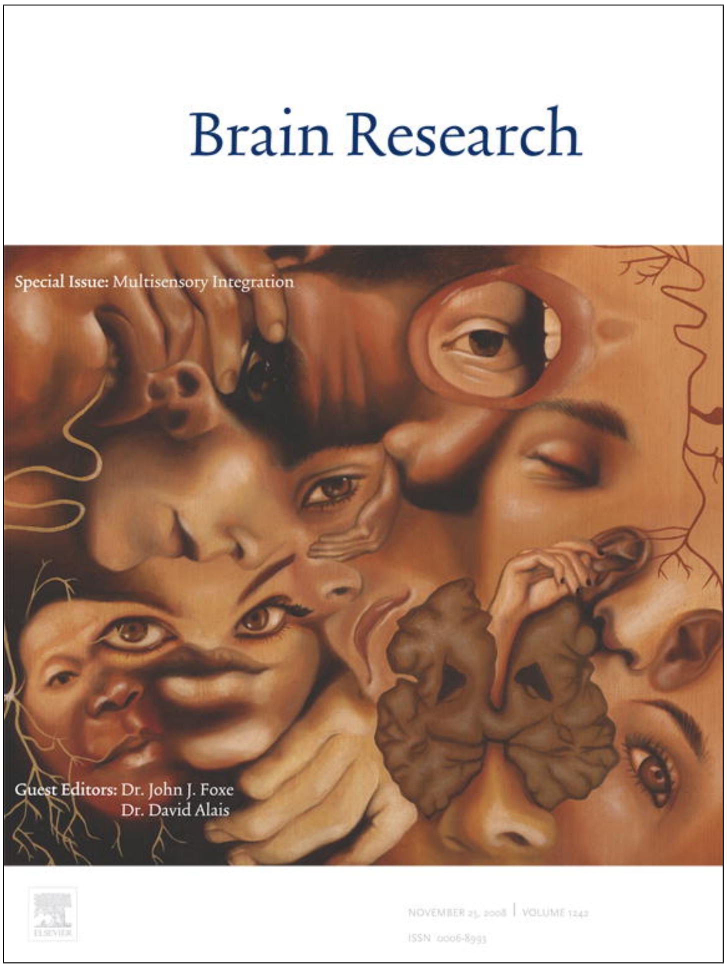

Brain Research

Special Issue: Multisensory Integration

Guest Editors: Dr. John J. Foxe
Dr. David Alais

NOVEMBER 25, 2008 | VOLUME 1242
ISSN 0006-8993

BRAIN RESEARCH 1242 (2008) 200–208

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

The effect of attention on the illusory capture of motion in bimodal stimuli

Ipek Oruc[a,*], *Scott Sinnett*[b], *Walter F. Bischof*[c], *Salvador Soto-Faraco*[d], *Karman Lock*[b], *Alan Kingstone*[b]

[a]*Department of Ophthalmology and Visual Sciences, University of British Columbia, Canada*
[b]*Department of Psychology, University of British Columbia, Canada*
[c]*Department of Computing Science, University of Alberta, Canada*
[d]*ICREA and Parc Científic de Barcelona, Universitat de Barcelona, Spain*

ARTICLE INFO

Article history:
Accepted 2 April 2008
Available online 16 April 2008

Keywords:
Multisensory integration
Apparent motion
Illusory capture
Motion perception
Crossmodal dynamic capture

ABSTRACT

A large body of work now exists that demonstrates the interaction between different sensory modalities when they are integrated into a single coherent percept. Yet, it is not yet clear whether attention plays a critical role in such crossmodal interactions. We investigated the effect of attention on the crossmodal integration of apparent motion signals using the crossmodal dynamic capture paradigm. The stimuli were bimodal apparent motion streams consisting of audio–visual, visual–tactile, or audio–tactile signals. The task was to indicate the direction of motion in one of the modalities, called the target modality, that could be congruent or incongruent with the direction of the second motion stream, called the distractor modality. The influence of the distractor modality on participants' response accuracy, called crossmodal capture, was assessed across three conditions of attentional manipulation. We found that attention does have an effect on how the motion signals are combined across modalities, but only when the susceptibility for capture between the two signals are comparable.

1. Introduction

Attention exerts a profound influence on many aspects of human perception and performance, ranging from the detection of simple sensory stimuli (e.g., Carrasco et al., 2000) to the implementation and control of complex behaviours like driving a vehicle (Underwood et al., 2003). Attention is also thought to play a critical role in the integration of different stimulus features *within* a sensory modality, such as vision (e.g., Treisman and Gelade, 1980), or in the processing of auditory sequences (e.g., Alain and Izenberg, 2003). Whether attention plays an important role in the binding of stimulus features *across* sensory modalities to achieve a unique coherent percept is, however, less clear. The present paper addresses this issue.

A prolific tradition in multisensory research has dealt with crossmodal attention phenomena, where the perception of targets in one sensory modality is measured as a function of cues in other senses. There is now overwhelming evidence for extensive influences between modalities (see Driver and Spence, 1998; Spence and Driver, 2004, for reviews). For example, visual events presented at a location previously occupied by a tactile, or an acoustic, cue are responded to more quickly and more

---

[*] *Corresponding author*. Fax: +1 604 875 4302.
E-mail address: ipek@psych.ubc.ca (I. Oruc).
URL: http://www.psych.ubc.ca/~ipek/ (I. Oruc).

