

Direction of visual apparent motion driven by perceptual organization of cross-modal signals

Warrick Roseboom

NTT Communication Science Laboratories,
Nippon Telegraph and Telephone Corporation, Atsugi,
Kanagawa, Japan



Takahiro Kawabe

NTT Communication Science Laboratories,
Nippon Telegraph and Telephone Corporation, Atsugi,
Kanagawa, Japan



Shin'ya Nishida

NTT Communication Science Laboratories,
Nippon Telegraph and Telephone Corporation, Atsugi,
Kanagawa, Japan



A critical function of the human brain is to determine the relationship between sensory signals. In the case of signals originating from different sensory modalities, such as audition and vision, several processes have been proposed that may facilitate perception of correspondence between two signals despite any temporal discrepancies in physical or neural transmission. One proposal, temporal ventriloquism, suggests that audio-visual temporal discrepancies can be resolved with a capture of visual event timing by that of nearby auditory events. Such an account implies a fundamental change in the timing representations of the involved events. Here we examine if such changes are necessary to account for a recently demonstrated effect, the modulation of visual apparent motion direction by audition. By contrast, we propose that the effect is driven by segmentation of the visual sequence on the basis of perceptual organization in the cross-modal sequence. Using different sequences of cross-modal (auditory and tactile) events, we found that the direction of visual apparent motion was not consistent with a temporal capture explanation. Rather, reports of visual apparent motion direction were dictated by perceptual organization within cross-modal sequences, determined on the basis of apparent relatedness. This result adds to the growing literature indicating the importance of apparent relatedness and sequence segmentation in apparent timing. Moreover, it demonstrates that, contrary to previous findings, cross-modal interaction can play a critical role in determining organization of signals within a single sensory modality.

Keywords: temporal ventriloquism, apparent motion, multisensory timing, cross-modal timing, perceptual grouping, perceptual organization, perceptual segmentation, cross-modal hierarchy

Citation: Roseboom, W., Kawabe, T., & Nishida, S. (2013). Direction of visual apparent motion driven by perceptual organization of cross-modal signals. *Journal of Vision*, 13(1):6, 1-13, <http://www.journalofvision.org/content/13/1/6>, doi:10.1167/13.1.6.

Introduction

Determining the relationship between different sensory signals is a critical function of the human brain. Many factors contribute to the perception that multiple signals, such as audition and vision, share a common source, but the most important cues appear to be spatial and temporal correspondence (Calvert, Spence, & Stein, 2004; Slutsky & Recanzone, 2001; Stein & Meredith, 1993). However, determining temporal correspondence is made difficult by differences in intrinsic and extrinsic transduction speeds for different sensory signals (see King, 2005). Several proposed processes may facilitate the combination of different sensory signals despite

temporal discrepancy between them (e.g., Temporal Recalibration; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Roseboom & Arnold, 2011; Vroomen, Keetels, de Gelder, & Bertelson, 2004; AV simultaneity window; Dixon & Spitz, 1980; Miner & Caudell, 1998; Navarra et al., 2005; Powers, Hillock, & Wallace, 2009; Vatakis & Spence, 2006). Of particular interest here is *temporal ventriloquism* (Fendrich & Corballis, 2001; Keetels & Vroomen, 2011a; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Parise & Spence, 2009; Scheier, Nijhawan, & Shimojo, 1999; Vroomen & Keetels, 2006). Temporal ventriloquism is often characterized as the capture of the apparent timing of one sensory signal by another. Under this proposal, temporal discrepancy between different sensory signals is resolved simply by diminish-

ing the apparent temporal distance between them. However, it is unclear how such a process might be implemented in the brain.

Typical demonstrations of temporal ventriloquism rely on the detection of small differences in the precision of temporal order judgments (TOJs) between successive visual presentations when presented with temporally adjacent, as compared with temporally synchronous, auditory events (Keetels & Vroomen, 2011a; Morein-Zamir et al., 2003; Parise & Spence, 2009; Scheier et al., 1999; Vroomen & Keetels, 2006). Recently, Freeman and Driver (2008) provided a much more compelling demonstration apparently supporting the existence of temporal ventriloquism. This study utilized a visual phenomenon wherein the repeated presentation of two successive, spatially offset, visual events can result in the appearance of directional motion from one location to the other. When the timing of the repetitive cycle is such that, for example, the duration between successive left-to-right presentations is shorter than that of the subsequent right-to-left portion of the cycle, a strong impression of rightward cycling motion is perceived (Movies 1 and 2; Dawson, 1991; Gepstein & Kubovy, 2007 for recent overview; Kolers, 1972; Ullman, 1979). If the duration between successive left-to-right-to-left presentations is symmetrical, the stimulus simply appears to switch between locations – a directionally ambiguous display. Freeman and Driver (2008) posited that the capture of visual event timing by temporally offset auditory events could induce an apparent temporal asymmetry in a directionally ambiguous visual display. Their results supported this proposal. For example, when the repetitive sequence consisted of a visual presentation on the left, accompanied by a temporally *trailing* auditory event, followed by a visual presentation on the right, accompanied by a temporally *leading* auditory event, participants reported a strong perception of left-to-right, *rightward*, motion (see Figure 1, Movie 3).

A critical question that remains for temporal ventriloquism, and timing phenomena generally, is whether changes in subjective timing perception necessitate changes in the corresponding neural event time course (e.g., Di Luca, Machulla, & Ernst, 2009; Keetels & Vroomen, 2011b; Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009). Temporal capture represents a strong account of temporal ventriloquism effects and suggests that the presence of an auditory event fundamentally alters the timing representation of a visual event (e.g., Keetels & Vroomen, 2011b; Vroomen & Keetels, 2009). Alternatively, subjective changes may reflect only a differing organization or interpretation of signals in relation to one another, with the apparent time course only marginally related to physical or neural event timing (Dennett & Kinsbourne, 1992; Fujisaki, Kitazawa, & Nishida, 2012; Johnston & Nishida, 2001).

