded into foams, which were placed just in front of the two speakers. Two LEDs were collocated with the two actuators, respectively. One red LED was placed at the center of the setup to serve as a fixation point. Participants made their responses by lifting left foot pedal for leftwards target motion or right foot for rightward motion. Accuracy rather than speed was emphasized. (B) Spatial and temporal correspondences between auditory input and visual/tactile target stimuli. The auditory beeps could occur either congruently or incongruently with the target motion stream, simultaneously or 500 ms later with respect to the visual/tactile targets. This figure is published in color in the online version.

cessing spatiotemporal information for different crossmodal interactions (Fujisaki and Nishida, 2009; Welch and Warren, 1980), we expected to observe differential auditory capture effects for the intermodal and intramodal apparent motion.

3. Method

3.1. Participants

Fourteen undergraduate and graduate students (5 females, average age 24.5 years) were tested. None of them reported any history of somatosensory or auditory deficits. They had normal or corrected-to-normal vision and were naïve to the purpose of this study. The experiment was performed in compliance with institutional guidelines set by Academic Affairs Committee, Department of Psychology at Peking University.

3.2. Apparatus and Stimuli

Two speakers were placed 30 cm from each other (center to center; see Fig. 1A). The tactile stimuli were produced using solenoid actuators in which the embedded cylinder metal tips, when the solenoid coils were magnetized, would tap the fingers to induce indentation taps (Heijo Research Electronics, UK). The maximum contact area is about 4 mm^2 and the maximum output is 3.06 W. Two solenoid actuators were put onto two foams that were laid directly in front of the speakers. Directly in front of each actuator there was a green LED. Thus, the presentations of auditory, visual and tactile stimuli could be essentially at the same spatial positions. The inputs to the LED (5 V, with duration of 50 ms) and to the actuators (with duration of 50 ms) were controlled *via* a parallel LPT port by software written with Matlab (Mathworks Inc.). Two footpedals attached to the floor directly beneath the participants' left and right feet were used to collect judgment responses.

An auditory apparent motion stream consisted of the presentation of two 50-ms tones (65 dB) with an ISI of 100 ms. The two tones used in a given apparent motion stream were of the same frequency, which was chosen randomly on a trial-by-trial basis from three possible frequencies: 450, 500 and 550 Hz. Likewise, a visual or a tactile stimulus lasted for 50 ms, with the ISI of 100 ms between the two stimuli. Participants were asked to wear a headset to prevent from hearing the faint noise (30 dB) from the actuators. However, the loudness of the auditory stimuli from the two speakers were intense enough (65 dB) such that the participants had no difficulty in perceiving the sounds.

3.3. Design and Procedures

A 4 (type of motion stream in the target modality: VV, VT, TV, TT) $\times$ 2 (congruency between the direction of apparent motion in the target modality and the direction of apparent motion in the auditory stream: congruent or incongruent) $\times$ 2 (temporal correspondence between the two streams: synchronous or with a delay of 500 ms) factorial design was adopted (see Fig. 1B). The experiment had a total of 384 trials, divided into six test blocks with each block having 4 trials from each experimental condition.

Participants sat in front of the two speakers at a distance of 40 cm and were instructed to keep their middle fingers on the tactile actuators (Fig. 1A) and their feet pressed down on the foot pedals. The room was kept dark throughout the experiment. Participants started each trial by lifting and putting down one foot. After an interval of 1300 ms, stimuli for the target stream and for the auditory stream (for synchronous presentation) were presented. After a pause of 750 ms participants made their judgment by lifting one foot corresponding to the direction of apparent motion in the target stream: left foot for leftward and right foot for rightward. After a foot response was made, a red LED at the middle position between the two speakers flashed for 200 ms and participants were asked to fixate at this position for the next trial.

Before the formal experiment, participants practiced discriminating the direction of apparent motion in the target stream in the absence of any auditory stimuli. All the types of apparent motion used in this experiment, including intramodal (VV, TT) and intermodal (VT, TV) stimuli, were presented in the practice session. Participants were asked after practice whether they perceived motion between the two consecutive stimuli in the visual and/or tactile modalities. They all reported to have a strong sense of motion for each type of stimuli (see also Discussion). All the participants reached the criterion of at least 90% correct judgments in their first 20 attempts.

4. Results

The average proportion of correct responses (with the associated standard error) for each condition is presented in Table 1. An analysis of variance (ANOVA) with motion type, congruency and synchrony as three within-participant factors found a significant main effect of congruency, $F(1, 13) = 25.89$, $p < 0.001$, $\eta_p^2 = 0.67$, with the congruent stimuli being judged more accurately (94.3%) than the incongruent stimuli (85.8%). The main effect of temporal synchrony was also significant, $F(1, 13) = 5.70$, $p < 0.05$, $\eta_p^2 = 0.31$, with the percentage of correct responses lower for the temporally synchronous presentation of auditory stimuli (88.7%) than for the delayed presentation of auditory stimuli (91.4%). Importantly, the main effect of motion type was significant, $F(3, 39) = 6.10$, $p < 0.01$, $\eta_p^2 = 0.35$. Bonferroni-corrected pairwise comparisons showed that the percentage of correct responses for VV motion (95.5%) was higher (ps < 0.05) than those for the other three types of motion which did not show difference between themselves (VT, 87.9%; TV, 87.5%; TT, 89.3%).