doi:10.1016/j.brainres.2008.04.014

accurately than those presented away from the cue (e.g., Gray and Tan, 2002; Kennett et al., 2001; McDonald et al., 2000; Spence and Driver, 1994). ERP studies in humans reveal that this orienting-related enhancement affects (target) stimulus processing very shortly after its onset, at functionally early stages. These findings, together with the results of fMRI studies, have led some researchers to propose that crossmodal effects might be expressed as early as in primary sensory areas of the cortex (Macaluso et al., 2000; see also Kayser et al., in press for recent evidence in single neurons). This pattern of results is reminiscent of the large multisensory enhancements at the single neuron level observed in animal studies, most notably in the Superior Colliculus (SC) of the cat and monkey (e.g., Stein, 1998; Stein and Meredith, 1993). These enhancements in neural firing are commonly seen when two sensory events co-occur in time and overlap in space within the receptive field of the neuron. The SC is crucially involved in a network that embodies orienting behaviours, and which also includes posterior parietal and frontal pre-motor areas (e.g., Posner and Petersen, 1990). These advances in crossmodal attention run parallel to the also mounting demonstrations of multisensory enhancement. In this line of work, evidence points strongly to the existence of dramatic enhancement in the processing of events if they are presented multisensorily rather than unisensorily (e.g., Forster et al., 2002; Nickerson, 1973; Zampini et al., 2007). These enhancements have been registered not only in terms of more accurate and faster behavioural reactions, but also in terms of earlier and stronger neural correlates using electrophysiology and brain imaging techniques (Beauchamp et al., 2004a,b; Calvert et al., 2000; Foxe et al., 2000, 2002; Molholm et al., 2002; Murray et al., 2005; Senkowski et al., 2006, 2007).

Despite the recent advances in crossmodal attention (discussed before), and the clear existence of multisensory enhancement in sensory processing, there is still some debate about the nature and consequences of the interplay between attention and the integration of multisensory events (e.g., McDonald et al., 2001). Note, for example, that in most of the studies cited above, attention to one or more stimuli/modalities has been manipulated in crossmodal situations, but crossmodal binding has not been measured as a function of attention (though see Talsma and Woldorff, 2005). In fact, up until recently, investigations have generally assumed, either explicitly or implicitly, that attention plays little or no role in multisensory integration as long as other pre-conditions apply (e.g., De Gelder and Bertelson, 2003; Welch, 1999). A classic example is provided by research on the McGurk effect (McGurk and MacDonald, 1976), where vision is shown to have a profound impact on speech perception, e.g., one perceives “da” when seeing the mouth-movement for “ga” combined with the sound “ba”. Previous investigations of this audio–visual phenomenon have led to the general conclusion that it occurs automatically, with attention playing an insignificant role in the effect (Massaro, 1998; Soto-Faraco et al., 2004c,d). However, recent research suggests that claims for the automaticity of the McGurk effect might actually be incorrect. Alsius et al. (2005; see also Alsius et al., 2007) reported that the McGurk effect breaks down when attentional resources are depleted. The implication of this finding is that for complex crossmodal phenomena like the McGurk effect, attention may be needed to bind features *across* modalities. It is therefore not unreasonable to hypothesize that other complex crossmodal phenomena, such as dynamic capture or the ventriloquist effect, may also be modulated by attention (see also Talsma and Woldorff, 2005, for evidence supporting an effect of attention on audio–visual integration).

Studies of the ventriloquist effect, where an auditory event is mislocalized toward a displaced visual event, have generally suggested that attention plays no role in crossmodal interactions in this specific paradigm. Specifically, the focus of attention (e.g., on a single modality or divided between modalities) has been found to be irrelevant, with a simple “flash” of light consistently capturing (ventriloquizing) a simultaneous “beep” (Bertelson et al., 2000; Vroomen et al., 2001). Furthermore, the same conclusion is reached for other sensory modality combinations. For instance, Caclin et al. (2002) found that a vibration presented to a fingertip will capture a sound whether or not spatial attention is focused on touch.

Interestingly, crossmodal ventriloquism as described above, does not appear to be confined to situations involving a single light, sound, or vibration. In *dynamic* displays where audio and visual stimuli could move between two locations, Soto-Faraco et al. (2002) reported that the direction of auditory *motion* was captured by the direction of visual *motion*, i.e., the direction of sound motion was perceived to be the same as the direction of visual motion even when they were actually in opposition. And as with static ventriloquism, this dynamic ventriloquism effect can occur for other modality combinations, for instance, with visual motion capturing tactile motion, and auditory and tactile motion capable of capturing one another (Sanabria et al., 2005; Soto-Faraco et al., 2004a).

However, whether attention plays a role in dynamic ventriloquism is currently unknown. Past studies of static ventriloquism would suggest that attention does not play a significant role in dynamic ventriloquism (see e.g., Bertelson et al., 2000; Caclin et al., 2002; Vroomen et al., 2001). Such an extrapolation from results stemming from experimentation with static ventriloquism to dynamic ventriloquism should be undertaken with caution. In fact, researchers have been explicit in their stance that static ventriloquism and dynamic ventriloquism are distinct phenomena: only dynamic ventriloquism may depend on brain regions that are uniquely tuned to motion signals (Soto-Faraco et al., 2004c,d), only dynamic ventriloquism may be mediated by the corpus callosum (Soto-Faraco et al., 2002), and stronger crossmodal integration effects are demonstrated consistently under dynamic conditions (Soto-Faraco et al., 2004b).

We examined the role of attention in crossmodal dynamic capture by manipulating whether participants focused on a single sensory modality or had to commit their attention to both sensory modalities. Participants were required to report motion direction in a target modality (e.g., sound), and to ignore motion in another distractor modality (e.g., vision or touch). In particular, participants were assigned to one of three groups.

In the Blocked Group, the target modality and the distractor modality were known in advance and held constant across a block of trials. This blocked design is the standard approach in previous studies of dynamic ventriloquism, enabling one to measure how participants perform when selectively attending a target modality while ignoring a distractor modality. The typical outcome is that visual motion captures concurrent audio or tactile motion, but neither audio nor tactile motion exerts an influence on visual motion. Interestingly, a more symmetrical relation exists between audio–tactile motion, as audio and tactile motion can capture one another.