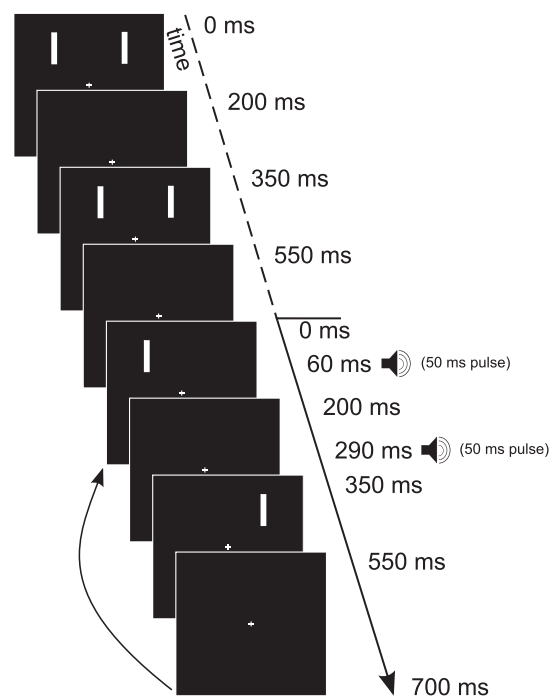


Figure 1. Example depiction of the stimulus sequence for a single trial presentation. Each trial began with presentations of the visual stimulus in both left and right positions. The visual stimulus subsequently cycled from side to side, left to right in example, eight times. The cross-modal flankers, here indicated by speaker icons, were presented offset from the visual stimulus presentation by 60 ms. Visual events were 200 ms in duration while cross-modal events were 50 ms in duration. The example configuration would result in a perception of visual motion cycling in a rightward direction.

An alternative explanation for temporal ventriloquism involves the grouping of apparently related auditory and visual sequences. Several recent studies have demonstrated the influence that apparent relatedness of within and across modality events can have on perceived timing. For example, it has been demonstrated that if a sequence of auditory events is alike, they tend to group together, strongly mitigating the likelihood of any auditory event in that sequence combining with signals from another modality (Cook & Van Valkenburg, 2009; Keetels, Stekelenburg, & Vroomen, 2007; Klink, Montijn, & van Wezel, 2011). These results provide evidence for the existence of a processing hierarchy wherein grouping is resolved within a modality prior to the possibility of any cross-modal combination. With regards to cross-modal grouping, factors such as content, semantic or spatial relation (e.g., Keetels & Vroomen, 2005; Parise & Spence, 2009; Vatakis & Spence, 2006; Zampini, Shore, & Spence, 2003; Zampini, Guest, Shore, & Spence, 2005) have been shown to contribute critically to the accuracy with which temporal judgments are made.

In this study we propose that the effect reported by Freeman and Driver (2008) is not based on temporal capture. Instead, we suggest that the effect may be driven by the grouping of successive auditory events that subsequently determine the grouping of visual events to imply directional motion. The proposed account differs significantly from the above mentioned demonstrations of the effect of grouping on subjective timing. Critically, we believe that grouping as determined by temporal proximity and feature similarity within one modality (e.g., auditory) may explicitly dictate the grouping within another (vision). This account implies that perceptual organization of a sequence in one modality may determine the specific arrangement of a sequence in another, contrary to previous indications of the processing hierarchy of multisensory signals (e.g., Cook & Van Valkenburg, 2009; Keetels et al., 2007; Klink et al., 2011; see Spence & Chen, 2012 for review).

To contrast these different hypotheses, we used a visual apparent motion display similar to that used by Freeman and Driver (2008). We examined the influence of different configurations of cross-modal cues (auditory or tactile) on perceived visual apparent motion. A temporal capture account predicts that any combination of cross-modal cues that can, individually, capture the timing of visual presentations should effectively modulate the direction of visual apparent motion. Consequently, perceived direction of visual motion should always be consistent with that implied by the *timing* of cross-modal relative to visual events. To preview the results, we find that this is not that case. When using different combinations and temporal configurations of auditory and tactile cross-modal cues, we find that a temporal capture explanation cannot account for the pattern of results. We believe the results are consistent with an explanation based on the grouping of auditory or tactile cues on the basis of similarity and temporal proximity. Once formed, the cross-modal groups dictate the grouping of visual events to determine perceived apparent motion direction. This account does not necessitate any changes in timing of event representations, but relies on the ability to flexibly change the apparent organization of a visual sequence based on the organization of cues presented in other modalities.

Materials and methods

Participants and apparatus

Participants included two of the authors (WR and TK) and six participants who were naïve as to the experimental purpose. All reported normal or corrected

to normal vision and hearing. Naïve participants received ¥1000 per hour for their participation. Ethical approval for this study was obtained from the ethical committee at Nippon Telegraph and Telephone Corporation (NTT Communication Science Laboratories Ethical Committee). The experiments were conducted according to the principles laid down in the Helsinki Declaration. Written informed consent was obtained from all participants except the authors.

Visual stimuli were generated using a VSG 2/3 from Cambridge Research Systems (CRS) and displayed on a 21" Sony Trinitron GDM-F520 monitor (resolution of 800×600 pixels and refresh rate of 100 Hz). Participants viewed stimuli from a distance of ~ 105 cm. Auditory signals were presented via a loudspeaker at a distance of ~ 60 cm, while tactile signals were presented via a vibration generator (EMIC Corp.) placed at a distance of ~ 50 cm from the participant. Participants placed their right arm on a cushioned arm-rest and rested their finger on the vibration generator. Audio and tactile stimulus presentations were controlled by a TDT RM1 Mobile Processor (Tucker-Davis Technologies). Auditory presentation timing was driven via a digital line from a VSG Break-out box (CRS), connected to the VSG, which triggered the RM1. Participants responded with their left hand using the left and right cursor keys of the keyboard.

Basic stimuli

The visual stimuli for all experimental phases consisted of white (CIE 1931 $x = 0.297$, $y = 0.321$, 123 cd/m^2) bars ($0.25 \times 1.55 \text{ dva}$) presented 3.35 dva either side, and 2 dva above a white central fixation point (0.25 dva in width and height), against a black ($\sim 0 \text{ cd/m}^2$) background (see Figure 1, for a graphic depiction). Individual visual stimulus presentations were 200 ms in duration. Broadband auditory noise was presented continuously throughout the experiment at $\sim 80 \text{ db SPL}$. Auditory signals consisted of a 50 ms pulse, containing 1 ms cosine onset and offset ramps of either a transient amplitude increase in the broadband noise ($\sim 85 \text{ db SPL}$) or a 1500 Hz sine-wave carrier (Pure tone stimulus). Tactile signals consisted of a 50 ms pulse, containing 1 ms cosine onset and offset ramps, of a 20 Hz sine-wave carrier. Participants wore Sennheiser HDA200 headphones for passive noise cancelling. The passive headphones, combined with high intensity background noise, were used in order to mask any audible noise produced by the tactile stimulator.

