BRILL

26 (2013) 19 51

MULTISENSORY RESEARCH

brill.com/msr

Mutual Influences of Intermodal Visual/Tactile Apparent Motion and Auditory Motion with Uncrossed and Crossed Arms

Yushi Jiang [1,2] **and Lihan Chen** [1,3,*]

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

[2] School of Economics and Management, Southwest Jiaotong University, Chengdu Sichuan, 610031, China

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

Received 13 March 2012; accepted 1 October 2012

Abstract

Intra-modal apparent motion has been shown to be affected or 'captured' by information from another, task-irrelevant modality, as shown in cross-modal dynamic capture effect. Here we created inter-modal apparent motion between visual and tactile stimuli and investigated whether there are mutual influences between auditory apparent motion and inter-modal visual/tactile apparent motion. Moreover, we examined whether and how the spatial remapping between somatotopic and external reference frames of tactile events affect the cross-modal capture between auditory apparent motion and inter-modal visual/tactile apparent motion, by introducing two arm postures: arms-uncrossed and arms-crossed. In Experiment 1, we used auditory stimuli (auditory apparent motion) as distractors and inter-modal visual/tactile stimuli (inter-modal apparent motion) as targets while in Experiment 2 we reversed the distractors and targets. In Experiment 1, we found a general detrimental influence of arms-crossed posture in the task of discrimination of direction in visual/tactile stream, but in Experiment 2, the influence of arms-uncrossed posture played a significant role in modulating the inter-modal visual/tactile stimuli capturing over auditory apparent motion. In both Experiments, the synchronously presented motion streams led to noticeable directional congruency effect in judging the target motion. Among the different modality combinations, tactile to tactile apparent motion (TT) and visual to visual apparent motion (VV) are two signatures revealing the asymmetric congruency effects. When the auditory stimuli were targets, the congruency effect was largest with VV distractors, lowest with TT distractors; the pattern was reversed when the auditory stimuli were distractors. In addition, across both experiments the congruency effect in visual to tactile (VT) and tactile to vi-

* To whom correspondence should be addressed. E-mail: CLH20000@gmail.com

DOI:10.1163/22134808-00002409

sual (TV) apparent motion was intermediate between the effect-sizes in VV and TT. We replicated the above findings with a block-wise design (Experiment 3). In Experiment 4, we introduced static distractor events (visual or tactile stimulus), and found the modulation of spatial remapping of distractors upon AA motion is reduced. These findings suggest that there are mutual but a robust asymmetric influence between intra-modal auditory apparent motion and intermodal visual/tactile apparent motion. We proposed that relative reliabilities in directional information between distractor and target streams, summed over a remapping process between two spatial reference frames, determined this asymmetric influence.

Keywords
Apparent motion, intermodal interaction, cross-modal capture, posture change

1. Introduction

Since the seminal investigation of Wertheimer (1912) towards visual apparent motion, the perception of apparent motion within each individual sensory modalities has been widely explored and documented (Burtt, 1917; Kirman, 1974; Kolers, 1972; Lakatos and Shepard, 1997; Strybel and Vatakis, 2004). Apparent motion as a movement illusion refers to the following perception: two spatially discrete stimuli are presented in rapid succession, although the stimuli themselves are not moving, the observer perceives a movement with the direction from the first occurring stimulus to the second one. It is proposed that apparent motion is primarily affected by three factors: the exposure time of the stimuli, and the temporal and spatial separation between them. Among the three factors, there are common constraints to apparent motion in visual, tactile and auditory space; the increased spatial separations between the stimuli in a given sensory modality would generally require an extended temporal interval between the stimuli, in order that a good discrimination of the direction of apparent motion can be achieved (Lakatos and Shepard, 1997). The time–spatial constraints in apparent motion has been reflected in Korte's third law (Korte, 1915), which originally states that the stimulus onset asynchrony (SOA) between two lights that produces optimal apparent motion is proportional to the distance between them.

The phenomenon and principles of intra-modal apparent motion have been robustly established. However, there has been a debate on whether there is intermodal apparent motion, in which two or more stationary stimuli of different sensory modalities are presented briefly from different spatial locations at appropriate inter-stimulus intervals. About half a century ago, several studies suggest the existence of inter-modal apparent motion with all possible combinations of auditory, visual and tactile stimuli (Galli, 1932; Zapparoli and Reatto, 1969), using personal feeling report towards the different combinations. Recently, Harrar . (2007, 2008) compared the properties of apparent motion between a light and a touch with apparent motion between either two

lights or two touches. Subjects rated the quality of apparent motion between each stimulus combination for a range of stimulus onset asynchronies (SOAs). Subjects reported perceiving apparent motion between all the following three stimulus combinations. For light-light visual apparent motion, it was consistent with Korte's third law. Touch-touch apparent motion also obeyed Korte's third law, but over a smaller range of distances, showing that proprioceptive information concerning the position of the fingers was integrated into the tactile motion system. The threshold and preferred SOAs for visuotactile apparent motion did not vary with distance, suggesting a different mechanism for multimodal apparent motion. The above studies revealed apparent motion saliency is dependent on different modalities. The findings of Harrar . were replicated by a recent study (Chen and Zhou, 2011), in which participants were asked to make a judgment of the direction of intra-modal (within auditory, visual or tactile sensory modality) apparent motion and inter-modal (visual to tactile or tactile to visual) apparent motion, the motion strength of different types of motion, as measured by a 6-point Likert scale, showed that the perceived strength of visual–visual apparent motion and tactile–tactile apparent motion is stronger than auditory-auditory apparent motion and inter-modal apparent motion (visual to tactile or tactile to visual apparent motion).

Apparent motion (AM), arising intra-modally (such as visual or tactile apparent motion) or inter-modally (visual to tactile apparent motion), has been successfully applied to investigate the multisensory interaction between dynamic events, typified in -*Sh* *Sh* . 'Cross-modal dynamic capture' refers to the phenomenon of directional information in one modality affecting the perceived direction of apparent motion in another modality (Soto-Faraco ., 2002). Typically, stimuli presented in the distractor modality are congruent or incongruent with stimuli in the target modality in terms of motion direction; the two motion streams could be presented synchronously or asynchronously. Participants are instructed to judge the motion direction of stimuli in the target modality while trying to ignore the stimuli in the distractor modality (Soto-Faraco ., 2002, 2004a, 2004b). An interesting finding in cross-modal dynamic capture effect is 'asymmetry' between the auditory, visual and tactile modalities. According to Soto-Faraco . (2003), with the dynamic spatial cross-modal capture, auditory apparent motion was strongly influenced by visual apparent motion (46%) and by tactile apparent motion (36%); tactile apparent motion was modulated by visual apparent motion (44%) and, to a smaller degree, by auditory apparent motion (15%); and finally, visual apparent motion was uninfluenced by auditory or tactile apparent motion. The direction of visual stimuli can capture the direction of auditory apparent motion but, on many occasions, the direction cues in the auditory stimuli cannot capture the direction of visual apparent motion (Soto-Faraco ., 2002; Soto-Faraco ., 2004b; Strybel and Vatakis, 2004). However,

under conditions of weak visual motion cues and attentional modulations, auditory motion can exert an influence on visual motion (Meyer and Wuerger, 2001; Oruc ., 2008; Sanabria ., 2007a). Likewise, tactile motion distractors have been revealed to impose a stronger influence upon the perception of auditory motion direction than auditory motion distractors do on tactile perception (Soto-Faraco ., 2004a); however, a recent study showed that the capture pattern between auditory and tactile motion depends on the precise stimuli parameters (Occelli ., 2009). For example, the cross-modal capture of tactile motion by audition was stronger with the more intense (less intense) auditory distractors and the capture effect exerted by the tactile distractor was stronger for less intense (than for more intense) auditory targets (Occelli ., 2009). In the case of interactions between visual motion and tactile motion, the visual motion distractors have a stronger influence on the perception of tactile motion direction than tactile motion distractors do on visual perception (Bensmaïa ., 2006; Craig, 2006; Lyons ., 2006). These asymmetries have been attributed to the differences in functional appropriateness and precision between different modalities under different experimental conditions (Alais and Burr, 2004; Freeman and Driver, 2008; Kafaligonul and Stoner, 2010; Shi ., 2010; Welch and Warren, 1980, 1986; Welch ., 1986). Furthermore, attentional biasing due to the features of stimuli (Occelli ., 2009) and the arrangement of sensory events being attended alters the perceptual grouping/segregation of unimodal stimuli in the distractor stimuli, and hence leads to the differential effect in cross-modal capture (Oruc ., 2008; Sanabria ., 2007b).