For a second set of participants, the Pre-cued Group, the modality pairings were held constant across a block of trials, but now the target modality was identified by a pre-cue at the start of each trial. Importantly, this pre-cue, and therefore the target and distractor modalities, varied randomly from trial to trial. Thus, like the Blocked Group, this Pre-cued Group always knew the target and distractor modalities before the motion stimuli were presented, but in contrast with the Blocked Group, now the attentional set varied from trial to trial. One would expect that this attentional switching would be more challenging (e.g., Meiran, 1996; Rogers and Monsell, 1995), but because the identity of the target modality was known in advance, as in the Blocked Group, the attentional set for the motion stimuli in the Pre-cued Group should approximate the attentional set in the Blocked Group.

For the third set of participants, the Post-cued Group, the modality pairings were again known in advance and held constant across a block of trials, but now the target modality was identified by a post-cue after each stimulus presentation. Importantly, this post-cue, and hence the target and distractor modalities varied randomly from trial to trial. What distinguishes the Post-cued Group from the Pre-cued Group, and by extension from the Blocked Group, is that in the Post-cued Group observers do not know which motion stream (target modality) they have to attend to and which to ignore (distractor modality) until *after* the motion stimuli have been presented. Therefore, unlike the other groups, in the Post-cued group both modality streams must be attended.

If, and how, this change in attentional set impacts performance on the dynamic ventriloquism task is the issue under investigation in the present study. If crossmodal interactions are unaffected by attention, then clearly the prediction is that the pattern of crossmodal interactions will be the same for all three groups. More specifically, observed levels of capture should not be affected by any differences in the attentional demands of the task. However, if attention to both modalities does play a significant role in crossmodal interactions, one might expect the integration of the motion signals into a single percept (motion stream) to be enhanced by committing attention to both signals. A conceptual parallel is that two different feature attributes are bound together into a common object when they are attended at the same time in visual search. In terms of crossmodal ventriloquism this would be revealed as an enhanced tendency to perceive motion streams that are physically presented in opposition as going in the same direction (e.g., a decline in response accuracy when the motion streams move in conflicting directions).

2. Results

The percent of correct responses for all three groups are presented in Table 1 as a function of the stimulus pairings and

Table 1 – Response accuracy for all three stimulus modality pairings (Audio–Visual, Visual–Tactile, Audio–Tactile) are shown as a function of congruency, synchrony, target modality, and group

Audio–Visual

	Blocked Group				Pre-cued Group				Post-cued Group			
% correct	Target: sound		Target: light		Target: sound		Target: light		Target: sound		Target: light	
	Con	Incon	Con	Incon	Con	Incon	Con	Incon	Con	Incon	Con	Incon
Synch	90	26.25	98.75	95	98.75	18.75	100	100	86.25	18.75	98.75	98.75
Asynch-a	85	86.25	98.75	98.75	91.25	95	98.75	93.75	72.5	81.25	95	100
Asynch-b	85	87.5	97.5	98.75	93.75	88.75	98.75	97.5	77.5	80	91.25	95

Visual–Tactile

	Blocked Group				Pre-cued Group				Post-cued Group			
% correct	Target: touch		Target: light		Target: touch		Target: light		Target: touch		Target: light	
	Con	Incon	Con	Incon	Con	Incon	Con	Incon	Con	Incon	Con	Incon
Synch	97.5	63.75	100	97.5	98.75	55	100	98.75	95	40	100	93.75
Asynch-a	93.75	95	97.5	100	98.75	96.25	97.5	96.25	81.25	92.5	97.5	97.5
Asynch-b	92.5	91.25	97.5	97.5	97.5	91.25	100	96.25	90	87.5	93.75	96.25

Audio–Tactile

	Blocked Group				Pre-cued Group				Post-cued Group			
% correct	Target: sound		Target: touch		Target: sound		Target: touch		Target: sound		Target: touch	
	Con	Incon	Con	Incon	Con	Incon	Con	Incon	Con	Incon	Con	Incon
Synch	92.5	71.25	97.5	77.5	98.75	55	100	80	95	50	93.75	51.25
Asynch-a	87.5	83.75	97.5	92.5	93.75	88.75	100	92.5	76.25	95	90	91.25
Asynch-b	86.25	86.25	96.25	95	91.25	91.25	98.75	95	86.25	90	93.75	93.75

target–distractor congruency. Below we report the analysis of variance (ANOVA) for each stimulus pairing (AV, AT, or VT) with Group (Blocked, Pre-cued, Post-cued) as a between-subjects factor and Target Modality (possible target modality depends on particular pairing being tested), Stimulus Congruency (Congruent, Incongruent), and Stimulus Synchrony (Synchronous, Asynchronous-before, and Asynchronous-after) as within-subjects factors.

2.1. *Audio-visual*

All main effects were significant (all $p<.05$), as was the 3-way interaction involving Target Modality × Congruency × Synchrony, $F(2, 54)=119.86$, $p<.0001$. Group did not interact significantly with any factor (all p's $>.05$), including a non-significant 4-way interaction, $F<1.2$, $p>.05$.

To determine the source of the Target Modality × Congruency × Synchrony interaction, we analyzed the data from each target modality separately. When participants responded to auditory motion (see Fig. 1A.), the Congruency × Synchrony interaction was significant, $F(2, 54)=157.70$, $p<.0001$. This interaction is due to the fact that a significant congruency effect appears only in the synchronous condition, with 91% correct responses to auditory motion in the congruent condition and 21% correct responses in the incongruent condition for a disparity of 70% ($p<.05$, Tukey–Kramer; all p's $>.05$ for the asynchronous conditions).

When the participants responded to the visual motion, the ANOVA yielded no significant effects, reflecting the fact that performance was excellent in all conditions, with the overall response accuracy to visual motion at 98%.