Each trial (Movies 3–8; Figure 1) began with presentation of the bars both left and right of fixation between three and six times, determined on a trial-by-trial basis. The visual inter stimulus onset asynchrony

(vSOA) was 350 ms. The purpose of this sequence was to minimize participants' bias to report direction of motion based on starting position. Following offset of the final presentation in this initial sequence, there was a 150 ms pause before the main stimulus sequence cycle. On 50% of trials, the main stimulus cycle began with presentation of the stimulus to the left of fixation followed, after the vSOA, by presentation of the visual stimulus to the right. On 50% of trials this was reversed. A single stimulus cycle took 700 ms in total and on each trial eight consecutive stimulus cycles were presented. As the vSOA was symmetrical (350 ms), the direction of motion in this display was ambiguous. Auditory or tactile events could be presented either synchronously with visual stimulus timing (1/3 of trials), or offset from the visual stimuli by 60 ms (2/3 of trials). Thus, while the vSOA was always 350 ms, the SOA between successive cross-modal flankers (fSOA) would on some trials match this (350 ms) or on other trials, for a given stimulus cycle, be shorter ($350 - 120 = 230$ ms), or longer ($350 + 120 = 470$ ms). Previous studies (Freeman & Driver, 2008; Kafaligonul & Stoner, 2010) have indicated that if the cross-modal flanker events both fall between successive left-then-right visual events, the display implies rightward apparent motion. Conversely, if the flanker events fall between right-then-left visual events, the display implies leftward apparent motion. And so, on 1/3 of trials the cross-modal flanker timing implied rightward motion, while on 1/3 of trials cross-modal timing implied leftward motion. On the remaining 1/3 of trials there was no timing discrepancy between visual events and cross-modal events. On the basis of relative timing cues alone, the direction of motion for these presentations would remain ambiguous. The timing of the first presented cross-modal flanker led the first visual event, and the second flanker lagged the second visual event, on 50% of trials, while this relationship was reversed on the remaining 50% of trials.

Cross-modal flanker configurations and experimental procedures

There were six different cross-modal flanker conditions; audio only (Movies 3 & 4; Figure 2; see Supplementary Materials for depictions of each movie timeline), tactile only, audio-tactile sequential (Figure 2), audio-tactile pairwise (Figure 2), noise-pure tone sequential (Movie 5) and noise-pure tone pairwise (Movies 6 - 8). In the audio only condition (see Figure 2), all cross-modal events were broadband auditory noise. This condition was effectively a replication of previous studies (Freeman & Driver, 2008; Kafaligonul & Stoner, 2010). In the tactile only condition, all cross-modal events were tactile stimuli. In the sequential

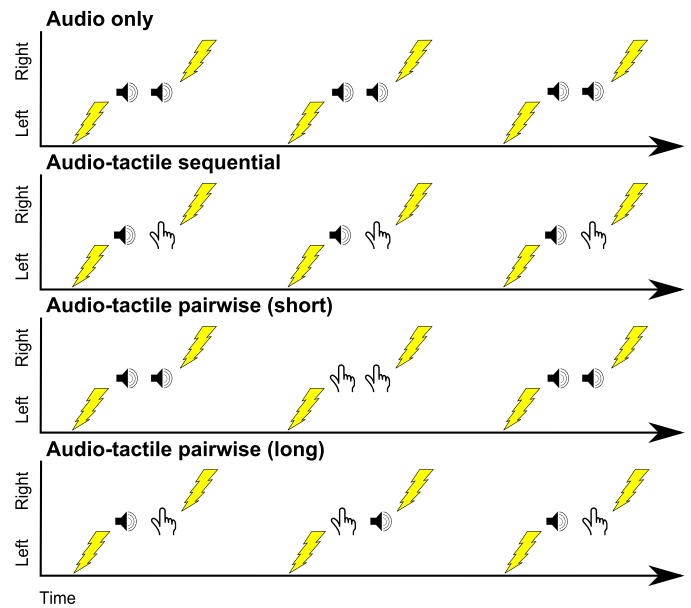


Figure 2. Schematic depiction of example timeline for different experimental conditions. Audio only condition contains only a single flanker signal type (tactile only condition is equivalent but contains only tactile signals). Audio-tactile sequential condition sequentially alternates between auditory and tactile signals (noise-pure tone condition is equivalent but contains pure-tone and broadband auditory noise signals). Audio-tactile pairwise conditions (short and long) alternate between auditory and tactile signals in a pairwise manner (noise-pure tone pairwise condition is equivalent but contains pure-tone and broadband auditory noise signals). See section *The importance of grouping between successive flanker events* and Figure 4 for the rationale behind separating pairwise conditions into “short” and “long” sets.

conditions, for both audio-tactile (see Figure 2) and noise-pure tone (Movie 5) combinations, the cross-modal events sequentially alternated between the different stimuli (e.g., A.T...A.T...; periods represent the passage of time). In the pairwise conditions, for both audio-tactile (see Figure 2) and noise-pure tone (Movies 6 - 8) combinations, the stimuli were sequentially alternated in a pairwise manner (e.g., A.A...T.T...). Participants completed each condition in a separate block of trials. Within a given block of trials, each possible configuration for that condition (e.g., a given trial starting with an auditory or tactile signal in the audio-tactile sequential condition) was presented with the same frequency. The order of presentation for the different configurations was determined on a trial-by-trial basis.

Each block consisted of 72 trials. Participants completed one block of trials for each of the conditions, except each of the pairwise conditions, for which they completed two blocks, eight blocks in total. Half of the participants completed the audio only condition first while the other half completed the tactile only

condition first. The remaining conditions were completed in a pseudo-random order. Each block of trials took approximately 20 min to complete. Participants were required to wait until the completion of the stimulus presentation on each trial before responding. The task was a binary forced-choice response; was the direction of visual motion left or right?

Prior to the experimental sessions, participants were shown the basic visual stimulus, without any accompanying cross-modal signals (Movies 1 and 2). The presentations could have either symmetrical vSOA, as described above, or asymmetrical vSOAs such that directional apparent motion would be induced. This was done to give participants an opportunity to observe a stimulus in which there physically was unambiguous directional motion. When successive vSOAs were asymmetrical, participants were able to identify in which direction the stimulus appeared to cycle.

Can simple cross-modal temporal capture drive visual apparent motion?

A simple temporal capture account would predict that, regardless of the sequence of different cross-modal events (see Figure 2), stimulus sequences with identical temporal profiles should elicit similar direction of visual apparent motion reports. Contrary to this account, an explanation based on the grouping of cross-modal events by temporal proximity and similarity would predict that sequences in which the type of cross-modal event is varied (e.g., audio-tactile sequential and pairwise sequences; see Figure 2) should exhibit strongly mitigated direction of motion effects. This prediction is made because it is possible for the temporal and similarity grouping cues to indicate opposite sequence configurations and therefore opposite directions of visual apparent motion (see Figure 6 and the Within-modal grouping can be dictated by cross-modal organization section of the General discussion for a complete explanation of the grouping hypothesis). Under these conditions the impression of directional visual apparent motion should become more ambiguous.