Moreover, the two-way interaction between motion type and synchrony was significant, $F(3, 39) = 10.94$, $p < 0.001$, $\eta_p^2 = 0.46$, as was the two-way interaction between congruency and synchrony, $F(3, 39) = 35.07$, $p < 0.001$, $\eta_p^2 = 0.73$ (Fig. 2A), and the three-way interactions between motion type, congruency and synchrony, $F(3, 39) = 5.27$, $p < 0.01$, $\eta_p^2 = 0.29$ (Fig. 2B). Further analysis showed that, for the delayed presentation of auditory stimuli, the main effects of motion type and congruency were both significant, $F(3, 39) = 5.62$, $p < 0.01$, $\eta_p^2 = 0.30$, and $F(1, 13) = 6.65$, $p < 0.05$, $\eta_p^2 = 0.34$; but interaction between the two factors was not, $F(3, 39) = 1.02$, $p > 0.1$, $\eta_p^2 = 0.07$. For the synchronous presentation of auditory stimuli, the main effects of motion type and congruency were significant, $F(3, 39) = 9.76$, $p < 0.001$, $\eta_p^2 = 0.43$, $F(1, 13) = 32.0$, $p < 0.001$, $\eta_p^2 = 0.71$, so was the interaction between the two factors, $F(3, 39) = 4.58$, $p < 0.01$, $\eta_p^2 = 0.26$. These analyses demonstrated that the congruency between the motion direction of the task-irrelevant auditory stream affected the judgment of motion direction in the

Table 1.
Percentages of correct responses in Experiment 1, with numbers in brackets indicating standard errors (VV: visual–visual motion; VT: visual–tactile motion; TV: tactile–visual motion; TT: tactile–tactile motion)

	VV		VT		TV		TT	
	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Synchronous	99.7	92.6	95.5	79.5	93.2	80.7	95.2	73.2
	(0.3)	(3.0)	(1.3)	(4.0)	(2.0)	(4.9)	(1.7)	(3.5)
Asynchronous	97.0	92.6	90.5	86.3	89.9	86.3	93.8	94.9
	(0.8)	(4.0)	(2.8)	(3.0)	(2.5)	(3.6)	(1.9)	(1.2)

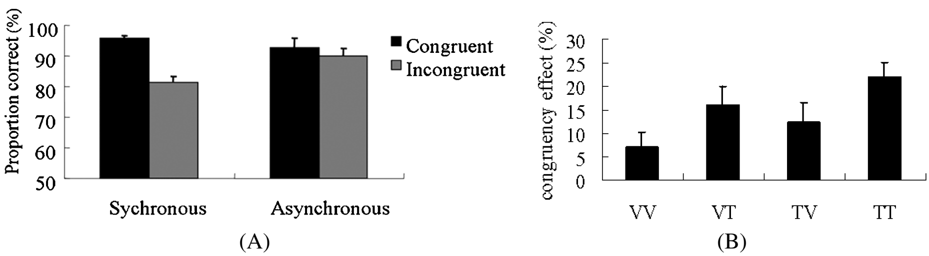

Figure 2. Percentages of correct responses in judging the direction of target apparent motion streams in Experiment 1. The error bars represent standard errors in different conditions. (A) Percentages of correct responses collapsed over types of target streams for synchronous and asynchronous presentations. (B) The congruency effects (incongruent *vs* congruent) for the four types of apparent motion in the synchronous condition (VV: visual–visual motion; VT: visual–tactile motion; TV: tactile–visual motion; TT: tactile–tactile motion).

target stream and this congruency effect was larger when the auditory stimuli was presented simultaneously with the target stimuli than when the auditory stimuli was delayed. Moreover, for different types of apparent motion in the target stream, the congruency effect remained the same when the auditory stimuli were delayed but was different when the auditory stimuli was synchronous. Furthermore, it is clear from Table 1 that the congruency effect was the smallest for VV motion (7.1%), the largest for TT motion (22.0%), and the intermediate for VT motion (16.0%) and for TV motion (12.5%), although all these effects were significant by themselves, $p < 0.05$ (VV, TV) or $p < 0.01$ (TT, VT).

To examine whether the modality of stimulus at the first and/or the second locations in the target stream made a difference in the auditory spatial capture effect, we conducted a $2 \times 2 \times 2$ ANOVA for the synchronous presentation of stimuli, with the first and the second factors referring to whether the stimulus at the first or the second location was visual or tactile and the third factor referring to the congruency of directions between target stimuli and auditory stimuli. We observed a main effect of stimulus modality at the first location, $F(1, 13) = 15.17$, $p < 0.01$, $\eta_p^2 = 0.54$, and at the second location, $F(1, 13) = 19.20$, $p < 0.01$, $\eta_p^2 = 0.60$, with a higher percentage of correct responses when the stimulus at the first location was visual (91.8%) than when it was tactile (85.6%) and when the stimulus at the second location was visual (91.5%) than when it was tactile (85.9%). The interaction between the modality of stimulus at the first location and the modality of stimulus at the second location did not reach significance, $F(1, 13) = 2.24$, $p > 0.1$, $\eta_p^2 = 0.15$, but the interaction between the modality of stimulus at the second location and congruence did, $F(1, 13) = 11.39$, $p < 0.01$, $\eta_p^2 = 0.47$. Further analysis showed that, when the stimuli at the second location was visual, the accuracy of apparent motion direction judgment was higher ($p < 0.01$) for the congruent stimuli (96.4%) than for the incongruent stimuli (86.6%); when the stimuli at the second location was tactile, the congruency effect was even larger (95.4% *vs* 76.3%).