With intermodal (visual and (or) tactile) apparent motion, Chen and Zhou (2011) demonstrated that both moving and asynchronous static sounds can capture intermodal (visual–tactile and tactile–visual) apparent motion; the auditory direction cues have less impact upon the perception of intra-modal visual apparent motion than upon the perception of intra-modal tactile or intermodal visual/tactile apparent motion. Their findings suggest intermodal apparent motion is susceptible to the influence of dynamic or static auditory information in similar ways as intra-modal visual or tactile apparent motion. Here we further ask conversely, whether the inter-modal visual/tactile apparent motion affects auditory apparent motion. In addition, we examine whether different spatial frames of reference in distractor (tactile) events, in addition to the temporal/spatial relations, as an additional constraint factor, could affect cross-modal dynamic capture. This issue is important because for certain stimuli, such as tactile stimuli, their spatial locations (and therefore, tactile motion direction) can be encoded at two categories of spatial frames of reference, i.e. the anatomical and external frames of reference. They could be either in spatial conflict (when the arms are crossed) or be aligned (arms uncrossed). Evidence has shown cross-modal links (coupling) in spatial attention in/between visual

and tactile stimuli would be dynamically updated as new body positions are adopted (Azañón ., 2008, 2010; Eimer ., 2001; Lloyd ., 2003). The empirical question/research purpose for the current study is two-fold:

, how the spatial remapping of distractor events (the hands-crossed condition served as bias prior) (Ernst and Di Luca, 2011), interacting with the traditional established principle of 'modality appropriateness and precision' (corresponding to 'likelihood' function) associated with the sensory events (Welch and Warren, 1980), determine the outcome of cross-modal dynamic capture. The spatial remapping could take two forms: correspondence between two 'activated' events from two opposite spatial locations, implemented the intra-modal or inter-modal distractor motion streams (we call it 'full-remapping', Experiments 1–3) or correspondence of the above two frames of reference, but with activated static events only in one spatial position (we name it 'half-remapping', Experiment 4). For the latter form, the conflicting correspondence for instance is that left finger touching stimuli in the right position and right finger touching the stimuli in the left position, but only one single stimulus (or combined visual/tactile stimuli in one location) appeared for a given experimental trial. Investigating into this issue would further reveal the constraint factor of spatial remapping in cross-modal dynamic capture.

, in contrast to the trial-by-trial design used in previous studies (as well as in current Experiment 1 and Experiment 2), a block-wise design (Experiment 3) was used to reveal whether the congruency effects robustly remained. The block-wise arrangement would reduce the frequent attentional shift between each sub-condition. If the cross-modal dynamic capture is susceptible to the adaptation of given fixed type of motion streams in a block, we would anticipate the repetition of trials within such a block would neutralize the congruency effects (Oruc ., 2008), otherwise, the cross-modal dynamic capture effect would still survive. This issue seems trivial, but to our best knowledge, it has not been rigorously tested within a same single study.

Here we adapted the same paradigm used in Chen and Zhou (2011) and presented participants with visual and tactile stimuli at two different locations (Fig. 1A) to create intermodal (visual–tactile, tactile–visual), intra-modal (visual–visual, tactile–tactile) apparent motion, and auditory apparent motion, which encoded direction information congruent or incongruent with the direction of intermodal or intra-modal apparent motion. In Experiment 1, the auditory stimuli were distractors, and participants were asked to make discriminations of the direction of visual and (or) tactile apparent motion, with arms crossed (as compared to Experiment 1 in Chen and Zhou (2011), where the participants' arms were not crossed). In Experiment 2, the auditory stimuli were targets, participants were asked to make discrimination of the direction of auditory apparent motion, irrespective of the visual and (or) tactile events, Experiment 2 included two sub-experiments: in Experiment 2a, the participants'

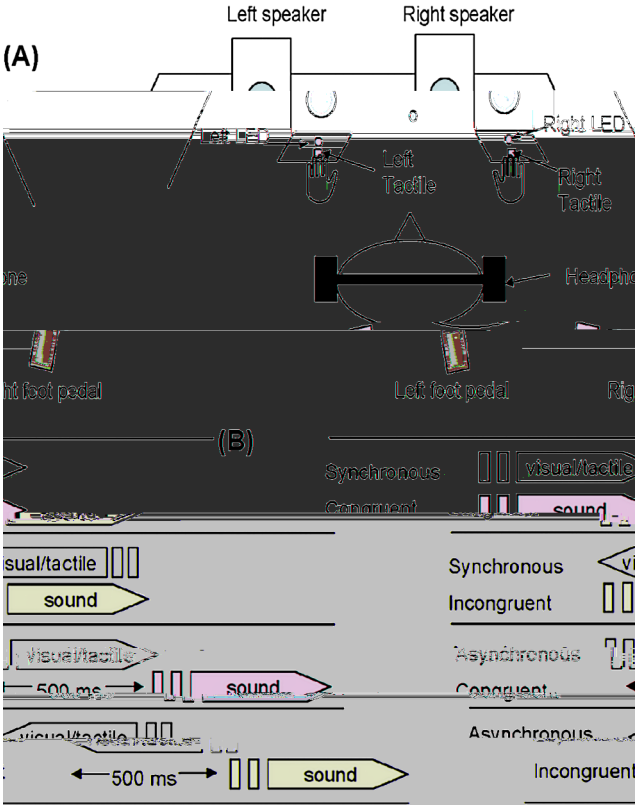


Figure 1. Experimental setup and temporal correspondence of motion streams used in Experiment 1 and Experiment 2. (A) The participant placed two middle fingers on the tactile actuators which were embedded into foams, which were placed just in front of the two speakers (the positions of two middle fingers were reversed in Experiment 2). 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 colour in the online version.

arms were not crossed, but in Experiment 2b, their arms were crossed. The empirical question was whether and how intermodal v

the impact of spatial remapping in visual/tactile distractors (without apparent motion) upon auditory apparent motion.

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 intra-modal (VV and TT) or the intermodal (VT and TV) rightward or leftward apparent motion (Chen and Zhou, 2011; Harrar and Harris, 2007; Harrar ., 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 intra-modal apparent motion. Participants put their left middle-finger on the surface of the right tactile stimulus and put the right middle-finger on the surface of the left tactile stimlus, they were instructed to judge the direction of the intermodal or intra-modal apparent motion while ignoring the auditory input.

The auditory stream could appear at the same time as the intermodal or intra-modal 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 ., 2004a).

2.1.

2.1.1.

Nineteen undergraduate and graduate students (8 females, average age 22.9 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.

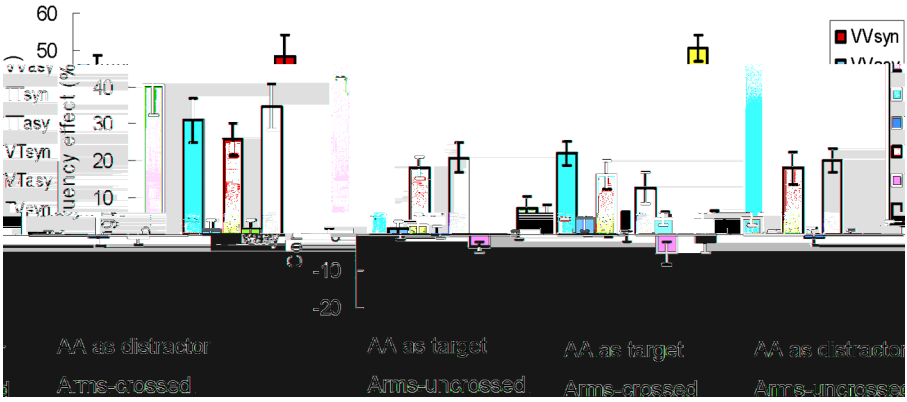
2.1.2.

