A single auditory tone alters the perception of multiple visual events

Yousuke Kawachi
Kansei Fukushi Research Institute, Tohoku Fukushi University, Sendai, Japan

Philip M. Grove
School of Psychology, The University of Queensland, Queensland, Australia

Kenzo Sakurai
Department of Psychology, Tohoku Gakuin University, Sendai, Japan

We aimed to show that a single auditory tone crossmodally affects multiple visual events using a multiple stream/bounce display (SBD), consisting of two disk pairs moving toward each other at equal speeds, coinciding, and then moving apart in a two-dimensional (2-D) display. The temporal offsets were manipulated between the coincidences of the disk pairs (0 to ±240 ms) by staggering motion onset between the pairs. A tone was presented at the coincidence timing of one of the disk pairs on half of the trials. Participants judged whether the disks in each of two pairs appeared to stream through or bounce off each other. Results show that a tone presented at either of the disk pairs’ coincidence points promoted bouncing percepts in both disk pairs compared to no-tone trials. Perceived bouncing persisted in the disk-pair whose coincidence was offset 60 ms before and up to more than 120 ms after the audiovisual coincidence timing of the other disk-pair. The temporal window of bounce promotion was comparable to that obtained with a conventional SBD. The interaction of a single auditory event and multiple visual events was also modulated by the kind of experimental task (the stream/bounce or simultaneity judgments). These findings suggest that, using a single auditory cue, the perceptual system resolves the ambiguity of the motion of multiple disk pairs presented within the conventional temporal window of crossmodal interaction.

Introduction

In most instances, the human perceptual system must process inputs from different sensory modalities, either integrating them into a single multimodal event, or segregating them into discrete events. The integration of multimodal sensory inputs into unified perception is complicated by the fact that the physical signals of light, sound, and touch arrive at the perceiver at different moments, and the processing speeds of the neural circuits specific to each modality are different. Moreover, the stimuli presented to different sensory modalities are separately processed in sensory modality-specific areas. Signals from these disparate areas must be integrated to generate a unified multimodal percept. Possible sites of integration include the superior temporal sulcus and posterior parietal lobule (for a review, see Driver & Noesselt, 2008).

Nevertheless, it has been repeatedly shown that physically synchronous, or near synchronous, auditory and visual signals are perceptually integrated as a single event (for a review, Spence & Squire, 2003). A representative example for this report is the so-called “stream/bounce effect” (Sekuler, Sekuler, & Lau, 1997). Typically, this effect involves a two-dimensional (2-D) visual display depicting two objects moving towards one another, coinciding, and then moving apart. Observers predominantly perceive streaming in visual-only displays (Bertenthal, Banton, & Bradbury, 1993; Sekuler & Sekuler, 1999). An additional transient (visual, auditory, or tactile) presented near the instant of coincidence induces perceived bouncing (e.g., Grove, Kawachi, & Sakurai, 2012; Grove & Sakurai, 2009; Kawabe & Miura, 2006; Kawachi & Miura, 2006, 2013; Sekuler et al., 1997; Watanabe & Shimojo, 1998). That is, the transient and the visual coincidence are perceptually integrated as a single collision event rather than two unimodal events occurring independently near the same instant in time.

Additional (contextual) events presented near the point of coincidence, however, greatly attenuate the induced bouncing. For example, the sound-induced bias towards bouncing disappears when an additional identical sound is presented just before and after the sound at the coincidence (Watanabe & Shimojo, 2001). These authors conclude that the auditory signals are perceptually grouped, reducing the salience of the sound that is synchronous with the visual coincidence. Similarly, Kawachi and Gyoba (2006, 2013) reported that a nearby moving object tracking alongside one of the objects in the stream/bounce display (SBD) diminishes the sound-induced bias towards perceived bouncing. They proposed that the presence of the nearby contextual objects and possible visual grouping between visual objects interferes with the crossmodal interaction underlying the bouncing percept.