5. Discussion

Results of this experiment indicated that the size of auditory capture effect on intermodal and intramodal apparent motion varies according to both the timing of the appearance of auditory direction cues and the type of the target motion streams. In the delayed presentation of auditory stimuli, we observed a small but significant congruency effect, indicating that the task-irrelevant direction cues encoded in the auditory stimuli may bias the direction judgment of the apparent motion in the target streams. This finding is inconsistent with Soto-Faraco *et al.* (2002, 2004b) which found that the delayed auditory stream imposes no influence on the judgment of visual or tactile apparent motion. We speculate that the small congruency effect in this experiment was purely postperceptual (Meyer and Wuerger, 2001) and it was probably caused by the variation of target motion streams which would render the direction judgment more difficult and more susceptible to influence of postperceptual bias. Nevertheless, it is important to note that this postperceptual bias occurred regardless of the type of target streams and cannot be used to account for the much larger auditory capture effects in the synchronous presentation of auditory stimuli.

The generally larger congruency effects for synchronous presentation varied as a function of the combination of modalities in the target streams. The smallest effect for the VV stream was consistent with earlier studies (Kitagawa and Ichihara, 2002; Soto-Faraco *et al.*, 2002, 2004b) that did not observe a significant impact of direction cues in the auditory stream upon the direction judgment of visual apparent motion. The largest effect for the TT stream was also consistent with earlier studies showing the capture effect of auditory stimuli upon tactile stimuli (Bresciani and Ernst, 2007; Bresciani *et al.*, 2005; Soto-Faraco *et al.*, 2004a). The novel finding in this experiment was that the direction cues in the auditory stimuli can also affect the direction judgment of intermodal apparent motion (i.e., visual–tactile, tactile–visual) and the size of this capture effect was between the effects for pure visual and pure tactile apparent motion. Moreover, it seems that it is the properties of stimuli at the second location that have a larger effect in determining the size of the spatial intermodal and intramodal capture effects, with the capture effect being larger for tactile stimuli than for visual stimuli.

The smallest congruency effect for the VV stimuli and the largest effect for the TT stimuli can be accounted for by the modality functional appropriateness hypothesis (Welch and Warren, 1980, 1986), according to which the visual modality is superior in spatial localization compared with auditory and tactile modalities. The dominance of visual directional cues in the discrimination of motion direction is likely to prohibit the influence of auditory temporal information upon the perception of visual apparent motion. On the other hand, tactile information and auditory information tend to be reciprocal and equivalent in both spatial and temporal processing (Soto-Faraco and Deco, 2009; Von Békésy, 1959) and the capture effect has been found to be bidirectional between audition and touch (Gesheider, 1970; Gillmeister and Eimer, 2007). Evidence has also shown that in judging the synchrony of two stimuli, the audio-tactile pair is superior in temporal resolution over the visuo–tactile or audio–visual pair (Fujisaki and Nishida, 2009). Thus, we can argue that the temporal information in the auditory input had imposed a noticeable influence on the TT apparent motion.

Importantly, for the intermodal visual/tactile apparent motion, we found that the auditory capture effect was intermediate between the effects for the VV and TT streams. This finding suggests that the strength of intermodal apparent motion is determined not by information from a single modality but by the integration of information of two modalities. This finding is consistent with the view that multisensory interaction is the result of integrative processes whereby information coming from each sensory modality is weighted during perception (e.g., Alais and Burr, 2004; Battaglia *et al.*, 2003; Bertelson, 1999; de Gelder and Bertelson, 2003; Ernst and Banks, 2002; Welch, 1999; Witten and Knudsen, 2005). The VT or TV streams containing the weaker tactile modality is hence more susceptible to the spatial capture of auditory stimuli than the VV stream.

One might argue that the differential auditory capture effects on the intermodal and intramodal apparent motion were simply due to the difference in the perceived quality or strength of these motions. To rule out this possibility, we asked ten participants (3 males, mean age of 24.8 years) who did not participate in the formal experiment to rate, on a 6-point Likert scale (6 = strongest motion, 1 = no motion at all), the strength of each of the four types of apparent motion after being presented with each stream for three times. The mean scores were 4.9 for the VV stimuli, 4.3 for the TT stimuli, 3.5 for either the VT or TV stimuli. The strength of apparent motion was significantly stronger for the VV and TT stimuli than for the VT and TV stimuli ($p < 0.05$; see also Harrar *et al.*, 2008), while in the present experiment the auditory capture effect was the largest for the TT stimuli and the smallest for the VV stimuli. This pattern is clearly inconsistent with the suggestion that the susceptibility of intramodal and intermodal apparent motion to auditory motion capture is solely determined by the strength of apparent motion.

One might also argue that the judgment of the direction of apparent motion was based on the perceived temporal sequence (and spatial locations) of visual–visual, tactile–tactile, visual–tactile and tactile–visual pairings rather than on the distinct motion percepts *per se*. To rule out this possibility, we asked the same ten participants to perform a temporal order judgment (TOJ) task on the same target streams with the same experimental setting as in Experiment 1. The mean percentages of correct judgment were 98.5% for VV, 94.5% for TT, 94.1% for VT and 94.3% for TV. Although the accuracy of TOJ for VV was significantly higher than the accuracy for TT, VT and TV (ps < 0.05), the latter three did not differ between themselves, inconsistent with the pattern of the auditory capture effects for the TT, VT and TV stimuli. Thus, the TOJ data suggest that the auditory motion capture effect upon the intermodal or intramodal apparent motion cannot be accounted for wholly by the difference in TOJ.