2.2. *Visual-tactile*

The percent correct responses for the visual–tactile pairing was analyzed as above, but now with Visual and Tactile as the potential target/distractor modalities. All main effects, except the main effect for Group, were significant (all p's $<.05$), as was the 3-way interaction involving Target Modality, Congruency, and Synchrony $F(2, 54)=40.03$, $p<.0001$. Group did not interact significantly with any factor (all p's $>.05$), with the 4-way interaction returning an $F<1.7$, $p>.05$.

To determine the source of the Target Modality × Congruency × Synchrony interaction, we analyzed the data from each target modality separately. When participants responded to tactile motion (see Fig. 1B), the Congruency × Synchrony interaction was significant, $F(2, 54)=56.97$, $p<.0001$. This interaction is due to a significant congruency effect in only the synchronous condition, with 97% correct responses to tactile motion in the congruent condition and only 53% correct in the incongruent

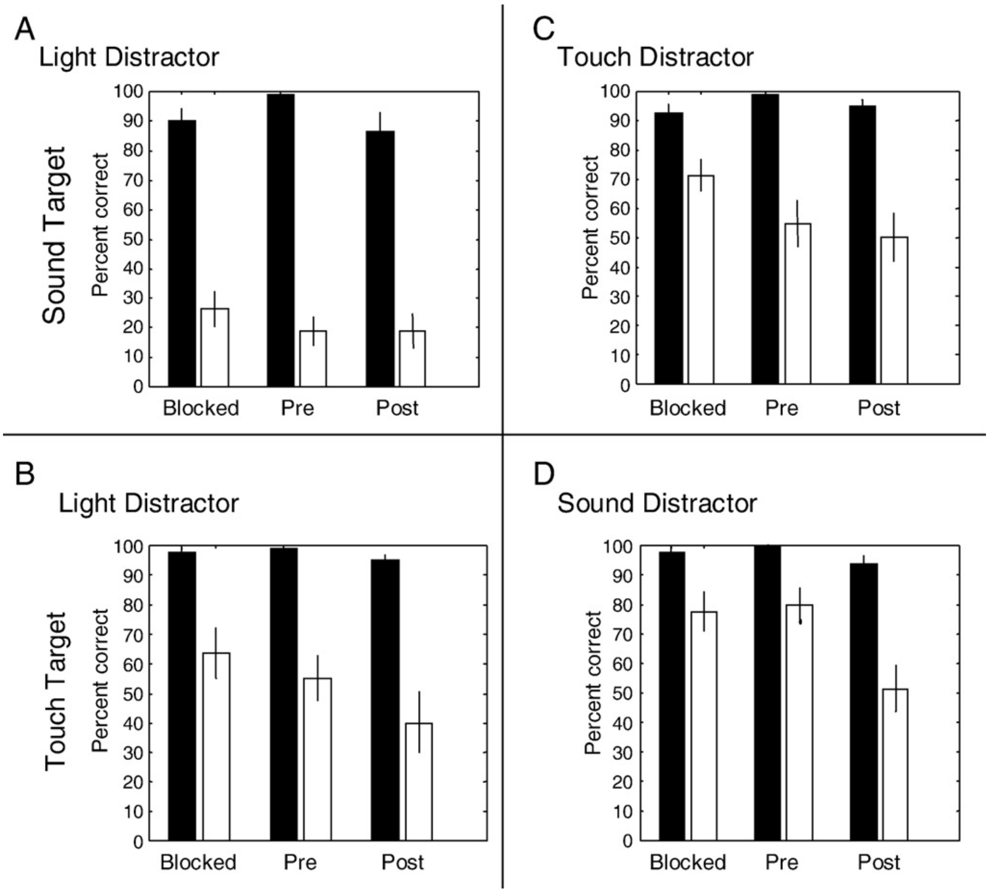

Fig. 1 – Response accuracy at the synchronous condition is plotted as a function of congruency (congruent in black, incongruent in white) for all three groups (Blocked, Pre-cue, Post-cue) for (A) Sound targets and Light distractors (B) Touch targets and Light distractors (C) Sound targets and Touch distractors, and (D) Touch targets and Sound distractors.

condition for a disparity of 44% ($p<.05$, Tukey–Kramer; all p's$>$.05 for the asynchronous conditions). Again, when the participants responded to the visual motion the overall accuracy was 98%.

2.3. *Audio-tactile*

The same ANOVA was performed as before but now with Audio and Tactile as potential target/distractor modalities. All main effects, except for Group, were significant (all $p<.05$), as was the 2-way interactions involving Congruency and Synchrony, $F(2,54)=70.43$, $p<.001$, and Group and Synchrony, $F(4,54)=3.01$, $p<.05$, and a 3-way interaction involving Target Modality, Congruency, and Synchrony, $F(2, 54)=3.44$, $p<.05$. For the first time we also recorded a highly significant interaction involving Group × Congruency × Synchrony, $F(4,54)=6.73$, $p<.001$. There were no other significant effects (all p's$>$.05).

To determine the cause of the Target Modality × Congruency × Synchrony interaction, we analyzed the data from each target modality separately. When participants responded to auditory motion (see Fig. 1C), the Congruency × Synchrony interaction was significant, $F(2, 54)=35.51$, $p<.0001$. This interaction is due to a significant congruency effect emerging only in the synchronous condition with response accuracy to auditory motion at 95% when in the congruent condition and 59% in the incongruent condition for a disparity of 36% ($p<.05$, Tukey–Kramer; all p's$>$.05 for the asynchronous conditions).