These previous studies demonstrate how additional sounds or an additional moving object diminish the effectiveness of a bounce-inducing sound in SBDs. In the present study, we investigate whether or not a single sound at the point of coincidence in a SBD can alter the perception of that display and an additional SBD for which the coincidence is synchronous or temporally offset from the sound, compared to viewing the same SBDs with no sound. If so, this implies that a single sound can exert influence over two visual events synchronous with the sound, as well as events where one visual event is synchronous with the sound and the other occurs sometime beforehand or afterward. This expands previous research on crossmodal interaction, which has looked almost exclusively at interactions between single auditory and single visual events, that is, a one on one combination of an auditory and a visual event (Fendrich & Corballis, 2001; Radeau & Bertelson, 1977; Sekuler et al., 1997; Spence, 2007; Welch & Warren, 1980).

Additionally, the present study contributes to the current knowledge on the temporal window of audiovisual interaction. Previous studies have demonstrated that the perceptual system’s temporal window for integration depends on different combinations of sensory inputs. For combinations of relatively simple auditory and visual stimuli, the integration window is approximately ±200 ms from physical audiovisual coincidence (Dixon & Spitz, 1980; Shimojo & Shams, 2001). For combinations of more complex auditory and visual stimuli such as speech and facial movements, the temporal window is wider (Spence & Squire, 2003). Recent investigations have shown that the temporal window of crossmodal interaction can be flexibly adjusted. Fujisaki, Shimojo, Kashino, and Nishida (2003) showed that the temporal audiovisual integration window can be recalibrated by adaptation to a temporal lag between audio and visual events. They concluded that the perceptual system continuously updates the audio/visual simultaneity window to reduce perceived time lags between two different sensory inputs from the same physical events. Here, we extend previous investigations that focus on a one-on-one interaction of auditory and visual signals, to show that audiovisual interaction occurs across multiple visual motion sequences and a single tone within the conventional temporal window of audiovisual interaction.

In this context, previous studies have primarily investigated a single isolated auditory stimulus and a single isolated visual event (Fendrich & Corballis, 2001; Radeau & Bertelson, 1977; Sekuler et al., 1997; Spence, 2007; Welch & Warren, 1980). However, a few recent studies have investigated the combination of a single auditory event and multiple visual events (e.g., Roseboom, Nishida, & Arnold, 2009; Roseboom, Nishida, Fujisaki, & Arnold, 2011; Van der Burg, Awh, & Olivers, 2013). These researchers demonstrated that interactions within the classic audiovisual simultaneity window are limited to one auditory event and a single visual event. However, it is possible that an auditory signal exerts an influence on the perceptual resolution of multiple visual events as long as they occur within the temporal window of crossmodal interaction. In the present study, we used a novel stimulus configuration and experimental task based on SBDs. We explored the temporal window of the audiovisual interaction in which the percepts of multiple visual events were altered by a single tone.

Here, we report a new phenomenon in which a single auditory tone alters the perceptual interpretation of two stream/bounce motion sequences whose coincidences are temporally offset. To anticipate, we demonstrate that an auditory tone synchronous with the coincidence of a pair of motion disks in a SBD strongly influences the resolution of that display and an additional SBD in which the time of coincidence is offset relative to the former. The influence of the auditory tone extends both to a coincident visual event and to a preceding/subsequent visual event.

**Experiment 1**

**Methods**

**Participants**

Twelve healthy adults participated in the experiment. Except for one author (YK), all participants were naive to the purpose of the experiment. All had either normal or corrected-to-normal vision and reported no hearing anomalies. Written informed consent was obtained from all participants according to the
guidelines of the Ethical Committee of Tohoku Fukushi University and the Declaration of Helsinki.