To examine the influence of the relative timing of cross-modal events on the visual display, we plotted the proportion of participants' reports that the visual stimulus appeared to cycle in a rightward direction as a function of whether the flanker timing implied rightward or leftward motion (Figure 3A). From this we were able to take a measure of the magnitude of influence that the timing of cross-modal events had on the perception of visual apparent motion. This was done by subtracting the proportion of "rightward" responses when the implied direction was leftward, from those when the implied direction was rightward

(Figure 3B). Here, values that approach one indicate that the perceived direction of visual apparent motion was consistent with being dictated by temporal capture of visual events by corresponding cross-modal flankers (negative 1 would imply that perceived motion direction is absolutely *inconsistent*).

As can be seen in Figure 3, the magnitude of cross-modal modulation was strong for the audio (AUD) and tactile (TAC) only conditions, consistent with previous studies (Freeman & Driver, 2008; Kafaligonul & Stoner, 2010). However, compared to the audio only condition, the influence of auditory and tactile flankers was significantly reduced for the conditions in which the flanker sequence was alternated, both sequentially (ATS; audio only = 0.839; audio-tactile sequential = 0.583, $t_7 = 6.85$, $p < 0.001$; paired samples-two tailed) and pairwise (ATP; audio only = 0.839; audio-tactile pairwise = 0.526, $t_7 = 7.27$, $p < 0.001$; paired samples-two tailed).

Because both the audio and tactile only conditions demonstrated a strong effect individually, the temporally offset auditory or tactile flanker should be able to adequately capture the timing of the nearby visual event regardless of the configuration. Consequently, a temporal capture account would predict that there is no difference between any of the four examined conditions. However, these results clearly indicate that the relationship between successive cross-modal flankers is a critical factor in the perceived direction of visual motion. The role of this relationship can be further examined through the conditions in which cross-modal flankers were alternated in a pairwise manner.

The importance of grouping between successive flanker events

The presented results raise significant doubt as to the possibility of a temporal capture account. An alternative account is that direction of visual motion is dependent on the apparent grouping of successive cross-modal flankers. The audio-tactile pairwise condition can be considered to have two cues that may contribute to the apparent grouping of flankers; the temporal distance between successive flankers, and whether successive flanker events are the same or different signals (similarity cue). On the basis of these two factors, the pairwise flanker sequences can be considered to be *short* or *long* pairs. For example, a pairwise flanker sequence may proceed as either A.A...T.T...A.A... (short) or as A.T...T.A...A.T... (long); see Figure 2; see also Figure 6 for a similar timeline depiction. When the flanker pairs are short, the timing and similarity cues can be considered to imply the same direction of motion. Moreover, a short cross-modal event pair implies the same direction of motion as implied by the simple temporal capture account.

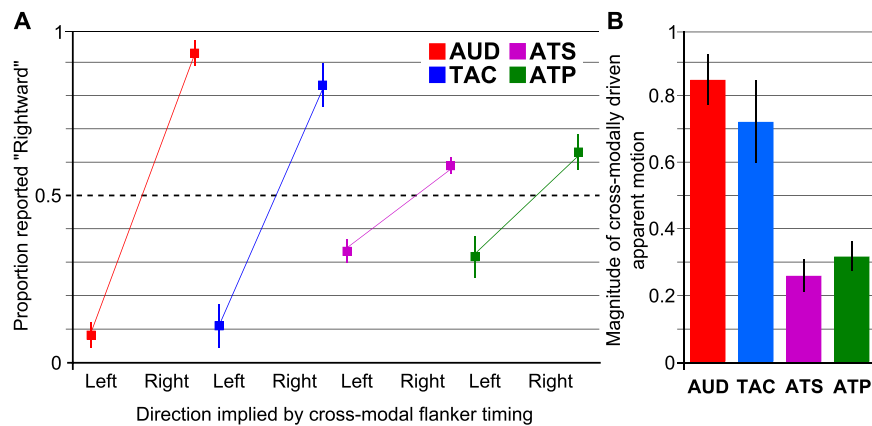


Figure 3. Graphical depictions of experimental data. (A) Averaged proportion of trials in which eight participants reported the perceived direction of motion as rightward when the timing of cross-modal flankers implied leftward or rightward motion; for the audio only (AUD), tactile only (TAC), audio-tactile sequential alternation (ATS), and audio-tactile pairwise alternation conditions (ATP). (B) Averaged estimate of cross-modal influence on visual apparent motion reports for eight participants for audio-tactile cross-modal flanker conditions. This measure was generated by subtracting the proportion of “rightward” responses when implied direction of motion was leftward, from that when it was rightward (see Figure 3A). A value of 1 would indicate that participants’ reports were perfectly consistent with cross-modal capture of visual event timing. Error bars indicate \pm standard error of the *mean*.

However, when the flanker pairs are long, the timing and similarity cues are in conflict with each other. Therefore, if the grouping of flankers on the basis of similarity is critical to the cross-modal modulation of visual apparent motion direction, the magnitude of this effect should be significantly reduced in the long compared to short pair condition.

As can be seen in Figure 4, when trial data is separated by the cross-modal flanker pair being *short* or *long* (see Figure 2), the short pair configuration for an audio-tactile pairwise sequence (ATP_S) shows strong modulation of perceived direction consistent with being driven by cross-modal flanker timing. However, for the long pair configuration, this result was not found. Here the perceived direction of visual motion was significantly different from what was found for the corresponding *short* condition (audio-tactile pairwise short = 0.745; audio-tactile pairwise long = -0.120; $t_7 = 4.59$, $p = 0.003$). This result indicates that the grouping of successive cross-modal flankers on the basis of their similarity is critical in driving the perceived direction of visual apparent motion.

Visual apparent motion driven by flanker grouping alone

The results thus far suggest that the grouping of cross-modal flanker pairs on the basis of similarity may drive the directional motion percept at least as strongly as the grouping induced by temporal proximity. However, no result explicitly reflects the influence of the similarity grouping cue by itself. To confirm that grouping by similarity alone can drive the direction of

visual motion, we can examine the trials in which auditory and tactile events alternated in a pairwise manner and cross-modal events were presented synchronously with corresponding visual events (ATP_{AMB}). In these trials, on the basis of timing alone, the direction of visual apparent motion should remain ambiguous. As these trials contain no temporal offset between visual events and the corresponding cross-modal events, they cannot be examined on the same basis as those reported above – direction implied by flanker-visual event temporal offset. However, an implied direction of motion can be determined by the cross-modal event sequence. The grouping hypothesis would suggest that a sequence beginning with a visual presentation on the left, coupled with a given cross-modal signal, followed by a visual presentation on the right, coupled with the same type of cross-modal signal, would imply a direction of motion from left to right – *rightward*. As shown in the right panel of Figure 4, participants’ reports of apparent motion direction were in fact consistent with this suggestion (audio-tactile pairwise ambiguous timing = 0.536, $t_7 = 4.56$, $p = 0.003$; single sample against 0 effect). This result demonstrates that temporal offsets between visual events and corresponding cross-modal events are not required to induce a perception of directional motion in a directionally ambiguous visual display.