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 (5V, with duration of 50 ms) and to the actuators (with duration of 50 ms, the delay for issuing tactile tap is about 3 ms, as confirmed by a personal communication with Heijo) were controlled a parallel LPT port by software written with Matlab (Mathworks Inc.) with Psychtoolbox (Brainard, 1997). 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

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; Congruent — the direction of AA motion and the target motion stream was congruent; Incongruent — the direction of AA motion and the target motion stream was incongruent; Norm — arms were not crossed; Cross — arms were crossed)

	VV				VT				TV				TT			
	Congruent		Incongruent		Congruent		Incongruent		Congruent		Incongruent		Congruent		Incongruent	
	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross
Synchronous	99.7	98.9	92.6	94.6	95.5	92.5	79.5	74.6	93.2	92.1	80.7	72.1	95.2	86.9	73.2	36.3
	(0.3)	(0.6)	(3.0)	(2.0)	(1.3)	(3.2)	(4.0)	(4.8)	(2.0)	(2.5)	(4.9)	(3.8)	(1.7)	(2.3)	(3.5)	(3.4)
Asynchronous	97.0	99.1	92.6	98.9	90.5	89.8	86.3	90.6	89.9	89.4	86.3	87.7	93.8	81.3	94.9	82.3
	(0.8)	(0.6)	(4.0)	(0.6)	(2.8)	(3.4)	(3.0)	(2.6)	(2.5)	(3.2)	(3.6)	(3.7)	(1.9)	(4.4)	(5.1)	(3.7)

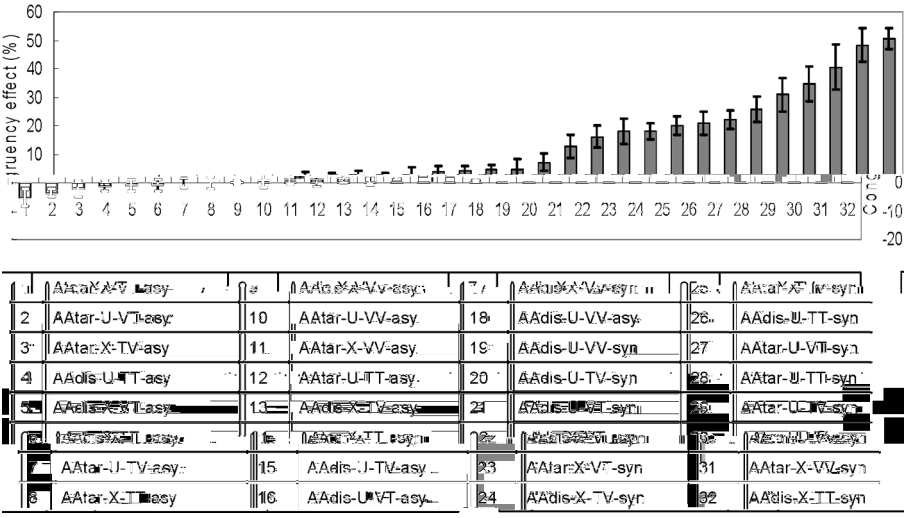


(A)

Figure 2. Congruency effects as a function of each sub-task/synchrony/modality combinations. For each sub-condition, the congruency effect was defined as the difference between proportion of correctly reporting the direction of target motion stream in the presence of directional congruent distractor motion stream and the proportion of correctly reporting the direction of target motion stream in the presence of directional incongruent distractor motion. (A) In the horizontal axis, from left to right there were four cluster of bars, representing the following four sub-tasks respectively: AA as target motion stream with the arms-uncrossed posture; AA as target motion stream with the arms-crossed posture; AA as distractor motion stream with arms-uncrossed posture; and AA as distractor motion stream with arms-crossed posture. Within each cluster, we used different color bars to depict the congruency effect for each modality/temporal combination: red for visual apparent motion with temporal synchrony condition (labeled as 'VVsyn'); purple for visual apparent motion with temporal asynchrony ('VVasy'); yellow for tactile apparent motion with temporal synchrony ('TTsyn'); green for tactile apparent motion with temporal asynchrony ('TTasy'); blue for visual to tactile apparent motion with temporal synchrony ('VTsyn'); gray for visual to tactile apparent motion with temporal asynchrony ('VTasy'); white for tactile to visual apparent motion with temporal synchrony ('TVsyn') and black for tactile to visual apparent motion with temporal asynchrony ('TVasy'). For the AA as distractor motion, each bar represents the congruency effect of a target motion — a given motion from combinations of visual/tactile modalities under certain temporal condition. For the AA as target motion stream, each bar represents the congruency effect of AA motion under all kinds of the distractor motion streams-visual/tactile AM. (B) A replot of sorting the congruency effect from the smallest to the largest across each sub-condition. The legends underneath the figure showed the detail of sub-condition. 'AAdis' refers to the condition that AA motion stream was the distractor event; 'AAtar' refers that AA motion stream was the target motion event; VV, TT, VT, TV is the modality combination as defined in the main text. 'U' refers to arms-uncrossed posture; 'X' refers to arms-crossed posture. This figure is published in colour in the online version.

(24.9%) than the one in arms-uncrossed condition (10.4%), $\blacktriangleright(1, 31) \neg 30.31$, < 0.001.

, for the interaction between arms posture conditions and temporal synchrony condition, at the synchrony level, the congruency effect in arms-crossed was significantly larger than the one in uncrossed condition



(B)

Figure 2. (Continued.)

(uncrossed: 14.4%; crossed: 23.2%), $F(1, 31) = 6.14$, $p < 0.05$. At the asynchrony level, the difference of percentages in arms-uncrossed was larger than the one in crossed condition, although both congruency effects were trivial (uncrossed: 2.8% and crossed: 0.1%), $F(1, 31) = 4.74$, $p < 0.05$. Likewise, at the arms-uncrossed condition, the congruency effect was larger in synchrony condition (syn: 14.4%; asyn: 2.8%), $F(1, 31) = 20.86$, $p < 0.001$; at the arms-crossed condition, the congruency effect was larger in synchrony condition (syn: 23.2%; asyn: 0.1%), $F(1, 31) = 110.84$, $p < 0.001$.

2.4.

For the influence of auditory distractors on the discrimination of the direction in the inter-modal visual/tactile apparent motion, we found some common characteristics, no matter whether the arms were crossed or not. Generally, when the directional cue in auditory motion and the directional cue in the inter-modal visual/tactile apparent motion are congruent, the accuracy of discrimination in the target motion streams is higher than the one in conflict directions. In addition, in both arms postures, when the two motion streams are presented synchronously, the discrimination accuracy in the target motion streams is lower than the one in asynchronous condition. This indicated that in cross-modal dynamic motion capture, the temporal factor (temporal synchrony) and spatial factor (directional congruency) are both important for the 'capture' effect to occur. Moreover, the congruency effect was larger when the two motion streams were synchronously presented. This finding is consistent with Soto-Faraco . (2002, 2004b) which found that the delayed auditory

stream imposes no influence on the judgment of visual or tactile apparent motion.

For different motion types, in both arms-uncrossed and arms-crossed conditions, the VV was least influenced by auditory distractors, and the TT was mostly influenced by auditory distractors. The smallest effect for the VV stream was consistent with earlier studies (Kitagawa and Ichihara, 2002; Soto-Faraco ., 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 ., 2005; Soto-Faraco ., 2004a). This result pattern could be accounted for by the modality appropriateness hypothesis (Welch and Warren, 1980, 1986), in which the visual modality dominates over tactile modality in spatial acuity, which renders the tactile events susceptible to the influence of auditory input. 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 by the integration of information of two modalities with associated weights during perception (Alais and Burr, 2004; Battaglia ., 2003; Chen and Zhou, 2011; Ernst and Banks, 2002; Witten and Knudsen, 2005).

Cross-experiments comparison revealed two major findings:

, we found a double dissociation between temporal synchrony condition and arms-posture in the congruency effects. In temporal synchronous presentation of two motion streams, the congruency effect was larger in arms-crossed posture than in arms-uncrossed posture; however, the result pattern was reversed in asynchrony condition, although the magnitude was largely reduced.

, the target motion stream in TT was interfered more in the arms-crossed condition than in the arms-uncrossed condition. This indicated that the spatial remapping between two conflicting spatial coordinates did have an additional and selective impact in cross-modal dynamic capture: the crossed-hands posture imposed a spatial remapping of sensory events (visual and/or tactile) between anatomical and external frames of reference. The congruency effect size was magnified when the two motion streams were presented synchronously, in which the remapping was not completed yet and thus made the motion direction judgment harder. Moreover, in the arms-crossed posture, the tactile events were mostly influenced due to the difficulty within tactile modality in resolving the conflicting of two spatial frames among all the motion combination in investigation (Azañón ., 2008, 2010).

3. Experiment 2

3.1.

Twelve undergraduate and graduate students (4 females, average age 24.2 years) attended Experiment 2a and another twelve undergraduate and graduate students (8 females, average age 24 years) attended Experiment 2b. 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. Participants were asked to make discrimination of the direction in auditory apparent motion in Experiment 2a (arms-uncrossed) and in Experiment 2b (arms-crossed), irrespective of the inter-modal visual/tactile apparent motion. The procedure was the same as in Experiment 1. In both sub-experiments, before the formal experiment, participants practiced discriminating the direction of apparent motion in the five different motion streams (AA — auditory motion, VV — visual motion, VT — visual to tactile motion; TV — tactile to visual motion; and TT — tactile motion). All the participants reached the criterion of at least 90% correct judgments in their first average 20 attempts for Experiment 2a (arms-uncrossed), and in their first average 30 attempts for Experiment 2b (arms-crossed).

3.2.

3.2.1. - /

( ) -

For the arms-uncrossed and arms-crossed conditions, the average proportion of correct responses (with the associated standard error) for each condition is presented in Table 2.