On the other hand, it seems that the functional appropriateness of stimuli at the second (ending) position of the intermodal apparent motion sequence has an important role in determining the auditory capture effect. It is possible that the larger auditory capture effect upon intermodal (and intramodal) apparent motion ending with a tactile stimulation, rather than a visual stimulation, at the second position is because the tactile modality is functionally weaker than the visual modality in processing spatial information and is more susceptible to the influence of auditory information. The spatial position of the second stimulation is likely to be more important than the spatial position of the first stimulation in determining the direction of apparent motion. Earlier studies have demonstrated that, in tasks of wayfinding and spatial localization, the sense of direction in animal and human subjects is seldom contributed by the first landmark but largely determined by the second landmark (Barlow, 1964; Howard and Templeton, 1966). Moreover, the position of the second stimulation is relatively more important than the first one in the development of direction cognition (Bridgeman, 2003).

6. Experiment 2

In Experiment 1, we adopted a dynamic capture paradigm in the spatial domain, with the intermodal visual/tactile apparent motion being readily captured by the conflicting but synchronous auditory motion stream. As demonstrated in recent studies, the perception of the direction of apparent motion in a target modality can be modulated by the temporal relationship between stimuli in the target modality and stimuli in a distractor modality (Bruns and Getzmann, 2008; Freeman and Driver, 2008; Getzmann, 2007; Kafaligonul and Stoner, 2010; Shi *et al.*, 2010). In Experiment 2, we examined whether auditory events can also capture the intermodal visual/tactile apparent motion when the temporal relations between auditory stimuli and visual/tactile events were manipulated.

We binaurally presented two beeps which were either synchronous or asynchronous with the two corresponding stimuli (visual and/or tactile) in the target motion stream (Fig. 3). For asynchronous presentation, the first sound preceded

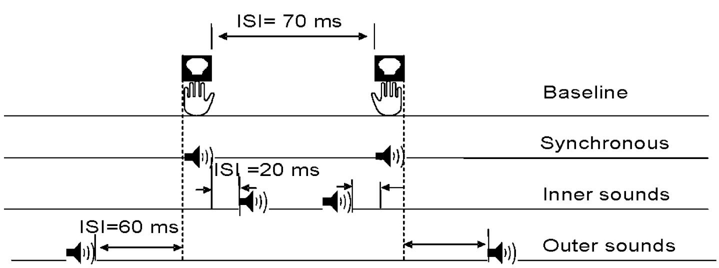

Figure 3. Temporal correspondences of stimuli in Experiment 2. Two biaurally presented auditory beeps were synchronized with visual/tactile stimulations (‘synchronous sounds’). Alternatively, the first beep was lagging the first stimulation and the second beep was leading the second stimulation (‘inner sounds’) or the first beep was leading the first stimulation and the second beep was lagging the second stimulation (‘outer sounds’). There was also a baseline condition in which no beeps were presented.

the onset of one target stimulus at the first spatial location and the second sound trailed the onset of the other target stimulus at the second spatial location (the 'outer sounds' condition), or the first sound trailed the onset of the target stimulus at the first location and the second sound preceded the target stimulus at the second location (the 'inner sounds' condition). For synchronous presentation, the onsets of the two sounds were aligned with the onsets of the target stimuli. Finally, a baseline condition was incorporated in which the target motion stream was presented in the absence of any sounds.

In the presence of spatially static but temporally asynchronous sound pairs, the mechanism of temporal ventriloquism may be at work (Morein-Zamir *et al.*, 2003; Shi *et al.*, 2010). Sounds may capture the intermodal apparent motion by biasing the perceived SOAs between stimuli in the target motion stream; that is, the 'outer sounds' would 'pull' apart the two stimuli in the target motion stream, making the determination of spatial locations of the two target stimuli hence the judgment of the motion direction easier than in the baseline or synchronous condition. Conversely, the 'inner sounds' would 'pull' closer the two target stimuli, making the determination of spatial location and the judgment of motion direction difficult.

7. Method

7.1. Participants

Fourteen undergraduate and graduate students (9 females, average age 22.5 years) were tested. None of them reported any history of somatosensory or auditory deficits. They had normal or corrected-to-normal vision and were naïve to the purpose of this study.

7.2. Apparatus and Stimuli

The same experimental setting as in Experiment 1 was used. However, the stimulus duration was curtailed to only 10 ms for all the visual, auditory and tactile stimuli and the ISI between two stimuli in the target motion stream was shortened to 70 ms. These parameters were determined through pilot tests to fulfill two purposes: one was to conveniently manipulate the temporal correspondence between crossmodal events with the curtailed duration of the individual stimulus, and the other was to obtain good apparent motion within the curtailed temporal interval between the target stimuli. The auditory pair (65 dB, 500 Hz) was presented to both ears simultaneously, providing no spatial information that could be used in judging the direction of the target apparent motion.

7.3. Design and Procedures

A 4 (type of apparent motion: VV, VT, TV, TT) $\times$ 4 (temporal relations between auditory events and target stimuli: synchronous, outer, inner and baseline) factorial design was adopted. For the 'outer sounds' condition, the onset of binaurally presented beeps preceded the onset of the first target stimulus at the first spatial location by 60 ms and the offset of the second beep trailed the onset of the second target stimulus at the second spatial location by 60 ms (Fig. 3). For the 'inner sounds' condition, the first beep trailed the onset of the first target stimulus at the first location by 20 ms and the second beep preceded the second target stimulus at the second location by 20 ms. The timing of stimulus presentation and data collection were the same as in Experiment 1. Participants were asked to judge the direction of the visual/tactile apparent motion, disregarding the auditory stimuli if they were presented. The experiment had a total of 384 trials, divided into six test blocks with each block having 4 trials from each experimental condition.