Is the modality of flankers critical for grouping?

In the results presented above, when the flanker type was alternated, it changed between different sensory

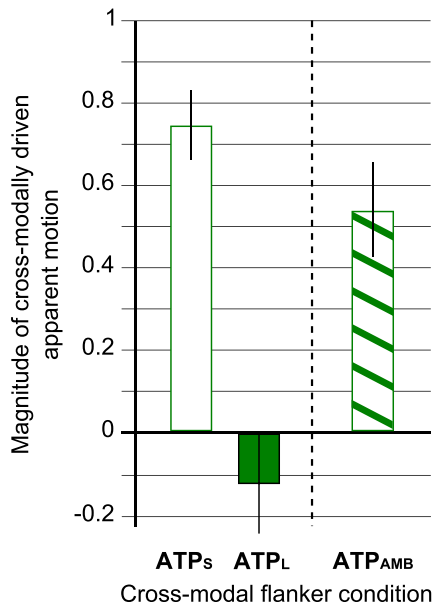


Figure 4. Bar plots depicting magnitude of cross-modal influence on perception of visual motion direction for the condition in which auditory and tactile flankers were alternated in a pairwise manner (ATP). In the left portion, these conditions are separated into short (ATP_S) and long (ATP_L) configurations on the basis of the timing relationship and similarity of signal type grouping cues. The right portion shows data from trials in which there was no temporal disparity between visual events and corresponding cross-modal flankers and, on the basis of timing alone, should remain directionally ambiguous. Error bars indicate \pm standard error of the *mean*.

modalities, auditory and tactile. Therefore, it is possible that the effects of flanker grouping may be limited to cases where flanker pairs are derived from separate sensory modalities. However, if equivalent results are obtained for the conditions in which the different signals originated in the same modality, but differ in feature (i.e., pure-tone and broadband noise), this outcome would indicate that similarity can be indicated by differences more specific than sensory modality.

Examining the basic conditions in which the flanker was alternated both sequentially (NPS) and pairwise (NPP), we can see that compared to the audio only condition the effect of cross-modal flankers on visual apparent motion was significantly reduced (audio only = 0.839; noise-pure tone sequential = 0.417, $t_7 = 4.47$, $p = 0.003$; and noise-pure tone pairwise = 0.453, $t_7 = 4.70$, $p = 0.002$; Figure 5; paired samples-two tailed). Likewise, when we separate the noise-pure tone pairwise condition into short and long combinations, we see a significant difference (noise-pure tone pairwise short = 0.787, noise-pure tone pairwise long = -0.016; $t_7 = 3.49$, $p = 0.01$; Figure 5; paired samples-two tailed). Finally, examining the noise-pure tone pairwise trials in which there is no temporal offset between visual events

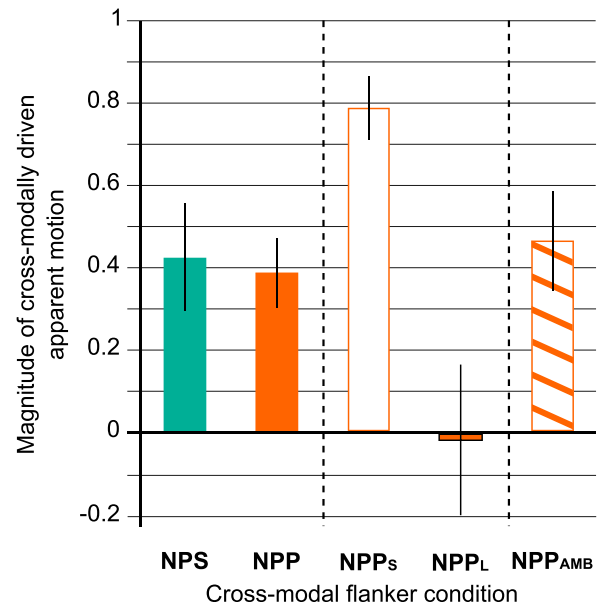


Figure 5. Bar plots depicting magnitude of cross-modal influence on perception of visual motion direction for the conditions in which both cross-modal flankers were auditory (pure-tone or broadband noise). In the left portion, the sequential alternation (NPS) and pairwise alternation (NPP) conditions are depicted. In the central portion, condition NPP is separated into short (NPP_S) and long (NPP_L) configurations on the basis of the timing relationship and similarity of signal type grouping cues. The right panel shows data from trials in which there was no temporal disparity between visual events and corresponding cross-modal flankers (NPP_{AMB}) and, on the basis of timing alone, should remain directionally ambiguous. Error bars indicate \pm standard error of the *mean*.

and the corresponding cross-modal events (NPP_{AMB}), we see that, as in the audio-tactile case, there was a significant modulation of visual apparent motion (noise-pure tone pairwise ambiguous timing; 0.464, $t_7 = 3.75$, $p = 0.007$; single sample against 0 effect). These results indicate that within-modal differences are sufficient to indicate flanker pairs, and subsequently dictate the direction of visual apparent motion.

General discussion

In this study we were interested in whether a recently reported effect, the modulation of visual apparent motion direction by temporally offset, though spatially uninformative, auditory signals could be accounted for by a temporal capture explanation. This explanation assumes a change in the timing of visual signals as induced by temporally offset auditory signals. An alternative explanation was that the grouping of flanker signals could dictate the grouping of visual events to imply a direction of motion. A critical difference of this proposal is that it does not invoke any changes in event