For the arms-uncrossed condition, an analysis of variance (ANOVA) with motion combination, temporal synchrony as two within-participant dependent-factors and congruency effect as dependent factor found a non-significant main effect of motion combination (VV: 20.8%, VT: 11.1%, TV: 17.4% and TT: 16.1%), $F(3, 33) = 2.284$, $= 0.097$, $\eta_p^2 = 0.172$. The main effect of temporal synchrony was also significant, with the congruency effect larger for the temporally synchronous presentation of auditory stimuli (33.0%) than for the delayed presentation of auditory stimuli (0.3%), $F(1, 11) = 38.099$, < 0.001, $\eta_p^2 = 0.776$. The interaction between motion combination and temporal condition was not significant, $F(3, 33) = 1.184$, $= 0.331$, $\eta_p^2 = 0.097$.

3.2.2. -

An analysis of variance (ANOVA) with motion combination, temporal synchrony as two within-participant dependent-factors and congruency effect as

Table 2.
Percentages of correct responses in Experiment 2, with numbers in brackets indicating standard errors

	VV				VT				TV				TT			
	Congruent		Incongruent		Congruent		Incongruent		Congruent		Incongruent		Congruent		Incongruent	
	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross	Norm	Cross
Synchronous.	95.5	97.2	54.9	49.0	91.7	87.8	66.0	69.8	94.1	91.7	59.4	70.8	91.7	81.9	60.8	84.4
	(1.9)	(1.5)	(8.3)	(6.6)	(3.2)	(2.6)	(5.2)	(3.2)	(1.8)	(3.2)	(7.6)	(5.2)	(2.1)	(3.7)	(6.2)	(3.6)
Asynchronous	92.4	92.0	91.3	90.6	89.6	85.8	93.1	91.0	89.6	88.5	89.6	90.6	93.1	92.4	91.7	92.4
	(2.1)	(2.6)	(3.4)	(2.1)	(3.4)	(5.6)	(3.3)	(4.0)	(3.5)	(5.4)	(4.1)	(3.0)	(2.4)	(3.2)	(1.8)	(2.8)

(VV: visual–visual motion; VT: visual–tactile motion; TV: tactile–visual motion; TT: tactile–tactile motion; Congruent — the direction of AA motion and the target motion stream was congruent; Incongruent — the direction of AA motion and the target motion stream was incongruent; Norm: arms were not crossed; cross: arms were crossed).

dependent factor found a significant main effect of motion combination (VV: 24.8%, VT: 6.4%, TV: 9.4% and TT: 1.2%), $F(3, 33) = 15.225$, $p < 0.001$, $\eta_p^2 = 0.581$. Pairwise comparisons showed that the congruency effect was largest in VV ($p < 0.05$). The main effect of temporal synchrony was also significant, with the congruency effect larger for the temporally synchronous presentation of auditory stimuli (21.2%) than for the delayed presentation of auditory stimuli (1.5%), $F(1, 11) = 66.256$, $p < 0.001$, $\eta_p^2 = 0.858$. The interaction between motion combination and temporal condition was not significant, $F(3, 33) = 27.235$, $p < 0.001$, $\eta_p^2 = 0.712$.

Further analysis showed that, for the delayed presentation of auditory stimuli, the congruency effects were not different among four motion combinations, $F(3, 33) = 1.373$, $p = 0.268$, $\eta_p^2 = 0.111$. For the synchronous presentation of the stimuli, the congruency effects were significantly different among four motion combinations, $F(3, 33) = 25.272$, $p < 0.001$; the congruency effect was largest in VV (48.3%) ($p < 0.05$), smallest in TT (2.4%) ($p < 0.01$), intermediate and no difference between VT (18.1%) and TV (20.8%), $p = 1$.

3.2.3. - vs -

We conducted cross-experiments comparisons between VTonAA Arms-uncrossed (Experiment 2a) VTonAA Arms-crossed (Experiment 2b) to compare the effect size of congruency effect. The arms posture was treated as between-subjects factor; the motion combination and temporal synchrony condition as two within-participant factors. The main effect of arms' posture was significant (congruency effect for arms uncrossed: 16.4% and for arms crossed: 9.9%), $F(1, 22) = 6$, $p = 0.023$, $\eta_p^2 = 0.214$. The main effect of motion combination was significant, the congruency effects are VV: 22.8%, VT: 8.8%, TV: 13.4% and TT: 7.5%, $F(3, 66) = 12.96$, $p < 0.001$, $\eta_p^2 = 0.371$. Pairwise comparison showed that the congruency effect was largest in VV, no difference between VT and TV, but smallest in TT ($p < 0.01$). The main effect of temporal synchrony was significant: the congruency effect was larger in temporal synchrony (27.1%) than in temporal asynchrony (0.9%), $F(1, 22) = 85.02$, $p < 0.001$, $\eta_p^2 = 0.794$. The two-way interaction between arms posture and motion combination was significant, $F(3, 66) = 5.207$, $p < 0.01$, $\eta_p^2 = 0.191$. The two-way interaction between arms posture and temporal synchrony was not significant, $F(1, 22) = 3.051$, $p = 0.095$, $\eta_p^2 = 0.122$. And the three-way interaction between arms posture, motion combination and temporal synchrony was significant, $F(3, 66) = 7.519$, $p < 0.01$, $\eta_p^2 = 0.255$.

We then conducted MANOVA simple effect analysis for the interaction between motion combination (VV, VT, TV, TT) and arms posture. On one hand, the congruency effect was not different among the four motion combinations in the arms-uncrossed condition (VV: 20.8%, VT: 11.1%, TV: 17.4% and TT:

16.1%), $F(3, 66) = 2.17$, $p = 1$, but the congruency effect differed in the arms-crossed condition (VV: 24.8%, VT: 6.4%, TV: 9.4% and TT: 1.2%), $F(3, 66) = 16$, $p < 0.001$. On the other hand, for AA as target, percentages of correct responses in TT were significantly different across different arms postures, with the congruency effect larger in the arms-uncrossed condition (16.1%) than the one in arm-crossed condition (1.2%), $F(1, 22) = 21.26$, $p < 0.001$.

For the interaction between arms posture conditions and temporal synchrony condition, at the synchrony level, the congruency effect in arms-crossed was significantly smaller than the one in uncrossed condition (uncrossed: 33.0%; crossed: 21.2%), $F(1, 22) = 4.68$, $p < 0.05$. At the asynchrony level, congruency effects were not different and trivial (uncrossed: 0.3% and crossed: 1.5%), $F(1, 22) = 0.54$, $p = 0.470$. Likewise, at the arms-uncrossed condition, the congruency effect was larger in synchrony condition (syn: 33.0%; asyn: 0.3%), $F(1, 22) = 60.14$, $p < 0.001$; at the arms-crossed condition, the congruency effect was also larger in synchrony condition (syn: 21.2%; asyn: 1.5%), $F(1, 22) = 27.93$, $p < 0.001$.

3.3.

For different motion types (distractors), different to Experiment 1, in the arms-uncrossed condition, the congruency effects among all the combinations of visual/tactile events were equal, it was different to the Sanabria . (2005a) results, where the distractors of bimodal distractor (visuotactile to visuotactile apparent motion) and visual apparent motion had a larger capture effect on auditory motion than did in the tactile apparent motion. We speculated that in Sanabria . (2005a), by introducing the bimodal motion (visuotactile to visuotactile apparent motion), the perceived relative motion saliency of different distractors might change and thus contribute to differential capture effect, such as the bimodal motion formed a more stronger multisensory motion signal and imposed a stronger capture effect. Conversely, for the arms-crossed condition, in comparison with Experiment 1 (AA as distractor, hand-crossed), the performance of discrimination of auditory motion was more harmed in the VV (24.0%) than in the TT (1.2%) condition, and the effect was obviously observed in temporal synchronous condition. Furthermore, in general, the directional congruency effect was larger in the arms-uncrossed condition (16.4%) than in the arms-crossed condition (9.9%). Typically, the congruency effect was larger in the TT-arms-uncrossed (16.1%) than in the TT-arms-crossed (1.2%) condition. The capture effect reported in the inter-modal distractors (visual and tactile events) with arms-crossed condition was modulated both by cross-modal interaction based on the functional appropriateness of individual sensory events and the spatial remapping process between different sensory events.

, with different sensory events, when dealing with the spatial attributes of multisensory stimuli, vision has been shown to dominate over both audition and touch, presumably because visual cues typically provide the most precise/appropriate information regarding stimulus location (Battaglia ., 2003; Bertelson and Aschersleben, 1998). In Experiment 1, the visual apparent motion stream was the target, so it was least influenced by the auditory motion stream, but in Experiment 2, the visual apparent motion being the distractor, it could impose a rather strong influence on the auditory motion. Hence we observed a reserved pattern of correct percentages between reporting of directions of VV and TT as targets and reporting directions of AA when VV and TT acted as distractors.