8. Results

Percentages of correct responses and standard deviations are presented in Table 2. A 4×4 repeated-measures ANOVA revealed a significant main effect of motion type, $F(3, 39) = 13.06$, $p < 0.001$, $\eta_p^2 = 0.50$. Bonferroni-corrected pairwise comparisons showed that the percentage of correct responses for the VV stream (87.1%) was higher than for the other three types of streams (VT, 75.1%; TV, 77.4%; TT, 79.5%), $p < 0.01$ (VV *vs* TT, VV *vs* VT) or $p < 0.001$ (VV *vs* TV), but that there were no differences between the latter three types, $p\text{s} > 0.1$. The main effect of temporal correspondence was also significant, $F(3, 39) = 19.13$, $p < 0.001$, $\eta_p^2 = 0.60$. The mean percentages of correct responses were 86.0, 74.6 79.6 and 78.9% for the outer, inner, synchronous and baseline conditions, respectively. Bonferroni-corrected pairwise comparisons showed that the synchronous auditory beeps did not influence the direction judgment of apparent motion in the target stream, as there was no difference between the synchronous and the baseline conditions, $p > 0.1$. However, the outer sounds led to an improvement of performance while the inner sounds induced an impairment of performance: outer *vs* baseline, $p < 0.001$; inner

Table 2.
Percentages of correct responses in Experiment 2, with numbers in brackets indicating standard errors (VV: visual–visual motion; VT: visual–tactile motion; TV: tactile–visual motion; TT: tactile–tactile motion)

	Baseline	Synchronous	Outer sounds	Inner sounds	Average
VV	88.1 (3.7)	85.7 (3.8)	91.7 (2.3)	82.7 (4.4)	87.1
VT	71.1 (3.9)	77.7 (4.4)	81.5 (5.1)	69.9 (4.0)	75.1
TV	76.2 (4.5)	77.7 (4.0)	83.9 (4.3)	71.7 (4.3)	77.4
TT	80.1 (4.1)	77.4 (3.2)	86.9 (3.5)	73.8 (3.0)	79.5
Average	78.9	79.6	86.0	74.6	

vs baseline, $p < 0.05$. The two-way interaction between motion type and temporal correspondence was not significant, $F(9, 117) = 0.92$, $p > 0.1$, $\eta_p^2 = 0.07$, indicating that the auditory temporal ventriloquism effect did not vary according to the type of apparent motion.

To examine whether the modality of stimulus at the first and/or the second locations in the target stream made a difference in the auditory temporal capture effect, we conducted a $2 \times 2 \times 4$ ANOVA, with the first and the second factors referring to whether the stimulus at the first or the second location was visual or tactile and the third factor referring to the type of temporal correspondence. We observed a main effect of stimulus modality at the first location, $F(1, 13) = 10.77$, $p < 0.01$, $\eta_p^2 = 0.45$, and at the second location, $F(1, 13) = 9.68$, $p < 0.01$, $\eta_p^2 = 0.43$, with higher percentage of correct responses when the stimulus at the first location was visual (81.2%) than when it was tactile (78.4%) and with higher percentage of correct responses when the stimulus at the second location was visual (82.2%) than when it was tactile (77.3%). However, the interaction between the modality of stimulus at the first location and the modality of stimulus at the second location was also significant, $F(1, 13) = 16.26$, $p < 0.001$, $\eta_p^2 = 0.56$,

Further tests showed that, across the four conditions, when the visual stimulus appeared at the first location, the percentage of correct responses was higher for VV than for VT, ps < 0.05 or 0.01; when the visual stimulus appeared at the second location, the percentage of correct responses was higher for VV than for TV. However, for the tactile stimuli, there was no such contingency.

9. Discussion

Consistent with previous studies demonstrating that an auditory event can influence the timing of the visual event (Burr *et al.*, 2009; Freeman and Driver, 2008; Getzmann, 2007; Morein-Zamir *et al.*, 2003; Shi *et al.*, 2010) and that auditory information can calibrate tactile events (Bresciani and Ernst, 2007; Bresciani *et al.*, 2005; Chen *et al.*, 2011), we observed temporal modulation of the perception of intramodal (VV, TT) apparent motion. It is likely that this modulation was due to the impact of intervals between paired auditory events upon the perceived time intervals between visual or tactile events (Bresciani and Ernst, 2007; Burr *et al.*, 2009; Shi *et al.*, 2010). Importantly, we extended this temporal ventriloquism effect to the intermodal visual/tactile interaction. According to the Korte's third law (Korte, 1915), with the fixed distance between two stimuli in the target motion stream, the timing information provided by auditory stimuli would affect the perception of the time interval between the two events in the target stream. The perceived short interval (for inner sounds) would elicit the percept of a short SOA between the two events in the target stream, leading to a difficulty in direction judgment; conversely, the perceived long interval (for outer sounds) would engender the percept of a long SOA between the two events in the target stream, rendering the judgment of the target motion direction easier (Harrar *et al.*, 2008). This subjective shortening or lengthening of the time interval between the two stimuli in the target stream might affect the perception of motion direction to the same extent for intramodal and intermodal interaction.

Detailed analyses taking into account the modality of stimuli at the first and the second locations in the target stream showed that, regardless of the temporal correspondence of auditory input and regardless of the location (or temporal order) of stimuli, the percentage of correct responses was higher when the stimulus was visual than when it was tactile. It is clear from Table 2 that this general effect of modality was caused mostly by the fact that the highest percentage of correct responses was to the VV stimuli. When tactile stimulation was involved, accuracy in the judgment of the motion direction was reduced for both intermodal (VT, TV) and intramodal (TT) apparent motion. These findings indicate that the visual event can act as a reliable cue for spatial localization while the tactile cue is susceptible to the influence of temporal interval in the auditory stream.