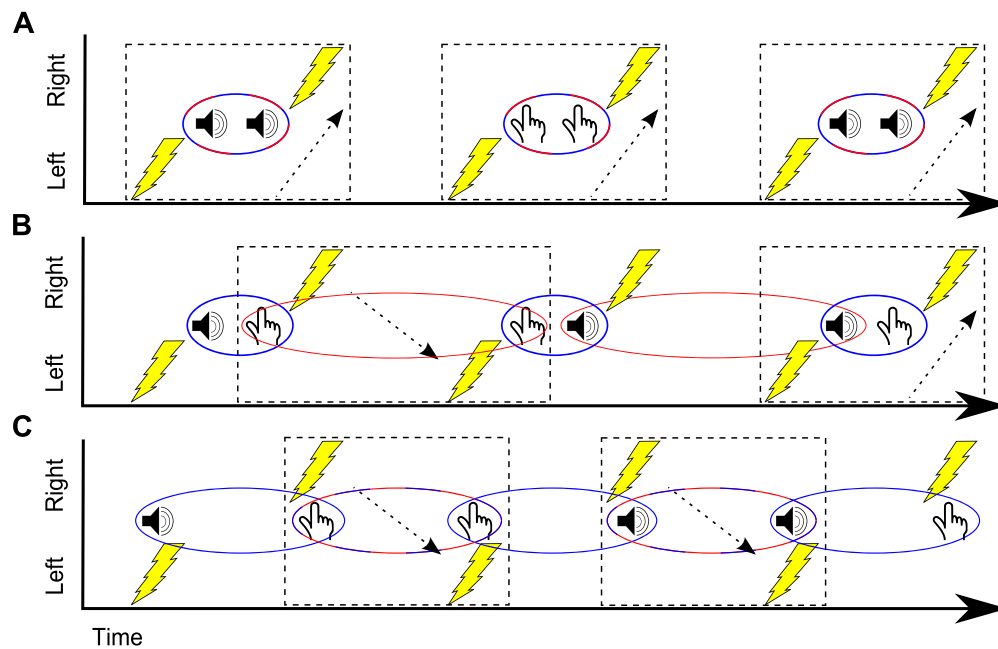


Figure 6. Schematic depiction of flanker grouping account of cross-modally driven visual apparent motion effect. (A–C) Grouping of auditory/tactile flankers by signal type (red ovals) and temporal proximity (blue ovals) dictates the cross-modal sequence segmentation (broken line box) to determine the grouping of visual events and generate directional visual apparent motion direction (broken shafted arrow). (A) When the two cues indicate the same grouping pattern, the segmentation is easily resolved. (B) At different relative cue strengths, the sequence segmentation becomes ambiguous and may be resolved on the basis of one cue or the other. (C) When the temporal proximity cue is not informative the sequence can still be resolved on the basis of similarity.

neural event timing. While we could replicate the original reported effect when cross-modal flanker sequences consisted of only a single signal type, the effect was strongly mitigated for all conditions in which the flanker sequence was alternated. These results are *inconsistent* with a temporal capture account that would predict no difference between sequences with identical temporal profiles.

Results from the conditions containing pairwise cross-modal sequences indicated that the grouping between successive cross-modal flankers on the basis of similarity was critical in determining the direction of apparent visual motion. This conclusion was confirmed by results from the pairwise conditions in which there was no temporal offset between a given visual event and the corresponding cross-modal signal. Despite containing no temporal offset, the direction of visual apparent motion was dictated by the sequence of cross-modal (tactile and/or auditory) events.

Within-modal grouping can be dictated by cross-modal organization

Our results demonstrate that the perceptual organization of tactile and/or auditory sequences can have a strong effect on the organization of visual sequences. Because similar results were obtained when the signals

differed only in feature (pure tone and broadband auditory noise) as when they differed in modality (auditory and tactile), we can characterize the grouping process as occurring supra-modally and being subject to feature information cues. In particular, we believe that the organization of the tactile and/or auditory events is facilitating the *segmentation* of the visual event stream. Figure 6 provides a schematic representation of how we propose such an effect might occur. Successive flanker events are grouped (red and blue ovals) and subsequently dictate the sequence of the combined multisensory events (broken line box). This configuration determines the grouping of visual events to generate directional visual apparent motion (indicated by the arrow). If the visual event pair begins on the left and ends on the right, the perceived direction of motion is rightward. Conversely, if the pair begins on the right and ends on the left, the perceived direction of motion is leftward. The strength of grouping for a given flanker pair is driven by temporal proximity (indicated by the blue ovals) and signal similarity (indicated by the red ovals). When two successive flankers are the same signal type, and close together in time (Figure 6A; audio-tactile pairwise short condition), both the similarity and temporal proximity cues indicate the same grouping configuration. In this case, the strength of grouping between flankers is strong and clearly dictates the visual event sequence. However, when the succes-

sive flankers are the same signal type, but further apart in time (Figure 6B; audio-tactile pairwise long condition), the grouping is less clear. Under these conditions, competition between the conflicting grouping cues determines the appropriate segmentation of the sequence. When there is no temporal offset between visual events and the corresponding cross-modal events, and the temporal distance between events is equal, the temporal proximity cue is not informative. However, the sequence can still be segmented according to the similarity cue (Figure 6C; audio-tactile pairwise ambiguous condition). The original demonstration of this effect (Freeman & Driver, 2008) can be considered an extreme version of the solution depicted in Figure 6A. In that case, as all flanker events are the same, they group based on temporal proximity alone.

The original study by Freeman & Driver (2008, experiment 4) included a similar experiment to the pairwise condition in this study. In that experiment, visual events were presented synchronously with the corresponding cross-modal event and different auditory signals, pure-tone pulses (defined by carrier frequencies of 480 Hz and 880 Hz) alternated in a pairwise manner. The precise value of the similarity grouping effect is not reported, only that it was significantly different from the standard temporally offset auditory condition. On the basis of this result, the authors concluded that an explanation based on cross-modal sequence grouping was unlikely. However, the modulation of visual motion by cross-modal events can be roughly estimated (as estimated in this study; see Figures 4-5) from figure S2 of that study as ~ 0.12 . This value is compared with values of 0.46 for the pure-tone and broadband noise combinations and 0.54 for the audio-tactile combinations in this study. As such, it would seem that there may have been some small effect of grouping by similarity in that study, though the difference in pitch between auditory signals was simply insufficient to promote effective grouping of the auditory sequence into distinct pairs.

Within-modal grouping and cross-modal processing hierarchy

Many studies have demonstrated the influence that within-modal grouping can have on potential cross-modal grouping (e.g., Cook & Van Valkenburg, 2009; Keetels et al., 2007; Klink et al., 2011; Sanabria, Soto-Faraco, Chan, & Spence, 2005; Watanabe & Shimojo, 2001). These studies support the existence of a processing hierarchy wherein within-modal operations are resolved prior to interaction with signals from another modality. Here we demonstrate that within-modal grouping (of visual events) can be determined by perceptual organization within other modalities. Little

evidence exists to support this possibility (however, see O’Leary & Rhodes, 1984 for a controversial example; Spence & Chen, 2012 for review). This result raises an interesting prospect. Equivalent apparent motion phenomena exist for both tactile and auditory sequences. Consequently, it seems unlikely that the cross-modal determination of within-modal grouping is limited to visual apparent motion, or indeed to interactions that uniquely affect vision. Rather, it is likely that similar operations may be identified in many combinations of multisensory events. Under this scenario, the likelihood of a given set of sensory cues being used to determine the organization of another would probably be determined by processes commonly found in cue combination (e.g., maximum likelihood estimation; Battaglia, Jacobs, & Aslin, 2003; Hillis, Ernst, Banks, & Landy, 2002; in comparison, see also Arnold, Tear, Schindel, & Roseboom, 2010; Roach, Heron, & McGraw, 2006). As such, the nature of the processing hierarchy may not be as concrete as implied by previous results.