, tactile events are more susceptible to the spatial remapping process (Azañón ., 2008, 2010) for the direction of tactile apparent motion (TT) incongruent to direction of auditory apparent motion, in the arms-uncrossed condition. There was no conflict between the internal spatial reference (anatomical positions of middle fingers) and the external spatial reference (the actual positions of stimuli), and the congruency of two spatial references has boosted the reliability of using TT directional cue and imposed a strong interference (capture) effect of TT over AA when the two streams were not congruent in directions. However, for the arms-crossed condition, the conflict and remapping between two spatial references has weakened the reliability of using directional cue affiliated with TT and the influence of TT over AA was decreased accordingly.

3.3.1. *Sh* *Sh* - *Sh* -

One might argue that the differential influences between auditory apparent motion and intermodal visual/tactile apparent motion with different arms postures were simply due to the differences in the perceived quality or strength towards these motion streams. To rule out this possibility, we asked an additional 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, with arms-uncrossed. 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 and 4.2 for AA (auditory-auditory apparent motion). We asked another group of ten participants (4 males, mean age of 22.7 years) to rate also 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, but with the arms-crossed (the left-middle finger on the right tactile actuator and right-middle finger on the left tactile actuator). The mean scores were 5.0 for the VV stimuli, 4.4 for the TT stimuli,

4.0 for either the VT or TV stimuli and 4.1 for AA. We took the arms posture as between-subjects factor and the five types of motion as within-subjects factors and made cross-experiments comparison. The main effect of motion types was significant, $F(4, 72) = 8.29$, $p < 0.001$. Bonferroni corrected comparisons showed that the strength of apparent motion was significantly stronger for the VV stimuli than for VT, TV and AA stimuli ($ps < 0.05$), but there was no significant difference between VV stimuli and TT stimuli ($p = 0.536$). The main effect of arms posture was not significant, $F(1, 18) = 0.75$, $p = 0.40$, and the interaction between motion type and arms posture also was not significant, $F(4, 72) = 0.70$, $p = 0.60$. The results suggested that the different motion capture pattern in arms-uncrossed and arms-crossed conditions is not largely determined by the perceived differences of the apparent motion across the two conditions.

4. Experiment 3

Across Experiment 1 and Experiment 2, we used between-subjects design and the sub- experimental conditions were trial-by-trial randomized. In Experiment 3, we tested explicitly for each condition combination (with each arm posture, temporal synchrony and modality combination), using block-wise design

4.1.

Twelve undergraduate and graduate students (7 females, average age 22.1 years) attended Experiment 3. Participants completed four sub-tasks. They were asked to make discrimination of the direction in auditory apparent motion (AA) or inter-modal visual/tactile apparent motion (VV, VT, TV, TT), with arms crossed or arms uncrossed. The orders of four sub-tasks among all the participants were arranged with the method of the Latin Square. The procedure was the same as in Experiment 1 and Experiment 2. Before the formal experiment, participants practiced discriminating the direction of apparent motion in the five different motion streams (AA, VV, VT, TV and TT).

4.2.

As suggested by one of the reviewers, the congruency effect should be given explicitly for each sub-condition (Fig. 3A). As in previous analysis, we sifted out the capture effect (congruent incongruent) with each sub-condition and re-plot the figure, with the order of increasing ‘capture size’ (see Fig. 3B), which gives a whole picture of the congruency effect across all kinds of sub-conditions, where we can see clearly an increasing congruency effect in the temporal synchronous condition.

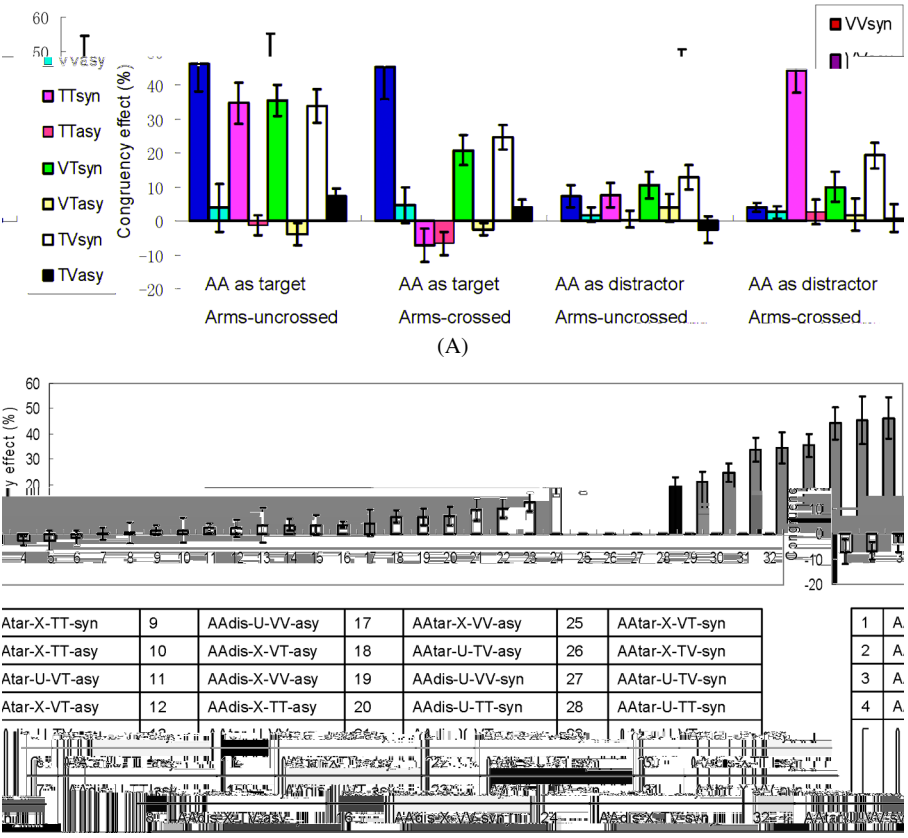


Figure 3. Congruency effects as a function of each sub-task/synchrony/modality combinations in Experiment 3. Here we used a block wise design. The designation and connotation of each individual cluster and bar is the same as in Fig. 2. This figure is published in colour in the online version.

We conducted repeated measures ANOVA. The main effect of the sub-tasks (AA as target with arms crossed/uncrossed; AA as distractor with arms crossed/uncrossed) was significant: $F(3, 44) = 9.047$, $p < 0.001$, $\eta_p^2 = 0.382$. Pairwise comparison showed that in the task AAtarget-uncrossed, the mean congruency effect is the largest (19.5%), $p < 0.05$. (AAtarget-crossed: 10.4%; AAdistractor-uncrossed: 5.2%; and AAdistractor-crossed: 10.6%). However, the main effect of modality combination (VV, TT, VT, TV) was not significant (VV (14.4%); TT (9.3%); VT (9.5%) and TV (12.4%)), $F(3, 132) = 2.563$, $p = 0.058$, $\eta_p^2 = 0.055$. The main effect of temporal synchrony was significant (Synchronous: 21.8% > Asynchronous: 1.0%), $F(1, 44) = 132.57$, $p < 0.001$, $\eta_p^2 = 0.751$. The interaction between sub-tasks and modality combination was significant, $F(9, 132) = 8.862$, $p < 0.001$,

$^{2}_{p}$, 0.377; the interaction between sub-tasks and temporal synchrony condition was significant, (3, 44) , 9.996, < 0.001, $^{2}_{p}$, 0.405. The three-way interaction between sub-tasks, modality combination and temporal synchrony was significant, (9, 132) , 6.724, < 0.001, $^{2}_{p}$, 0.314.

We then conducted MANOVA simple effect analysis towards the interaction between motion combination (VV, VT, TV, TT) and arms posture. For AAtarget-uncrossed, the main effect of modality combination was not significant, (3, 132) , 1.85, , 0.141; for AAtarget-cross, the main effect of modality combination was significant, (3, 132) , 18.59, < 0.001. The congruency effect (congruency incongruence) relation was VV(25%) > VT(9.2%) , TV(14.2%) > TT(6.9%) (VV TT, < 0.01; TT VT, < 0.01; TT TV, < 0.001). For AAdistractor-uncrossed, the main effect of modality combination was not significant, (3, 132) < 1, , 0.894; for AAdistractor-crossed, the main effect of modality combination was significant, (3, 132) , 8.50, < 0.001. The congruency effect (congruency incongruence) relation was TT(23.3%) > VT(5.8%) , TV(10.0%) > VV(3.1%). (VV TT, < 0.01; TT VT, < 0.05; TT TV, < 0.05). On the other hand, we compared each motion combination across four sub-tasks. For VV, (3, 44) , 7.50, < 0.001, the relation of congruency effects over four sub-tasks were: AAtarget-uncrossed (25%) , AAtarget-crossed (25%) > AAdistractor-uncrossed (4.4%) , AAdistractor-crossed (3.1%); for VT, (3, 44) , 2.15, , 0.108; for TV, (3, 44) , 6.77, < 0.01, the relation of congruency effects were: AAtarget-uncrossed (20.4%) , AAtarget-crossed (14.2%) > AAdistractor-uncrossed (5.2%) , AAdistractor-crossed (10.0%). For TT, (3, 44) , 21.22, < 0.001, AAdistractor-crossed (23.3%) > AAtarget-uncrossed (16.7%) > AAdistractor-uncrossed (4.0%) > AAtarget-crossed (6.9%).