Note that we did not observe any difference in the accuracy of motion direction judgment between the ‘no sounds’ and ‘synchronous sounds’ conditions. This null effect seems to be inconsistent with Getzmann (2007) in which participants were asked to subjectively categorize different types of apparent motion. Visual streams accompanied with synchronous sounds were less likely to be judged as having ‘continuous motion’ than visual streams without sounds (see also Shi *et al.*, 2010). The author suggested that the two sounds emphasize the distinctness of the two visual stimuli and, thus, hinder the percept of continuous motion. Our motion direction judgment task, however, needs only minimal information concerning the direction of apparent motion and hence the impaired percept of continuous motion, due to the presence of the synchronous sounds, has not noticeable impact upon the direction judgment.

10. General Discussion

Previous investigations into dynamic crossmodal capture have been confined to two given modalities, such as dynamic direction cues or static events in one modality affecting the perception of motion stream in another modality. The present study demonstrated that both spatially (directionally) informative (Experiment 1) and spatially uninformative but temporally asynchronous (Experiment 2) sounds can capture intermodal visual–tactile or tactile–visual apparent motion.

It is known that crossmodal integration takes place within a certain, limited temporal and spatial range (Alais and Burr, 2004; Bresciani *et al.*, 2006; Gepshtein *et al.*, 2005; Shi *et al.*, 2010; Slutsky and Recanzone, 2001; Spence *et al.*, 2007). This range is to a large extent dependent on the functional appropriateness and precision of information processing in a particular sensory modality although experimental conditions and task demands can affect the manifestation of crossmodal integration (Alais and Burr, 2004). According to the functional appropriateness and precision hypothesis (Welch and Warren, 1980, 1986), different modalities possess different dominant functions, with the highest spatial resolution or temporal acuity in one modality dominating the perception of events in the other modality during multisensory integration. This dominance of one modality over the other was observed in the present study. In Experiment 1 with spatial manipulations, the perception of visual apparent motion (VV), compared with other types of motion streams, was the least affected (i.e., with the smallest congruency effect) by direction cues in the auditory input; in Experiment 2 with temporal manipulations, given that audition is stronger in temporal acuity than either touch or vision, it is not surprising that the tactile and visual apparent motion was affected to the same extent by temporal cues in auditory stream.

Interestingly, an important finding here was that the effect of direction cues in the task-irrelevant auditory input upon the perception of intermodal visual/tactile apparent motion was intermediate between its effect upon the perception of intramodal visual (VV) and tactile (TT) apparent motion. This is reminiscent of an earlier finding that when visual stimuli are severely blurred, sound can capture vision; for less blurred visual stimuli, neither sense dominates and perception follows a mean spatial position (Alais and Burr, 2004). The spatial resolution of intermodal visual/tactile apparent motion appears to be the mean of the resolutions of vision and touch. Hence, the effect of auditory direction cues appears to be the mean of the susceptibilities of vision and touch, even though the modality of stimulus at the ending (second) position of the intermodal apparent motion seems to play a more prominent role in the computation of this mean. Alternatively, according to the rules of *inverse effectiveness* in multisensory integration (Holmes and Spence, 2005; Stein and Meredith, 1993), especially in spatial ventriloquism, the crossmodal dynamic capture effect is stronger for functionally weaker signals (in spatial localization). In this way, the auditory capture effect in Experiment 1 was stronger for the tactile stimuli than for the visual stimuli, as was the mean of the susceptibility for the visual–tactile or the tactile–visual stimuli.

On the other hand, in Experiment 2 with temporal manipulations, the timing of auditory input affected to the same extent the perception of intramodal visual and tactile apparent motion (VV, TT) and the perception of intermodal visual/tactile apparent motion (VT, TV). The finding of auditory temporal structures affecting the perception of visual or tactile apparent motion is consistent with earlier studies (Freeman and Driver, 2008; Getzmann, 2007; Kafaligonul and Stoner, 2010; Shi *et al.*, 2010). According to Korte's third law (Korte, 1915), the percept of apparent motion is determined by exposure time, spatial separation and inter-stimulus onset interval (ISI) of the two stimuli in investigation. Although the ISI between the two target stimuli were fixed at 70 ms in this experiment, this interval was subjectively prolonged or curtailed by different auditory temporal structures ('outer sounds' or 'inner sounds'). This made the separation of the spatial location of stimuli in the target stream easier or more difficult and the response accuracy regarding the direction of target apparent motion higher or lower (see also Burr *et al.*, 2009; Shi *et al.*, 2010; van Erp and Werkhoven, 2004). The equivalent effects upon the four types of apparent motion suggest that in temporal ventriloquism there might be a common crossmodal temporal binding process that combines information from the target stream and from the task-irrelevant auditory stream. This process is largely independent of combinations of individual modalities in the target stream, but is dependent upon the perceived temporal interval between the two stimuli in the target stream (Fujisaki and Nishida, 2010; Keetels and Vroomen, 2008; Vroomen and Keetels 2006; Zampini *et al.*, 2005).