Grouping by relatedness, sequence segmentation and apparent timing

An interpretation of temporal ventriloquism based on the segmentation of sensory sequences by apparent relatedness is supported by a growing number of recent studies. These studies demonstrate that changes in appearance or apparent timing can be determined purely by apparent relatedness. For example, it has recently been demonstrated that similarity between visual elements can dominate spatio-temporal information to drive motion perception in visual apparent motion displays such as the Ternus and split-motion display (Hein & Cavanagh, 2012; Hein & Moore, 2012). These results may be considered a within-modal analogy of what we report for the multisensory case, with visual feature correspondence used to segment the visual event sequence and determine the direction of apparent motion.

Separately, many studies have demonstrated, using either visual signals alone (Nicol & Shore, 2007), or audio and visual combinations (Keetels & Vroomen, 2005; Parise & Spence, 2009; Spence, Baddeley, Zampini, James, & Shore, 2003; Vatakis & Spence, 2007; Zampini et al., 2003; Zampini et al., 2005), that decreases in apparent relatedness through spatial (Keetels & Vroomen, 2005; Nicol & Shore, 2007; Spence et al., 2003; Zampini et al., 2003; Zampini et al., 2005) or content/semantic differences (Parise & Spence, 2009; Vatakis & Spence, 2007; though see also Keetels & Vroomen, 2011a), can enhance the precision of TOJs between the signals. These studies demonstrate that the apparent relatedness of two events plays a critical role

in the segmentation of sequences into discrete events, regardless of source modality.

Finally, it has been shown that the temporal context in which an audio-visual pair is presented can profoundly alter the perception of their relatedness. For example, the point of subjective synchrony for an audio-visual pair can be shifted away from the timing of an additional, temporally adjacent visual (or auditory) event (Roseboom, Nishida, & Arnold, 2009). This shift results in a given audio-visual pair appearing to be asynchronous at physical timing relationships that, in the absence of the additional event, had seemed synchronous. This result was attributed to a *selective ventriloquism* process wherein the ability to determine which of two discrete visual events was *most* synchronous (related) with a single auditory event facilitated segmentation of the event sequence (Roseboom et al., 2009; Roseboom, Nishida, Fujisaki, & Arnold, 2011).

These studies collectively demonstrate that segmentation of event sequences on the basis of apparent relatedness can have a strong influence on apparent timing. In combination with the results presented in this study, a compelling case is being built to support the suggestion that shifts in event temporal position may not be necessary to explain changes in apparent timing.

Remaining role for temporal capture?

While it is clear that a simple temporal capture account cannot explain the results we have obtained, we are unable to explicitly exclude the possibility that it still has some role. One limitation of the current paradigm is that changes in the temporal offset between visual events and corresponding cross-modal flankers will also change the temporal distance between successive cross-modal flanker events themselves. Therefore, it is not possible to manipulate the influence of one proposed process (e.g., temporal capture) without also changing that of the other (e.g., grouping). We have demonstrated in this study that it is not necessary to invoke temporal capture in order to demonstrate an effect of cross-modal events on visual apparent motion. Critically, this is not to say that there is not some additional contribution of temporal capture beyond the effect of grouping. In fact, a result indicating the auditory capture of visual timing has recently been demonstrated using a type of visual motion stimuli usually considered to be processed pre-attentively and so, presumably, is not subject to the grouping processes we describe here (Kafaligonul & Stoner, 2012). In any case, as the paradigm used in our study cannot directly dissociate the role of grouping by temporal proximity and that of a putative temporal capture, it is necessary to investi-

gate such a contrast in alternative experimental paradigms, such as the traditional temporal ventriloquism task (e.g., Morein-Zamir et al., 2003).

Traditional temporal ventriloquism and sequence segmentation

A critical question that remains is whether the traditional demonstrations of temporal ventriloquism (e.g., Morein-Zamir et al., 2003) can be explained by a method similar to that described above. Such an outcome would add to doubts about the nature of cross-modal temporal capture in a broader context. As the TOJ task employed in typical demonstrations of temporal ventriloquism is fundamentally concerned with segmentation of the visual signal sequence, we believe that sequence segmentation likely contributes to the effect. Any cue that assists in indicating that the two visual signals constitute discrete events is likely to enhance judgment precision. One could consider the fact that precision of TOJs between two audiovisual pairs is enhanced relative to presentations without auditory signals (e.g., Keetels & Vroomen, 2011a) as evidence for this premise. From this perspective, shifting the auditory events further apart in time will likely strengthen the cue to segment the visual sequence and further enhance the precision of TOJs between two audio-visual pairs (i.e., the temporal ventriloquism effect as demonstrated by Morein-Zamir et al., 2003). We believe that manipulations that do not require changes in the temporal profile of the sequence, but assist in segmentation of the stimulus, will yield similar results. For example, it may be possible to enhance the cue to segment the visual stream into two discrete events simply by altering the properties of the cross-modal signals, such as we have done in this study. We will be looking to confirm this possibility in future studies.

Conclusions

Investigations of human perception often focus on the discovery of examples of low level interactions between sensory signals. In the temporal domain, this approach often leads to proposals wherein a change in subjective timing report is explained as a consequence of some change in the representation of temporal position or brain time (see Johnston & Nishida, 2001) for a given event. Phenomena, such as temporal ventriloquism and the corresponding temporal capture account (see Vroomen & Keetels, 2009), or perceptual latency accounts of temporal recalibration (Di Luca et al., 2009; Navarra et al., 2009; nevertheless, see also

Roseboom & Arnold, 2011), color-motion asynchrony (Linares & Lopez-Moliner, 2006; Moutoussis & Zeki, 1997; however, see also Arnold, Clifford, & Wenderoth, 2001; Nishida & Johnston, 2002), and the flash-lag effect (see Linares & Lopez-Moliner, 2007; Whitney & Murakami, 1998; however, see also Arnold, Ong, & Roseboom, 2009) are all examples of where such an approach has been applied. In this study, we provide evidence that, at least for one particular demonstration of temporal ventriloquism, such an explanation proves inadequate. Whether an account such as that we propose proves useful in describing other temporal capture or assimilation effects, such as the original temporal ventriloquism demonstration (e.g., Morein-Zamir et al., 2003) or the cross-modal double flash illusion (Shams, Kamitani, & Shimojo, 2000), remains an empirical question; however, we suspect it likely that the effect of grouping interactions of the form demonstrated in this study will prove to be ubiquitous in human perception.

Acknowledgments

The authors would like to thank Daniel Linares for comments and discussions throughout the course of the project.

Commercial relationships: none.

Corresponding author: Warrick Roseboom.

Email: wjroseboom@gmail.com.

Address: NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Kanagawa, Japan.