The analysis of the auditory data also revealed a significant Group × Congruency × Synchrony interaction, $F(4,54)=3.24$, $p<.05$. To determine the source of this interaction we analyzed the data separately for each synchrony condition. This revealed that for the synchronous condition, there was a significant Group × Congruency interaction, $F(2,27)=3.91$, $p<.05$, reflecting the fact that performance on auditory motion was only 50% in the incongruent post-cued condition, which was significantly worse than the blocked condition, 71% ($p<.05$, Tukey–Kramer). There were no significant differences between the incongruent pre-cued condition, and the incongruent blocked or post-cued conditions (p's$>$.05). The effect of Group was not significant in the congruent synchronous conditions, or in any of the asynchronous congruent, and asynchronous incongruent conditions (all p's$>$.05).

When participants responded to the tactile motion (see Fig. 1D), the Congruency × Synchrony interaction was significant, $F(2,54)=31.37$, $p<.001$. This was due to a significant congruency effect emerging for tactile motion only in the synchronous condition, with accuracy at 97% when tactile motion is congruent with sound motion and 70% when it is incongruent ($p<.05$, Tukey–Kramer; all other p's$>$.05).

This analysis also returned a significant Group × Congruency × Synchrony interaction, $F(4, 54)=4.40$, $p<.05$. To determine its source we analyzed the data separately for each synchrony condition. This revealed that for the synchronous condition there was a significant Group × Congruency interaction, $F(2,27)=4.21$, $p<.05$, reflecting that performance for tactile motion was only 51% in the incongruent post-cued condition, which was significantly worse than either the blocked or pre-cued conditions, 78% and 80% respectively ($p<.05$, Tukey–Kramer). There was no difference between blocked and pre-cued conditions, or between any of the target conditions in the congruent or asynchronous conditions (all p's$>$.05).

3. Discussion

The aim of the present study was to examine whether crossmodal dynamic capture can be modulated by the distribution of attention across sensory modalities. The Blocked Group results replicate the standard finding in the literature: when a target–distractor modality pairing is held constant throughout a block of trials, then visual motion captures synchronous audio or tactile motion, nothing captures visual motion, and synchronous audio motion and tactile motion can capture each other.

For the Pre-cued Group the modality pairings remained constant for a block of trials, but now which modality was the target and which was the distractor varied from trial to trial. Importantly, however, participants were pre-cued as to the target modality, and therefore in principle their attentional set should approximate the standard blocked condition, and by extension, the nature of the crossmodal interactions should be the same as what is typically found; and so they were. As before, visual motion captured audio and tactile motion, nothing captured vision, and both audio and tactile motion could capture each other. Most importantly, on no occasion was there any significant difference between the Blocked and Pre-cued Groups as a function of any crossmodal combination.

A different outcome, however, was observed for the Post-cued Group. In this case, participants could not narrow their attentional set to any one modality before hand, and had therefore to attend to both sensory modalities. While performance remained the same as before for crossmodal pairings involving visual motion, a significant change emerged from interactions involving audio–tactile integration. For these combinations, crossmodal dynamic capture was accentuated, and participants misperceived the direction of sound motion when it was paired with an opposing tactile motion; and similarly, participants misperceived the direction of tactile motion when it was paired with an opposing sound motion. The strong interpretation of these effects is that attention can play a significant role in the binding of stimulus features across modalities.

It is worth noting that one might question this interpretation and posit instead that response errors are just more likely to occur when a decision must be made about two possible target motion streams (post-cue condition) than one target motion stream (blocked and pre-cued conditions). However, the data do not support this position. First, this account makes the prediction that response errors would increase across the board, for all possible post-cue conditions, regardless of particular modality pairings. And secondly, it makes the prediction that response errors for all modality pairings should be increased for asynchronous and synchronous conditions alike. This was clearly not the case. Post-cued motion stream judgment errors were elevated selectively, only for the auditory–tactile modality pairings and only when they were presented synchronously, when crossmodal binding is optimal.

Calvert and Thesen (2004) suggest that for multisensory integration to take place, items from each modality must correspond spatially and be presented synchronously. These authors claim that if two or more stimuli occur at the same time and in the same spatial positions, they are inevitably bound into a single percept. This seems to be the case for our stimuli, except when vision is involved. That is, the current results demonstrate that the direction of the visual stream is consistently judged correctly, regardless of the presence of incongruent and synchronous auditory or tactile dynamic streams. However, judgments of the direction of auditory and tactile targets capture each other, with synchronous tactile and auditory dynamic streams being misperceived as congruent.

Why should visual motion be unaffected by incongruent crossmodal stimulation when auditory and tactile motion are affected? One possibility is that the acuity of the visual system for spatial stimuli is greater than the acuity of either the tactile or auditory sensory modalities. Accordingly, the visual system can successfully resolve the direction of visual motion easily and efficiently. In other words, the quality of motion information is better for the visual modality than the auditory and tactile modalities. As a result visual motion captures auditory and tactile motion. This has been referred to as the Modality Appropriateness Hypothesis, which states that the sensory modality that has the highest acuity for the stimulus being investigated dominates the percept (see Welch, 1999; Welch and Warren, 1980). Hence, the visual sensory system would dominate in our experiment, as the quality of information coming from the dynamic visual signal is greatest (Woods and Recanzone, 2004).

How attention is directed can have consequences on the amount of binding that occurs. In keeping with the above modality appropriateness hypothesis (Welch, 1999; Welch and Warren, 1980), our data indicate that whether vision is attended to or not, the quality of visual motion information is much greater than auditory or tactile motion. In other words, attending to, and enhancing, auditory or tactile motion information does not negate the quality advantage enjoyed by visual motion information. In contrast, when the motion information between modalities is comparable, as it is for auditory and touch, committing attention to both motion signals brings them into convergence. This is evidenced by an increase in capture in the post-cue condition for both modalities.