For the interaction between sub-tasks and temporal condition, we tore apart into temporal synchrony and temporal asynchrony conditions. The interaction between sub-tasks and temporal synchrony was significant, (3, 44) , 12.77, < 0.001, with the congruency effect in AAtarget-uncrossed (37.5%) being the largest. However, the interaction between sub-tasks and temporal asynchrony did not reach significance, (3, 44) , 0.22, , 0.883.

4.3.

With block design, we replicated the cross-modal dynamic capture pattern as in the trial-by-trial design (Expt. 1 and Expt. 2), the congruency effect was larger in temporal condition than in temporal asynchrony condition. Interestingly, across the four sub-tasks (AAtarget-uncrossed, AAtarget-crossed, AAdistractor-uncrossed, AAdistractor-crossed), the congruency effect was largest in - condition, and it was mainly observed with temporal condition in synchrony. This pattern reflects the general

temporal-spatial law governing the cross-modal dynamic capture: temporal synchrony make it hard to resolve the correspondence between motion streams, while the segregation of two motion streams in temporal asynchrony condition has lead to perceptual separation of the two streams, which has reduced the interaction between the motion streams (Soto-Faraco ., 2004a, b). Sound signals are generally weak in spatial localization, so discrimination of the direction in auditory apparent motion was readily interfered in the presence of visual/tactile apparent motion. Furthermore, the arms-uncrossed posture provides a pure foundation for the interaction of different motion streams, free from the additional impact of ambiguity in resolving two spatial frames and thus to maximize the congruency effect in cross-modal dynamic capture: we will come to this point in detail in General Discussion.

For different modality combinations (VV, VT, TV and TT), however, since there were opposite trends of congruency effects across AA as distractor (congruency effect: VV < VT ¯ TV < TT) and AA as target conditions (congruency effect: VV > VT ¯ TV > TT), in which the differences among the four motion combinations were seen in AAtarget-crossed and AAdistractor-crossed conditions, respectively, the collapsed main effects of motion combinations were neutralized and did not attain statistical significance. It can be inferred that among all the modality combinations, visual apparent motion and tactile apparent motion are two signatures to depict the modality appropriateness in cross-modal dynamic capture, dependent on whether the auditory signals serve as targets or distractors. Interestingly, the interaction effect in modality combination collapsed over arms-posture and sub-tasks conditions were observed typically in arms-crossed posture, this indicated that in addition to the common underlying cross-modal interaction between different motion streams, the spatial remapping (especially for tactile events) has imposed the differential congruency effect among the four modality combinations (VV, VT, TV and TT).

5. Experiment 4

For the observed influence of distractors (inter-modal apparent motion) upon the auditory motion stream in the cross-hands condition, the conflicting spatial information was given by two (activated) stimuli bearing conflicting spatial coordinates (external coordinate anatomical coordinate), the two stimuli consisted of 'apparent motion'. It would be interesting to examine whether and how the conflicting spatial information, activated by only single stimulus, i.e. static distractors (visual or tactile stimulus), influences the cross-modal apparent motion capture. Furthermore, one might argue that the capture effect produced by the cross-modal distractors (Experiments 1–3) was probably contributed by the effect of the static components of the distractor on the per-

ception of the auditory stimuli, the purpose of carrying out Experiment 4 also addresses this alternative possible account. We took the work of Sanabria . (2005a). (Experiment 2, . .) and used within-subject design, but curtailed the distractor types to incorporate key Distractor conditions and reduced experimental length, as described in the following session.

5.1.

Fourteen undergraduate and graduate students (5 females, average age 22.6 years) attended Experiment 4. Participants completed two sub-tasks. They were asked to make discrimination of the direction in auditory apparent motion (AA) in the presence of five types of distractors (as in contrast to the 8 conditions in Sanabria ., 2005), with arms crossed or arms uncrossed. The conditions of Distractor types are: (1) V1-T2. Visual stimulus was presented with the first sound in the auditory apparent motion stream and the tactile stimulus was presented with the second sound; (2) T1-V2. Tactile stimulus was presented with the first sound and visual stimulus at the second sound; (3) V1. Only one visual stimulus was presented with the first sound; (4) T1. Only one tactile stimulus was presented with the first sound; (5) V1-ST1. Both visual stimulus and tactile stimulus (at same spatial location) presented simultaneously with the first sound. In conditions of distractors with apparent motion (condition 1 and condition 2), the direction of distractor motion (VT or TV) could be congruent or incongruent with the direction of AA motion. For the static distractors (V or T), the congruent trials were defined as the static distractors occurred at the same location as the first sound of AA motion stream, and the incongruent trials were defined as the static distractors appeared at the opposite location to the first sound in AA motion stream. The distractors stimuli could be presented synchronously or asynchronously with the AA motion stream, as implemented in the Experiments 1–3. Participants attended two sessions of Experiment 4: Experiment 4a (hands-uncrossed) and Experiment 4b (hands-crossed). The order of the two sessions was randomized and counterbalanced across all participants. The experiment procedure was the same as in Experiment 3. Before the formal experiment, participants received practice about discriminating the direction of apparent motion in the three different motion streams (AA, VT, TV).

5.2.

We conducted cross-experiments comparisons between VTonAA Arms-uncrossed (Experiment 4a) VTonAA Arms-crossed (Experiment 4b) to compare the effectsize of congruency effect. The arms posture was treated as between-subjects factor; the Distractor types and temporal synchrony condition as two within-participant factors.

The main effect of arms posture was not significant (congruency effect for arms uncrossed: 13.5% and for arms crossed: 9.4%), $F(1, 26) = 1.378$, $= 0.251$, $\eta_p^2 = 0.050$. The main effect of temporal synchrony condition was significant, $F(1, 26) = 50.867$, < 0.001, $\eta_p^2 = 0.662$, with high congruency effect in synchronous condition than the one in asynchronous condition (29.1% 6.1%, respectively). The main effect of Distractor type was significant, $F(4, 104) = 10.540$, < 0.001, $\eta_p^2 = 0.288$. The congruency effects are V1-T1: 26.5%; T1-V2: 6.2%; V1: 13%, T1: 2.2% and V1-ST1: 13.8%. Pairwise comparison showed that the congruency effect was largest in V1-T1 (s < 0.01), but no difference between V1-T1 and V1-ST1 ($= 0.551$).The congruency effect was higher in V1 than in T1 (< 0.01), and the congruency effect was higher in V1-ST1 than the one in T1 condition (< 0.001). The interaction between Distractor type and hand posture was not significant, $F(4, 104) = 0.468$, $= 0.759$, $\eta_p^2 = 0.018$; the interaction between temporal synchrony and hand posture was also not significant, $F(1, 26) = 2.597$, $= 0.119$, $\eta_p^2 = 0.091$.

The interaction between Distractor type and synchrony was significant, $F(4, 104) = 7.606$, < 0.001, $\eta_p^2 = 0.226$. We then separated the analysis into synchrony and asynchrony conditions. In the synchrony condition, the main effect of Distractor type was significant, $F(4, 104) = 15.026$, < 0.001, $\eta_p^2 = 0.366$. The congruency effects were V1-T1: 46.1%, T1-V2: 16.1%; V1: 35.0%, T1: 7.3% and V1-ST1: 41.1%. Pairwise comparison showed the congruency effect was largest in V1-T1 (s < 0.01), smallest in T1 (s < 0.01), but no difference among the three conditions- V1-T1, V1 and V1-ST1 (s > 0.2). In the asynchronous condition, the main effect of Distractor type was significant, $F(4, 104) = 3.620$, < 0.01, $\eta_p^2 = 0.122$. The congruency effects are V1-T1: 7.0%; T1-V2: 3.6%; V1: 8.9%, T1: 11.8% and V1-ST1: 13.4%. Pairwise comparison showed the congruency effects were statistically the same across the five Distractor types (s > 0.3) (Fig. 4).

5.3.