References

- Arnold, D. H., Clifford, C. W. G. & Wenderoth, P. (2001). Asynchronous processing in vision. *Current Biology*, *11*, 596–600.
- Arnold, D. H., Ong, Y., & Roseboom, W. (2009). Simple differential latencies modulate, but do not cause the flash-lag effect. *Journal of Vision*, *9*(5):4, 1–8, <http://www.journalofvision.org/content/9/5/4>, doi:10.1167/9.5.4. [PubMed] [Article]
- Arnold, D. H., Tear, M., Schindel, R., & Roseboom, W. (2010). Audio-visual speech cue combination. *PLoS One*, *5*, e10217.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America*, *20*(7), 1391–1397.
- Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processing*. Cambridge, MA: MIT Press.
- Cook, L. A., & Van Valkenburg, D. L. (2009). Audio-visual organization and the temporal ventriloquism effect between grouped sequences: Evidence that unimodal grouping precedes cross-modal integration. *Perception*, *38*(8), 1220–33.
- Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, *98*, 569–603.
- Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: the where and when of consciousness in the brain. *Behavioural Brain Science*, *15*, 183–247.
- Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12):7, 1–16, <http://www.journalofvision.org/content/9/12/7>, doi:10.1167/9.12.7. [PubMed] [Article]
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*, 719–721.
- Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception and Psychophysics*, *63*, 719–725.
- Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology*, *18*(16), 1262–1266.
- Fujisaki, W., Kitazawa, S., & Nishida, S. (2012). Multisensory timing. In B. E. Stein (Ed.), *The new handbook of multisensory processes* (pp. 301–317). Cambridge, MA: MIT Press.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*, 773–778.
- Gepshtein, S., & Kubovy, M. (2007). The lawful perception of apparent motion. *Journal of Vision*, *7*(8):9, 1–15, <http://www.journalofvision.org/content/7/8/9>, doi:10.1167/7.8.9. [PubMed] [Article]
- Hein, E., & Cavanagh, P. (2012). Motion correspondence in the Ternus display shows feature bias in spatiotopic coordinates. *Journal of Vision*, *12*(7):16, 1–14, <http://www.journalofvision.org/content/XX/XX/16>, doi:10.1167/12.7.16. [PubMed] [Article]
- Hein, E., & Moore, C. M. (2012). Spatio-temporal priority revisited: The role of feature identity and similarity for object correspondence in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 975–988.

- Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, *298*, 1627–1630.
- Johnston, A., & Nishida, S. (2001). Time perception: Brain time or event time? *Current Biology*, *11*, R427–R430.
- Kafaligonul, H., & Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals. *Journal of Vision*, *10*(12):31, 1–13, <http://www.journalofvision.org/content/10/12/31>, doi:10.1167/10.12.31. [PubMed] [Article]
- Kafaligonul, H., & Stoner, G. R. (2012). Static sound timing alters sensitivity to low-level visual motion. *Journal of Vision*, *12*(11):2, 1–9, <http://www.journalofvision.org/content/12/11/2>, doi:10.1167/12.11.2. [PubMed] [Article]
- Keetels, M., Stekelenburg, J., & Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. *Experimental Brain Research*, *180*, 449–456.
- Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research*, *167*, 635–640.
- Keetels, M., & Vroomen, J. (2011a). No effect of synesthetic congruency on temporal ventriloquism. *Attention, Perception, & Psychophysics*, *73*(1), 209–18.
- Keetels, M., & Vroomen, J. (2011b). Sound affects the speed of visual processing. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 699–708.
- King, A. J. (2005). Multisensory integration: strategies for synchronization. *Current Biology*, *15*, R339–R341.
- Klink, P. C., Montijn, J. S. & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception, & Psychophysics*, *73*, 219–236.
- Kolers, P. A. (1972). *Aspects of motion perception*. Oxford, UK: Pergamon.
- Linares, D., & López-Moliner. (2006). Perceptual asynchrony between color and motion with a single direction change. *Journal of Vision*, *6*(9):10, 974–981, <http://www.journalofvision.org/content/6/9/10>, doi:10.1167/6.9.10. [PubMed] [Article]
- Linares, D., & López-Moliner. (2007). Absence of flash-lag when judging global shape from local positions. *Vision Research*, *47*, 357–362.
- Miner, N., & Caudell, T. (1998). Computational requirements and synchronization issues of virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, *7*, 396–409.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154–163.
- Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London B*, *264*, 393–399.
- Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. In *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9169–9173.
- Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*, 499–507.
- Nicol, J. R., & Shore, D. I. (2007). Perceptual grouping impairs temporal resolution. *Experimental Brain Research*, *183*(2), 141–148.
- Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology*, *12*(5), 359–368.
- O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception Psychophysics*, *35*(6), 565–569.
- Parise, C., & Spence, C. (2009). “When birds of a feather flock together”: synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, *4*, e5664.
- Powers, R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*, *29*, 12 265–12 274.
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society of London B*, *273*, 2159–2168.
- Roseboom, W., & Arnold, D. H. (2011). Twice upon a time: Multiple, concurrent, temporal recalibrations of audio-visual speech. *Psychological Science*, *22*, 872–877.
- Roseboom, W., Nishida, S., & Arnold, D. H. (2009). The sliding window of audio-visual simultaneity. *Journal of Vision*, *9*(12):4, 1–8, <http://www.journalofvision.org/content/9/12/4>

- journalofvision.org/content/9/12/4, doi:10.1167/9.12.4. [PubMed] [Article]
- Roseboom, W., Nishida, S., Fujisaki, W., & Arnold, D. H. (2011). Audio-visual speech timing sensitivity is enhanced in cluttered conditions. *PLoS One*, *6*, e18309.
- Sanabria, D., Soto-Faraco, S., Chan, J., & Spence, C. (2005). Intramodal perceptual grouping modulates multisensory integration: Evidence from the cross-modal dynamic capture task. *Neuroscience Letters*, *377*(1), 59–64.
- Scheier, C. R., Nijhawan, R., & Shimojo, S. (1999). Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science*, *40*, S792.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*, 788.
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*, 7–10.
- Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception Psychophysics*, *65*, 318–328.
- Spence, C., & Chen, Y. C. (2012). Intramodal and cross-modal perceptual grouping. In B. E. Stein (Ed.), *The new handbook of multisensory processes*. Cambridge, MA: MIT Press (pp. 265–281).
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, *1111*, 134–142.
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Perception and Psychophysics*, *69*(5), 744–756.
- Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1063–1071.
- Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychologica*, *130*, 58–63.
- Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*, 32–35.
- Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, *12*, 109–116.
- Whitney, D., & Murakami, I. (1998). Latency difference, not spatial extrapolation. *Nature Neuroscience*, *1*, 656–657.
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67*, 531–544.
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*, 198–210.