According to the traditional view of multisensory integration, information regarding motion in each sense is initially processed independently in modality-specific (or unimodal) brain areas, converging later at higher-order association areas. Visual motion processing has repeatedly been shown to involve visual area V5/MT1; auditory motion has been isolated to the planum temporale, the inferior and superior parietal cortices, and the right insula (Pavani et al., 2002); and tactile motion perception has been localized to the primary and secondary somatosensory areas located in the postcentral gyrus (e.g., Hagen et al., 2002). These motion signals may then converge in the intraparietal sulcus, as well as the precentral gyrus. When taken together, the evidence increasingly supports the idea that a network of brain areas is critically involved in processing motion information from more than one sensory modality. Whether the effects of attention on motion perception should be attributed to enhancement of signals at unimodal or multimodal brain areas is uncertain, and will unquestionably be the focus of future investigation.

Our findings converge with the recent discovery of Alsius et al. (2005; see also Alsius et al., 2007), suggesting that attention may have a much broader and more fundamental role to play in crossmodal interactions than previously thought. Indeed, attention has been found to modulate other processes under both unimodal and crossmodal presentations that were previously considered automatic, such as statistical learning (Toro et al., 2005), written and auditory word perception (Rees et al., 1999; Sinnett et al., 2006), or auditory stream segregation (Carlyon et al., 2001). Much of this research suggests that by depleting attentional reservoirs, a modulation of the effect can be observed (see for example, Lavie, 2005). In the present study, by requiring participants to monitor two streams, we ensured that both modalities were attended to, creating a condition in which completion of crossmodal integration is facilitated. Thus, in the case of the crossmodal dynamic capture, when auditory and visual motion streams are presented in opposition, the misperception of sound motion as a consequence of integration would be more likely. Given the wealth of studies that have failed to find a role of attention in crossmodal interactions, and the present investigation's discovery that attention does play a significant role, it will be important for future investigations to uncover what factors are critical to whether an attention effect is observed or not.

The current study adds to a relatively small, but growing, body of work demonstrating that attention can modulate crossmodal feature binding. We set out to answer whether attention has an impact on crossmodal dynamic ventriloquism. The data presented in this study support the idea that attention does indeed have an impact on crossmodal dynamic ventriloquism. It does not appear for all crossmodal combinations, but rather it appears to be specific to modality pairings that involve comparable levels of capture.

4. Experimental procedures

4.1. *Participants*

Thirty undergraduates (23 women; age = 19.9 ± 3.2; 26 right-handed) from the University of British Columbia participated in the experiment for course credit. All participants had normal hearing and normal or corrected-to-normal vision. The participants were randomly assigned to one of three groups: Blocked (8 women; age = 20.6 ± 4.8; 7 right-handed), Pre-cued (9 women; age = 19.9 ± 1.9; 10 right-handed), or Post-cued (6 women; age = 19.3 ± 2.4; 9 right-handed).

4.2. *Apparatus*

Two red Light Emitting Diodes (LEDs) were placed 50 cm from the participant and 15 cm away from either side of a centre fixation orange LED. All LEDs were fixed at a height of 5 cm above the table, each mounted on a loudspeaker cone. Two silent vibro-tactile stimulators (Sanko Electric #1E120, miniature enclosed vibrating motor) were attached to the left and right loudspeakers through an extended cable such that a vibrating motor was suspended 1 cm under each red LED, without actually touching the speaker cone. This was done to avoid any cross-talk between the auditory and tactile stimuli. Two foot-pedals on the floor were used to collect participants' responses (see Fig. 2). The LEDs and the foot-pedals were connected to the parallel port of the computer through a custom-made relay box. The loudspeakers and the vibro-tactile stimulators were

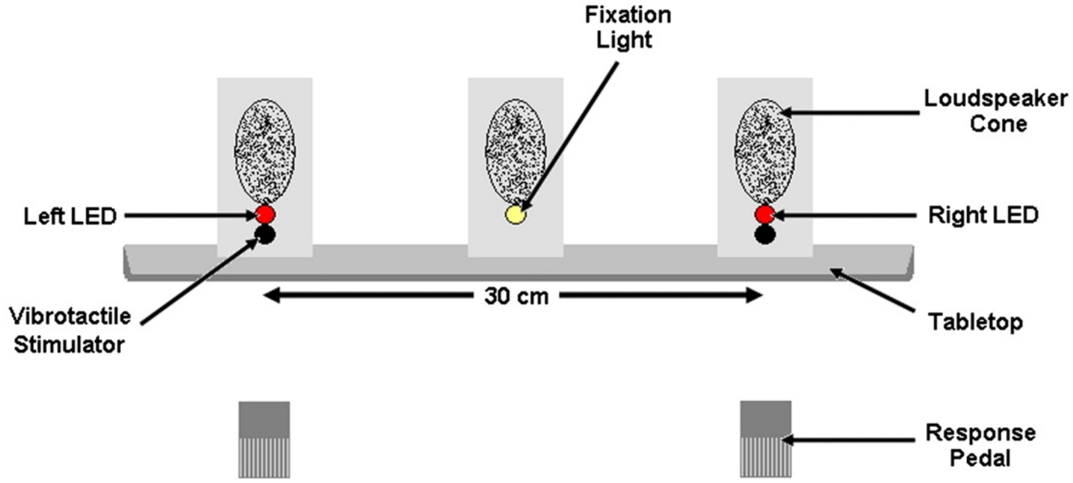

Fig. 2 – A schematic illustration of the experimental setup.

connected to the computer's multi-channel sound card (SB Audigy, 2ZS Audio, D400). The experiment was programmed using Presentation software (Neurobiological Systems).

Auditory apparent motion displays consisted of two 50-ms tones (57-dB, 500 Hz) separated by a 100-ms inter-stimulus-interval (ISI), each originating from a different loudspeaker. Visual apparent motion displays consisted of two 50-ms flashes separated by a 100-ms ISI, each presented on a red LED (luminous intensity = 500 mcd). Touch apparent motion consisted of two 50-ms vibrations (100 Hz) separated by a 100-ms ISI, each presented on a vibro-tactile stimulator. The vibration amplitudes were approximately 5.7 times the amplitudes at detection threshold.