Across the two experiments, we found both moving and static sounds impose a capture effect upon the perception of target motion streams composed of visual and/or tactile stimuli. However, the underlying mechanisms for the two types of capture effects may be different. In Experiment 1, a typical *spatial* ventriloquism can account the findings, while in Experiment 2, *temporal* ventriloquism is at work. In spatial ventriloquism, the perception of the motion direction of the target stream is biased by the incongruent auditory motion direction through the integration of direction cues from the target and the auditory streams. In temporal ventriloquism, the perception of motion direction of the target stream is affected by the perceived ISI between the two stimuli in the target stream, and the perception of this ISI is 'captured' by the interval between two concurrently presented sounds. Although the four types of target streams showed the same pattern of variation for the overall accuracy in motion direction judgment, the congruency between the directions of target and auditory streams (Experiment 1) and the temporal correspondence between the target and auditory streams (Experiment 2) had different patterns of impact upon the judgment of target motion direction. Moreover, the modality of the second stimulus in the target stream (Experiment 1) and the modality of both the first and the second stimuli in the target stream (Experiment 2) played an important role in determining the capture effect, with the tactile stimuli more susceptible to auditory capture. These differences strongly suggest different functions of spatial and temporal information in crossmodal integration.

In the above discussions, we have implicitly assumed that the effects we observed for either spatial or temporal manipulations were perceptual in nature, i.e., the intermodal integration taking place at the perceptual level. However, as we demonstrated in Experiment 1, postperceptual processes or response bias might contribute also to these effects (see Meyer and Wueger, 2001). Nevertheless, it is clear that the postperceptual processes cannot be wholly responsible for the interactive pattern of the auditory capture effects in Experiment 1. For Experiment 2, the fact that the auditory input did not provide direction cues that were directly related to the judgment of the direction of apparent motion in the target streams helps us to rule out a simple postperceptual account.

To conclude, by presenting direction cues or temporal information in the task-irrelevant auditory modality and by asking participants to judge the direction of apparent motion caused by the sequential presentation of visual and/or tactile stimuli at two spatial locations, we demonstrated that perception of both intramodal (visual or tactile) and intermodal (visual/tactile) apparent motion can be affected by the task-irrelevant spatial (apparent motion) or temporal information (of static events) in another modality, expanding the scope of crossmodal dynamic capture and intermodal integration.

Acknowledgements

This research was supported by grants from the Natural Science Foundation of China (30770712, 90920012) and the Ministry of Science and Technology of China (2010CB833904) to Xiaolin Zhou and a grant from China Postdoctoral Science Foundation (20100470) to Lihan Chen. We thank two anonymous reviewers for their constructive comments and helpful suggestions.

References

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262.

Barlow, J. S. (1964). Inertial navigation as a basis for animal navigation, *J. Theor. Biol.* **6**, 76–117.

Bensmaïa, S. J., Killebrew, J. H. and Craig, J. C. (2006). Influence of visual motion on tactile motion perception, *J. Neurophysiol.* **96**, 1625–1637.

Battaglia, P. W., Jacobs, R. A. and Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1391–1397.

Bertelson, P. (1999). Ventriloquism: a case of crossmodal perceptual grouping, in: *Cognitive Contributions to the Perception of Spatial and Temporal Events*, Aschersleben, G., Bachmann, T. and Müsseler, J. (Eds), pp. 347–362. Elsevier, Amsterdam.

Bertelson, P. and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location, *Psychonomic Bull. Rev.* **5**, 482–489.

Brainard, D. H. (1997). The psychophysics toolbox, *Spat. Vis.* **10**, 433–436.

Bresciani, J. P., Dammeier, F. and Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events, *J. Vision* **6**, 554–564.

Bresciani, J. P., Ernst, M. O. (2007). Signal reliability modulates auditory-tactile integration for event counting, *Neuroreport* **18**, 1157–1161.

Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V. and Kheddar, A. (2005). Feeling what you hear: auditory signals can modulate tactile tap perception, *Exper. Brain Res.* **162**, 172–180.

Bridgeman, B. (2003). *Psychology & Evolution: The Origins of Mind*. Sage Publications, Inc., Thousand Oaks, CA.

Bruns, P. and Getzmann, S. (2008). Audiovisual influences on the perception of visual apparent motion: Exploring the effect of a single sound, *Acta Psychol. (Amst.)* **129**, 272–283.

Burr, D., Banks, M. S. and Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration, *Exper. Brain Res.* **198**, 49–57.

Caclin, A., Soto-Faraco, S., Kingstone, A. and Spence, C. (2002). Tactile "capture" of audition, *Percept. Psychophys.* **64**, 616–630.

Chen, L., Shi, Z. and Müller, H. J. (2011). Interaction of perceptual grouping and crossmodal temporal capture in tactile apparent-motion, *PLoS ONE* **6**, e17130; doi:10.1371/journal.pone.0017130.

Craig, J. C. (2006). Visual motion interferes with tactile motion perception, *Perception* **35**, 351–367.

de Gelder, B. and Bertelson, P. (2003). Multisensory integration, perception and ecological validity, *Trends Cogn. Sci.* **7**, 460–467.

Driver, J. and Spence, C. (2000). Multisensory perception: beyond modularity and convergence, *Curr. Biol.* **10**, R731–R735.

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

Exner, S. (1875). Uber das Sehen von Bewegungen und die Theorie des zusammengesetzten Auges, *Sitzungsberichte der Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse, Abteilung* **3**, 156–190.

Freeman, E. and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound, *Curr. Biol.* **18**, 1262–1266.

Fujisaki, W. and Nishida, S. (2009). Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception, *Exper. Brain Res.* **198**, 245–259.

Fujisaki, W. and Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities, *Proc. R. Soc. B* **277**, 2281–2290.

Galli, P. A. (1932). Über mittelst verschiedener Sinnesreize erweckte Wahrnehmung von Scheinbewegungen [On the perception of apparent motion elicited by different sensory stimuli], *Archiv. fur die gesamte Psychologie* **85**, 137–180.

Gepshtein, S., Burge, J., Ernst, M. O. and Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity, *J. Vision* **5**, 1013–1023.