With the introduction of static stimuli (visual or tactile stimulus), we replicated the main findings in Sanabria . (2005a). The capture effect of inter-modal apparent motion (V1-T2) was largely contributed by the static components — a single visual stimulus; and the first distractor (V1) seemed to induce a static ventriloquist effect on the temporally coincident sound; the visual stimulus impaired more significantly on the AA motion than did the tactile stimulus. Interestingly, the capture effect of inter-modal apparent motion (V1-T2) was equal to the effect imposed by compound stimuli (visual stimulus together with tactile stimulus at the same location, i.e. V1-ST1). However, in stark contrast to the previous Experiment 2 and Experiment 3, we did not observe the modulation effect of hands- posture of distractors upon AA motion. It

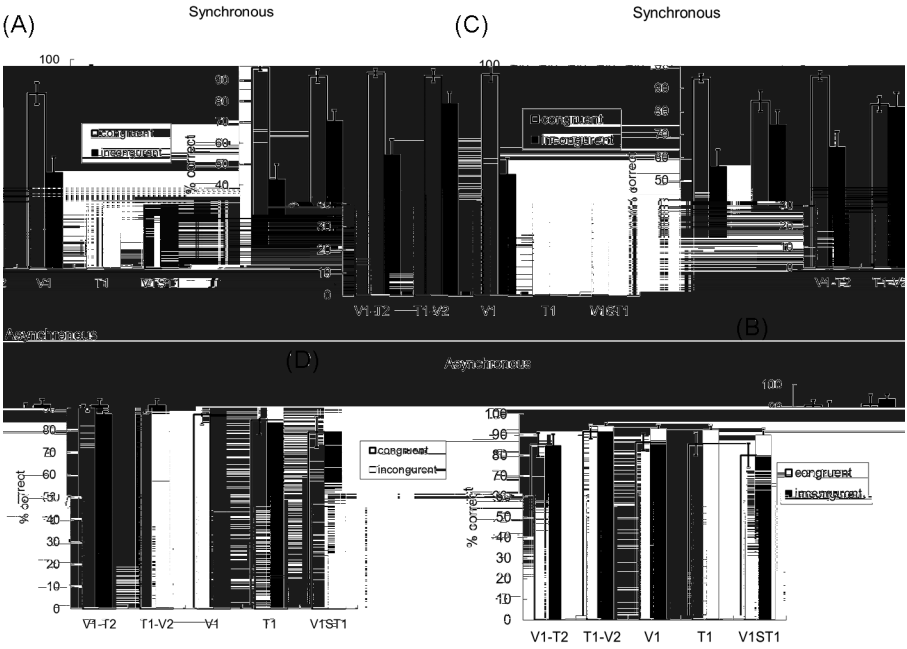


Figure 4. Graph showing the mean accuracy (SE) in discriminating the direction of auditory apparent motion in the uncrossed hand posture (A, B) and crossed hand posture (C, D) in Experiment 4, as a function of factors with Distractor type, temporal synchrony and congruency. The conditions of Distractor types are: (1) V1-T2: Visual stimulus was presented with the first sound in the auditory apparent motion stream and the tactile stimulus was presented with the second sound; (2) T1-V2: Tactile stimulus was presented with the first sound and visual stimulus at the second sound. (3) V1: visual stimulus was presented with the first sound; (4) T1: tactile stimulus was presented with the first sound; (5) V1-ST1: both visual stimulus and tactile stimulus (at same spatial location) presented with the first sound.

indicates that the activated spatial relation between two positions (as in Experiment 1–3), given by the distractors of apparent motion, is crucial for the spatial remapping (in visual and tactile events) to take effect in cross-modal dynamic capture.

6. General Discussion

Previous investigations into dynamic cross-modal capture have been mostly confined to two given modalities, such as dynamic direction cues in one modality affecting the perception of motion stream in another modality. The present study extended Chen and Zhou (2011) and explored mutual influences of inter-modal visual/tactile apparent motion and auditory apparent motion, taking into account the spatial reference frames of sensory events. In conformity to previous bunches of studies, the general findings are that the temporal

synchrony and spatial congruency as well as the modality appropriateness still work in the mutual influences in cross-modal dynamic capture (Alais and Burr, 2004; Bresciani ., 2006; Calvert ., 2000; Gepshtein ., 2005; Macaluso ., 2004; Shi ., 2010; Slutsky and Recanzone, 2001; Spence ., 2007; Stein and Meredith, 1993). The congruency effect (direction-incongruent condition direction-congruent condition) was obviously observed when the two motion streams were in the temporal synchrony condition. For different modality combinations, the VV motion imposed a large capture effect on AA motion but AA motion as distractor had a lesser influence on VV motion. In the arms-uncrossed posture, tactile motion affected the auditory motion judgments while the effect was largely reduced in the arms-crossed posture. The capture size of auditory events VT and TV or TV and VT over auditory events was intermediate between those in VV and TT, with the exception that TV motion impact AA motion, to a large extent resemblance of TT motion over AA motion (Sanabria ., 2005b).

It is widely agreed that cross-modal interactions depend on the relative reliability of the sensory signals contributing to the perception of a particular multisensory event (Ernst and Bülthoff, 2004). Accordingly, the cross-modal capture effect results from an optimal integration of the spatial information from different sensory events. Here in cross-modal dynamic capture, the reliability of spatial information is co-determined by two sources:

, the bias prior (a type of prior knowledge) of the spatial remapping of distractor events with different hands-postures (Ernst and Di Luca, 2011).

, the reliability of spatial/temporal function associated with a specific sensory event, established as 'functional appropriateness and precision' (Welch and Warren, 1980, 1986). In the cross-modal dynamic capture, for the second source (corresponding to an 'integration' process), the highest spatial resolution or temporal acuity of sensory events in one modality dominates the perception of events in the other modality. The visual events are more superior in the spatial acuity than are the tactile events, so that when the visual events and tactile events are targets (Experiment 1), the discrimination of visual apparent motion was least influenced by auditory apparent motion and tactile apparent motion was most affected by auditory apparent motion. Conversely, when the visual events and tactile events are distractors (Experiment 2), the discrimination of direction in auditory apparent motion was more influenced by visual apparent motion than was by tactile apparent motion.

More importantly, in addition to the 'integration' process, the 'remapping' of the two spatial coordinates in the distractors mediates the cross-modal dynamic capture. Different sensory events could be encoded in two spatial references, that is, in their spatial locations (absolute reference frames) or in $\$_n$ positions ($\$_n$ reference frames). The tactile events

tile stimuli are remapped into externally defined coordinates and this spatial remapping takes longer to achieve when external and anatomically centered codes are in conflict as when the hands adopt a crossed-hands posture (Eimer ., 2001; Shore ., 2002; Maravita ., 2003; Yamamoto and Kitazawa, 2001). When the two coordinates were not in consensus, a spatial realignment (remapping), usually from the somatotopic reference frames (takes short time) to the external reference frames (takes longer) would resolve this conflict (Azañón and Soto-Faraco, 2008). In Experiment 1, for TT as target, the congruency effect size was larger in the arms-crossed condition (24.9%) than in the arms-uncrossed condition (10.4%), the result pattern was reversed in Experiment 2, the congruency effect of AA in the presence of TT distractors with arms-crossed is lower (1.2%) than the one with arms-uncrossed (16.1%). The result pattern could be explained by an interplay mechanism: 'remapping' of the two spatial coordinates in distractor events and 'integration' of the relatively reliabilities of spatio-temporal information across different cross-modal events. Specifically, in Experiment 1 (**AAonVT**), the auditory stimuli as distractors are weak in the spatial localization, so that the spatial relations between visual/tactile events were mainly modulated by the spatial remapping between visual/tactile events, i.e. the intra-targets perceptual grouping might have the upper hand over the cross-modal interaction (auditory influence on visual/tactile events). In the arms-crossed condition, the conflict between two spatial references has made it difficult for discrimination of the direction of tactile apparent motion, so that the reliability of using directional cues in the visual/tactile events was decreased. For Experiment 2 (**VTonAA**), the distractors were visual/tactile events, which are functionally stronger in spatial localization, and the target auditory motion stream was functionally weak in spatial localization (susceptible to the influence of visual/tactile distractors). The conflict between two spatial references (arms-crossed) might have decreased the intra-distractors perceptual (directional) grouping of tactile events, the reliability of using directional cues in tactile distractors was decreased, hence the effectiveness of the cross-modal capture of tactile events upon auditory apparent motion is interfered. Previous studies using spatial ventriloquism have also suggested that crossing hands may decrease the relative reliability of tactile spatial information due to the conflict between anatomical and external coordinates (Shore ., 2002; Yamamoto and Kitzawa, 2001), and thereby reduces the influence of tactile information on multisensory integration as well (Bruns and Rőder, 2010; Sanabria ., 2005a, b; Soto-Faraco ., 2004a, b). Similarly, the current study (Experiment 2 and Experiment 3) suggests that the reduced reliability of using directional cues for tactile motion stream (in the arm-crossed situation) constrained the 'capture' effect.