4.3. *Procedure*

Participants were seated approximately 50 cm from the display in a dimly lit room. The participants positioned their hands such that their left and right index finger tips lightly touched the vibro-tactile stimulators from below. The participants were also instructed to rest their feet on the foot-pedals throughout the experiment.

Each trial began with a 1 s visual fixation stimulus (centre orange LED), followed by a 1 s delay before a pair of apparent motion displays was presented. The participant was then required to respond, indicating the direction of the apparent motion (left or right) by releasing the left or right foot-pedal. If the response was incorrect, the subject received error-feedback in the form of a 100-ms vibration on both vibro-tactile stimulators. On any given trial, apparent motion stimuli were presented in only two modalities: either sounds and lights, sounds and touches, or lights and touches.

A trial could be congruent or incongruent depending on whether the two motion streams moved in the same direction or in opposite directions. In addition, the two motion displays could be temporally synchronous or asynchronous. In synchronous trials, the two motion streams were presented concurrently (i.e., target–distractor SOA = 0). In asynchronous trials, there was a 500-ms lag between the onsets of the two apparent motion streams. When the target motion stream preceded the distractor motion stream, the trial was called asynchronous-before (i.e., target–distractor SOA = −500). Similarly, when the target motion stream followed the distractor stream, the trial was called asynchronous-after (i.e., target–distractor SOA = 500). The 2 types of congruency (congruent and incongruent) and the 3 types of synchrony (synchronous, asynchronous-before, and asynchronous-after) produced 6 different types of trials, each of which was presented an equal number of times in randomized order.

4.4. *Blocked group*

At the start of each block the participant was instructed to judge, throughout the block, the direction of motion of one target modality (either visual, auditory, or tactile) and ignore the other distractor modality. The participant was also informed that the two motion stimuli could move in the same or opposite directions in different trials, and that the direction of the distractor modality contained no information regarding the direction of the target modality. The participant was instructed to maximize response accuracy over response speed. Each block contained 48 trials consisting of 4 repetitions of the 6 types of trials described above, in 2 possible directions of motion (leftward or rightward). The target and distractor modalities remained the same within a block. Each participant completed 6 blocks, with three types of modality pairings (audio–visual, visual–tactile, audio–tactile) and two possible target modalities. The order of modality pairings was randomized across participants. As a result a total of 288 (48 × 6) trials were completed by each participant.

4.5. *Pre-cued group*

The procedure was the same as the Blocked Group with the exception that a *pre-cue* announced the target modality before the target and distractor stimuli were presented. This pre-cue was a female voice that spoke either the word “sound”, “light”, or “touch” (only two pre-cue words were possible within a block). A 500-ms delay separated the offset of the pre-cue and the onset of the motion stream. Within a block of 96 trials the pre-cue varied randomly between the two possible stimulus modalities. Thus one block in the Pre-cued condition (96 trials) was equivalent to two blocks (48×2) with the same modality pairing in the Blocked condition where the trials were randomly intermixed. The order of the blocks was randomized for each participant and blocks. Each participant completed 3 blocks resulting in a total of 288 (96×3) trials.

4.6. *Post-cued group*

The procedure was the same as the Pre-cued Group with the exception that the pre-cue was replaced by a *post-cue* that announced the target modality 500 ms following the offset of the motion stream.

REFERENCES

Alain, C., Izenberg, A., 2003. Effects of attentional load on auditory scene analysis. J. Cogn. Neurosci. 15 (7), 1063–1073.

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

Alsius, A., Navarra, J., Soto-Faraco, S., 2007. Attention to touch weakens audiovisual speech integration. Exp. Brain Res. 183 (3), 399–404.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004a. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neurosci. 7 (11), 1190–1192.

Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004b. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41 (5), 809–823.

Bertelson, P., Vroomen, J., de Gelder, B., Driver, J., 2000. The ventriloquist effect does not depend on the direction of deliberate visual attention. Percept. Psychophys. 62 (2), 321–332.

Caclin, A., Soto-Faraco, S., Kingstone, A., Spence, C., 2002. Tactile "capture" of audition. Percept. Psychophys. 64 (4), 616–630.

Calvert, G.A., Thesen, T., 2004. Multisensory integration: methodological approaches and emerging principles in the human brain. J. Physiol. (Paris) 98 (1–3), 191–205.

Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10 (11), 649–657.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H., 2001. Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. 27 (1), 115–127.

Carrasco, M., Penpeci-Talgar, C., Eckstein, M., 2000. Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. Vis. Res. 40 (10–12), 1203–1215.

De Gelder, B., Bertelson, P., 2003. Multisensory integration, perception and ecological validity. Trends Cogn. Sci. 7 (10), 460–467.

Driver, J., Spence, C., 1998. Attention and the crossmodal construction of space. Trends Cogn. Sci. 2 (7), 254–262.

Forster, B., Cavina-Pratesi, C., Aglioti, S.M., Berlucchi, G., 2002. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. Exp. Brain Res. 143 (4), 480–487.

Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., Schroeder, C.E., 2000. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res. Cogn. Brain Res. 10 (1–2), 77–83.

Foxe, J.J., Wylie, G.R., Martinez, A., Schroeder, C.E., Javitt, D.C., Guilfoyle, D., et al., 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J. Neurophysiol. 88 (1), 540–543.

Gray, R., Tan, H.Z., 2002. Dynamic and predictive links between touch and vision. Exp. Brain Res. 145 (1), 50–55.

Hagen, M.C., Franzen, O., McGlone, F., Essick, G., Dancer, C., Pardo, J.V., 2002. Tactile motion activates the human middle temporal/V5 (MT/V5) complex. Eur. J. Neurosci. 16 (5), 957–964.