Gesheider, G. A. (1970). Some comparisons between touch and hearing, *IEEE Transactions on Man-machine Systems* **11**, 28–35.

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion, *Perception* **36**, 1089–1103.

Gillmeister, H. and Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness, *Brain Res.* **1160**, 58–68.

Harrar, V. and Harris, L. R. (2007). Multimodal ternus: visual, tactile, and visuo–tactile grouping in apparent motion, *Perception* **36**, 1455–1464.

Harrar, V., Winter, R. and Harris, L. R. (2008). Visuotactile apparent motion, *Percept. Psychophys.* **70**, 807–817.

Hirsh, I. J. and Sherrick, C. E. J. (1961). Perceived order in different sense modalities, *J. Exper. Psychol.* **62**, 423–432.

Holmes, N. P. and Spence, C. (2005). Multisensory integration: space, time and superadditivity, *Curr. Biol.* **15**, R762–R764.

Howard, I. P. and Templeton, W. B. (1966). *Human Spatial Orientation*. Wiley, New York.

Kafaligonul, H. and Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals, *J. Vis.* **10**(12), 31; doi: 10.1167/10.12.31.

Keetels, M. and Vroomen, J. (2005). The role of spatial disparity and hemifields in audio–visual temporal order judgments, *Exper. Brain Res.* **167**, 635–640.

Keetels, M. and Vroomen, J. (2008). Tactile–visual temporal ventriloquism: no effect of spatial disparity, *Percept. Psychophys.* **70**, 765–771.

Kitagawa, N. and Ichihara, S. (2002). Hearing visual motion in depth, *Nature* **416**, 172–174.

Korte, A. (1915). Kinematoskopische Untersuchungen, *Zeitschrift fur Psychologie* **72**, 193–206.

Lyons, G., Sanabria, D., Vatakis, A. and Spence, C. (2006). The modulation of crossmodal integration by unimodal perceptual grouping: a visuotactile apparent motion study, *Exper. Brain Res.* **174**, 510–516.

Meyer, G. F. and Wueger, S. M. (2001). Cross-modal integration of auditory and visual motion signals, *Neuroreport* **12**, 2557–2560.

Morein-Zamir, S., Soto-Faraco, S. and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism, *Cogn. Brain Res.* **17**, 154–163.

Oruc, I., Sinnett, S., Bischof, W. F., Soto-Faraco, S., Lock, K. and Kingstone, A. (2008). The effect of attention on the illusory capture of motion in bimodal stimuli, *Brain Res.* **1242**, 200–208.

Shi, Z., Chen, L. and Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: the influence of time interval, *Exper. Brain Res.* **203**, 723–735.

Slutsky, D. A. and Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect, *Neuroreport* **12**, 7–10.

Soto-Faraco, S. and Deco, G. (2009). Multisensory contributions to the perception of vibrotactile events, *Bahav. Brain Res.* **196**, 145–154.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C. and Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities, *Cogn. Brain Res.* **14**, 139–146.

Soto-Faraco, S., Kingstone, A. and Spence, C. (2003). Multisensory contributions to the perception of motion, *Neuropsychologia* **41**, 1847–1862.

Soto-Faraco, S., Spence, C. and Kingstone, A. (2004a). Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture, *Cogn. Affect. Behav. Neurosci.* **4**, 208–217.

Soto-Faraco, S., Spence, C. and Kingstone, A. (2004b). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities, *J. Exper. Psychol. Hum. Percept. Perform.* **30**, 330–345.

Spence, C., Sanabria, D. and Soto-Faraco, S. (2007). Intersensory Gestalten and crossmodal scene perception, in: *The Psychology of Beauty and Kansei: New Horizons of Gestalt Perception*, Noguchi, K. (Ed.). Nihon University College of Humanities and Sciences, Tokyo.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA.

Strybel, T. Z. and Vatakis, A. (2004). A comparison of auditory and visual apparent motion presented individually and with crossmodal moving distractors, *Perception* **33**, 1033–1048.

van Erp, J. B. and Werkhoven, P. J. (2004). Vibro-tactile and visual asynchronies: Sensitivity and consistency, *Perception* **33**, 103–111.

Von Békésy, G. (1959). Similarities between hearing and skin sensations, *Psychol. Rev.* **66**, 1–22.

Vroomen, J. and de Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect, *J. Exper. Psychol. Hum. Percept. Perform.* **30**, 513–518.

Vroomen, J. and Keetels, M. (2006). The spatial constraint in intersensory pairing: no role in temporal ventriloquism, *J. Exper. Psychol. Hum. Percept. Perform.* **32**, 1063–1071.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psychol. Bull.* **88**, 638–667.

Welch, R. B. and Warren, D. H. (1986). Intersensory interactions, in: *Handbook of Perception and Human Performance*, Vol. 1, Boff, K. R., Kaufman, L. and Thomas, J. P. (Eds), pp. 1–36. Wiley, New York.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions, in: *Cognitive Contributions to the Perception of Spatial and Temporal Events*, Aschersleben, G., Bachmann, T. and Musseler, J. (Eds), pp. 371–387. Elsevier, Amsterdam.

Witten, I. B. and Knudsen, E. I. (2005). Why seeing is believing, *Neuron* **48**, 480–496.

Zapparoli, G. C. and Reatto, L. L. (1969). The apparent movement between visual and acoustic stimulus and the problem of intermodal relations, *Acta Psychol. (Amst.)* **29**, 256–267.

Zampini, M., Brown, T., Shore, D. I., Maravita, A., Röder, B. and Spence, C. (2005). Audiotactile temporal order judgments, *Acta Psychol. (Amst.)* **118**, 277–291.