For the inter-modal distractors (inter-modal apparent motion), a general finding here was that the effect of direction cues in intermodal visual/tactile

apparent motion (VT and TV) was intermediate between its effect upon the perception of intra-modal visual (VV) and tactile (TT) apparent motion. Two factors might operate to achieve the 'intermediate' effect — one is the spatial resolution of intermodal visual/tactile apparent motion appears to be the mean of the resolutions of vision and touch. Henceforth the effect of auditory direction cues appears to be the mean of the susceptibilities of vision and touch (Experiment 1) and the degree of intra-distractors perceptual grouping also seems to be the mean between intra-modal VV and TT apparent motion. The second factor is the spatial remapping between visual/tactile events, although the tactile events are more influenced by the spatial conflict between two spatial references, the visual events were less influenced by this conflict and thus mitigated the influences of spatial cues in intermodal visual–tactile or tactile–visual apparent motion streams. Nevertheless, the impact of VT motion upon AAmotion somehow resembles the effect of TT motion over AA motion.

One might argue that the cross-modal dynamic capture effect could be contributed by the static stimuli in the distractor motion stream. Sanabria . (2005a) found the capture effects of TV and TT over AA were similar, and their finding suggests the first stimulus of the distractor motion stream play a key role in the cross-modal dynamic capture effect. In Experiment 4, we introduce the key static events (visual or tactile stimuli). The results suggest that the capture effect of inter-modal apparent motion (V1-T2) is largely contributed by the static components. A single visual stimulus or compound stimuli (visual stimulus with tactile stimulus at the same location, i.e. V1-ST1) and the first distractor (V1) seems to induce a static ventriloquist effect on the temporally coincident sound, replicating the results of Sanabria . (2005a). For the present study, we cannot rule out this possibility of static ventriloquist-like effect, that is, the location of first auditory stimulus might have been mislocalized toward the location of the first distractor stimulus (visual or tactile). Importantly, in stark contrast to the full-remapping with two opposite spatial locations (Experiments 2–3), the modulation effect of hands-posture on AA motion was not observed in half-remapping condition (with static distractor events). It indicates a full-remapping of a directional cue, given by the intra-modal or inter-modal apparent motion, is crucial and would maximize the effect of spatial remapping in cross-modal dynamic capture.

Using both trial-by-trial design and block-wise design, we found the similar picture of cross-modal dynamic capture. This suggests the interaction of motion streams from different modalities is robust and less influenced by the adaptation of a given motion combination in a short time (Soto-Faraco ., 2004). Moreover, the mutual influence between auditory apparent motion and visual/tactile apparent motion could not be simply reduced to the differential motion saliency of the different motion types across arms-uncrossed and arms-crossed conditions. We compared the subjective reports of the 'strength'

of apparent motion: the strength of apparent motion was significantly stronger. The strength of apparent motion was significant stronger for the VV stimuli than for VT, TV and AA stimuli (s < 0.05), but there was no significant difference between VV stimuli and TT stimuli. The main effect of arms posture was not significant nor the interaction between motion type and arms posture. The results suggested that the different motion capture pattern in arms-uncrossed and arms-crossed conditions is not simply determined by the perceived differences of the apparent motion across the two conditions.

To conclude, by presenting direction cues between auditory stimuli and visual/tactile stimuli and asking participants to judge the direction of apparent motion caused by the sequential presentation of visual and/or tactile stimuli or auditory stimuli (distractors) at two spatial locations, we have demonstrated there are mutual but asymmetric influences between auditory events and visual/tactile events in cross-modal motion capture. The relative reliabilities of spatial/temporal information with the stimuli signals, especially from the distractor modalities, together with a spatial remapping process, could account for the asymmetric cross-modal dynamic capture.

This research was supported by grants from the Natural Science Foundation of China (71102113, 71090402, 31200760), China Postdoctoral Science Foundation (20100480112, 201104032), the Fundamental Research Funds for the Central Universities, Sichuan Social Sciences Funding Program (SC12C009) and National High Technology Research and Development Program of China (863 Program) (2012AA011602).

References

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

Azañón, E. and Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events, . . **18**, 1044–1049.

Azañón, E., Camacho, K. and Soto-Faraco, S. (2010). Tactile remapping beyond space, . . . **31**, 1858–1867.

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

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

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

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

Bresciani, J. P. and Ernst, M. O. (2007). Signal reliability modulates auditory–tactile integration for event counting, **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, . . **162**, 172–180.

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

Bruns, P. and Rőder, B. (2010). Tactile capture of auditory localization is modulated by hand posture, . . **57**, 267–274.

Burtt, H. E. (1917). Auditory illusions of movement. A preliminary study, . . . **2**, 63–75.

Calvert, G. A., Campbell, R. and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of cross-modal binding in the human heteromodal cortex, . . **10**, 649–657.

Chen, L. and Zhou, X. (2011). Capture of intermodal visual/tactile apparent motion by moving and static sounds, **24**, 369–389.

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

Eimer, M., Cockburn, D., Smedley, B. and Driver, J. (2001). Cross-modal links in endogenous spatial attention are mediated by common external locations: evidence from event-related brain potentials, . . **139**, 398–411.

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

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, . . **8**, 162–169.

Ernst, M. and Di Luca, M. (2011). Multisensory perception: from integration to remapping, in: , J. Trommershäuser (Ed.), pp. 225–250. Oxford University Press, New York, NY, USA.

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

Galli, P. A. (1932). Über mittelst verschiedener Sinnesreize erweckte Wahrnehmung von Scheinbewegungen [On the perception of apparent motion elicited by different sensory stimuli], . *Si* **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, . **5**, 1013–1023.

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

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

Kafaligonul, H. and Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals, . **10**, 1–13.

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

Kirman, J. H. (1974). Tactile apparent movement: the effects of number of stimulators, . . . **103**, 1175–1180.

Kolers, P. A. (1972). . Pergamon Press, Oxford, UK.

Korte, A. (1915). Kinematoskopische Untersuchungen, *Z* **72**, 193–206.

Lakatos, S. and Shepard, R. N. (1997). Constraints common to apparent motion in visual, tactile and auditory space, . . .: ′ . . **23**, 1050–1060.

Lloyd, D. M., Merat, N., McGlone, F. and Spence, C. (2003). Cross-modal links between audition and touch in covert endogenous spatial attention, . . **65**, 901–924.

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

Macaluso, E., George, N., Dolan, R., Spence, C. and Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study, **21**, 725–732.

Maravita, A., Spence, C. and Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach, . **13**, R531–R539.

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

Occelli, V., Spence, C. and Zampini, M. (2009). The effect of sound intensity on the audiotactile cross-modal dynamic capture effect, . . **193**, 409–419.

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, . **1242**, 200–208.

Sanabria, D., Soto-Faraco, S. and Spence, C. (2005a). Assessing the effect of visual and tactile distractors on the perception of auditory apparent motion, . . . **166**, 548–558.

Sanabria, D., Soto-Faraco, S. and Spence, C. (2005b). Spatiotemporal interactions between audition and touch depend on hand posture, . . . **165**, 505–514.

Sanabria, D., Soto-Faraco, S. and Spence, C. (2007a). Spatial attention modulates audiovisual interactions in apparent motion, . . .: ′ . **33**, 927–937.

Sanabria, D., Spence, C. and Soto-Faraco, S. (2007b). Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: a signal detection study, **102**, 299–310.

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

Shore, D. I., Spry, E. and Spence, C. (2002). Confusing the mind by crossing the hands, . . **14**, 153–163.

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

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, . . **14**, 139–146.

Soto-Faraco, S., Kingstone, A. and Spence, C. (2003). Multisensory contributions to the perception of motion, **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, ... **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, . . .: ′ . **30**, 330–345.

Spence, C., Sanabria, D. and Soto-Faraco, S. (2007). Intersensory Gestalten and cross-modal scene perception, in: : ′ - , K. Noguchi (Ed.). Nihon University College of Humanities and Sciences, Tokyo, Japan.

Stein, B. E. and Meredith, M. A. (1993). , MIT Press, Cambridge, MA, USA.

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

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

Welch, R. B. and Warren, D. H. (1986). Intersensory interactions, in: ′ ′ , Vol. 1, K. R. Boff, L. Kaufman and J. P. Thomas (Eds), pp. 1–36. John Wiley and Son, New York, USA.

Welch, R. B., Duttonhurt, L. D. and Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception, . . **39**, 294–300.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung,*Z* **61**, 161–165.

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

Yamamoto, S. and Kitazawa, S. (2001). Reversal of subjective temporal order due to arm crossing, . **4**, 759–765.

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