Kayser, C., Petkov, C.I., Logothetis, N.K., in press. Visual modulation of neurons in auditory cortex. Cereb. Cortex. (Advance access published on January 6, 2008). doi:10.1093/cercor/bhm187.

Kennett, S., Taylor-Clarke, M., Haggard, P., 2001. Noninformative vision improves the spatial resolution of touch in humans. Curr. Biol. 11 (15), 1188–1191.

Lavie, N., 2005. Distracted and confused? selective attention under load. Trends Cogn. Sci. 9 (2), 75–82.

Macaluso, E., Frith, C.D., Driver, J., 2000. Modulation of human visual cortex by crossmodal spatial attention. Science 289 (5482), 1206–1208.

Massaro, D.W., 1998. Perceiving Talking Faces: From Speech Perception to a Behavioral Principle. MIT Press, Cambridge, MA.

McDonald, J.J., Teder-Salejarvi, W.A., Hillyard, S.A., 2000. Involuntary orienting to sound improves visual perception. Nature 407 (6806), 906–908.

McDonald, J.J., Teder-Salejarvi, W.A., Ward, L.M., 2001. Multisensory integration and crossmodal attention effects in the human brain. Science 292 (5523), 1791.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264 (5588), 746–748.

Meiran, N., 1996. Reconfiguration of processing mode prior to task performance. J. Exp. Psychol., Learn., Mem., Cogn. 22 (6), 1423–1442.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res. Cogn. Brain Res. 14 (1), 115–128.

Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., et al., 2005. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb. Cortex 15 (7), 963–974.

Nickerson, R.S., 1973. Intersensory facilitation of reaction time: energy summation or preparation enhancement? Psychol. Rev. 80 (6), 489–509.

Pavani, F., Macaluso, E., Warren, J.D., Driver, J., Griffiths, T.D., 2002. A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr. Biol. 12 (18), 1584–1590.

Posner, M.I., Petersen, S.E., 1990. The attention system of the human brain. Ann. Rev. Neurosci. 13, 25–42.

Rees, G., Russell, C., Frith, C.D., Driver, J., 1999. Inattentional blindness versus inattentional amnesia for fixated but ignored words. Science 286 (5449), 2504–2507.

Rogers, R.D., Monsell, S., 1995. Costs of a predictable switch between simple cognitive tasks. J. Exp. Psychol. Gen. 124 (2), 207–231.

Sanabria, D., Soto-Faraco, S., Spence, C., 2005. Assessing the effect of visual and tactile distractors on the perception of auditory apparent motion. Exp. Brain Res. 166 (3–4), 548–558.

Senkowski, D., Molholm, S., Gomez-Ramirez, M., Foxe, J.J., 2006. Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: a high-density electrical mapping study. Cereb. Cortex 16 (11), 1556–1565.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C.S., Woldorff, M.G., 2007. Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45 (3), 561–571.

Sinnett, S., Costa, A., Soto-Faraco, S., 2006. Manipulating inattentional blindness within and across sensory modalities. Q. J. Exp. Psychol. (Colchester) 59 (8), 1425–1442.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., Kingstone, A., 2002. The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. Brain Res. Cogn. Brain Res. 14 (1), 139–146.

Soto-Faraco, S., Spence, C., Kingstone, A., 2004a. Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture. Cogn. Affect. Behav. Neurosci. 4 (2), 208–217.

Soto-Faraco, S., Spence, C., Kingstone, A., 2004b. Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. J. Exp. Psychol. Hum. Percept. Perform. 30 (2), 330–345.

Soto-Faraco, S., Navarra, J., Alsius, A., 2004c. Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. Cognition 92 (3), B13–B23.

Soto-Faraco, S., Spence, C., Lloyd, D., Kingstone, A., 2004d. Moving multisensory research along: motion perception across sensory modalities. Curr. Dir. Psychol. Sci. 13 (1), 29–32.

Spence, C., Driver, J., 1994. Covert spatial orienting in audition— exogenous and endogenous mechanisms. J. Exp. Psychol. Hum. Percept. Perform. 20 (3), 555–574.

Spence, C., Driver, J., 2004. Crossmodal Space and Crossmodal Attention. Oxford University Press, Oxford, UK.

Stein, B., 1998. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. Exp. Brain Res. 123 (1–2), 124–135.

Stein, B., Meredith, M.A., 1993. The Merging of the Senses. MIT Press, Cambridge, MA.

Talsma, D., Woldorff, M.G., 2005. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J. Cogn. Neurosci. 17 (7), 1098–1114.

Toro, J.M., Sinnett, S., Soto-Faraco, S., 2005. Speech segmentation by statistical learning depends on attention. Cognition 97 (2), B25–B34.

Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. Cogn. Psychol. 12 (1), 97–136.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., Crundall, D., 2003. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics 46 (6), 629–646.

Vroomen, J., Bertelson, P., de Gelder, B., 2001. The ventriloquist effect does not depend on the direction of automatic visual attention. Percept. Psychophys. 63 (4), 651–659.

Welch, R.B., 1999. Meaning, attention, and the “unity assuption” in the intersensory bias of spatial and temporal perceptions In: Ashersleben, G., Bachmann, T., Müsseler, J. (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events. Elsevier, Amsterdam, pp. 371–387.

Welch, R.B., Warren, D.H., 1980. Immediate perceptual response to intersensory discrepancy. Psychol. Bull. 88 (3), 638–667.

Woods, T., Recanzone, G., 2004. Cross-modal interactions evidenced by the ventrioloquism effect in humans and monkeys. In: Calvert, G.A., Spence, C., Stein, B.E. (Eds.), The Handbook of Multisensory Processes. MIT Press, Cambridge, MA, pp. 35–48.

Zampini, M., Torresan, D., Spence, C., Murray, M.M., 2007. Auditory-somatosensory multisensory interactions in front and rear space. Neuropsychologia 45 (8), 1869–1877.