Stimuli

All stimuli were generated on a PC (Dell Precision T5500; Dell, Austin, TX) running Matlab 2007a (The MathWorks, Inc., Natick, MA) and the Cogent Graphics package (http://www.vislab.ucl.ac.uk/cogent.php). Visual stimuli were presented on a CRT monitor (Sony GDM-F520, Sony, Tokyo, Japan; refresh rate 100 Hz; resolution 1024 × 768 pixels). As shown in Figure 1, the stimuli consisted of a white fixation dot, presented at the center of the monitor (76.45 cd/m², 0.53 deg of visual angle), on a gray background (39.29 cd/m²). SBDs consisted of a red (13.20 cd/m²) and a green disk-pair (13.22 cd/m²). Each disk was 0.53 deg of visual angle in diameter. The red and green disk pairs traced orthogonal oblique (±45°) trajectories crossing the fixation point. The assignment of disk color (red/green) and that of the oblique trajectories (±45°) were randomized. All disks were initially presented 5.47 s from the fixation point. The disks moved at a constant speed of 15.2°/s. In order to describe the relative timing of the point of coincidence in red and green disk pairs, we designate one as the standard SBD (StdSBD). The initial motion onset of the StdSBD is time 0. The other disk-pair is designated as the test SBD (TstSBD). The motion onset of the TstSBD could occur from 240 ms before to 240 ms after the motion onset of the StdSBD. Disks of each of the StdSBD and TstSBD were presented simultaneously and were stationary before the motion onset and after the motion offset. The total presentation duration of the StdSBD and TstSBD was constant across trials. Each of the two SBDs was assigned to the StdSBD or TstSBD in a random manner on a trial-by-trial basis. Therefore, participants were not aware which of the two SBDs was the StdSBD or TstSBD. A brief tone (1800 Hz, 10 ms, 68 dB SPL; background noise level: ~40 dB) was introduced only at the coincidence point of the stdSBD through headphones (HDA 200; Sennheiser, Wedemark, Germany). The physical simultaneity of auditory and visual stimuli was assessed with an oscilloscope (TBS 1102; Tektronix, Pittsfield, MA). Deviations from simultaneity did not exceed 5 ms.

Procedure

The participants were seated with their heads resting on a chin rest at a viewing distance of 57 cm. Participants held their gaze on the fixation dot for the duration of the visual stimulus. On a given trial, participants began the stimulus presentation by pressing one of the assigned keys. At the end of the stimulus presentation, participants were asked to make two responses by pressing the appropriate buttons. One response was to indicate whether the StdSBD (red or green disk-pair) appeared to stream or bounce. The other was to indicate whether the TstSBD (green or red disk-pair) appeared to stream or bounce. Participants completed trials for each of the temporal offsets (9) × presence/absence of a tone (2) × disk color (2) × oblique trajectory (2). The StdSBD and TstSBD were combined factorially with all temporal offset and tone presence/absence. The StdSBD with a tone, TstSBD with a tone, the StdSBD with no tone, and TstSBD with no tone conditions are referred to hereafter as the StdSBD-t, TstSBD-t, StdSBD-nt, and TstSBD-nt conditions, respectively.

To replicate past results that have employed conventional SBDs, participants performed trials in which only the TstSBD with a tone was presented, hereafter referred as SnglTstSBD-t condition. The SnglTstSBD-t was presented with each of the temporal offsets (9) × disk color (2) × oblique trajectory (2). The tone was presented with one of nine temporal offsets from the coincidence of the TstSBD (0, ±30, ±60, ±120, ±240 ms). As a baseline, participants completed trials in which only the StdSBD was presented, trials consisting of disk color (2) × oblique trajectory (2; single StdSBD with no tone condition, hereafter referred to as SnglStdSBD-nt condition). Within an experimental block, each of all trial types was repeated two times in random order for a total of 224 trials, on a trial-by-trial basis. Participants completed three experimental blocks with a rest between blocks. In total, each of the display-tone combination conditions (StdSBD-t, TstSBD-t, StdSBD-nt, TstSBD-nt, and SnglTstSBD-t) was repeated 24 times for each temporal offset condition.

Results

Group mean data are shown in Figure 2. We tabulated the proportion of bounce responses in each condition for each participant and these were used as the units for statistical analyses. Although we controlled the extraneous factors of disk color and trajectory angle of the disks by randomization, we conducted paired t tests to examine whether the differences of disk color or trajectory angle affect the proportion of bouncing response. We found no significant differences, color: t(11) = −1.50, p = 0.162; trajectory angle: t(11) = −0.890, p = 0.392). Therefore, we combined data from both color of disk-pair and orientation of motion trajectory for each condition. Inspection of the figure reveals that the proportions of the bounce response in the TstSBD-t condition is very similar to that in the SnglTstSBD-t condition for all
temporal offsets except 240 ms, suggesting that a tone with or without the StdSBD has a similar modulatory effect on the perceptual judgments of the TstSBD. As reported in many previous studies (e.g., Sekuler et al., 1997), the StdSBD-t is predominantly judged as bouncing across all temporal offset conditions compared to the StdSBD-nt condition. Moreover, the promotion of bouncing is most pronounced at the 0-ms offset in both the StdSBD and TstSBD conditions. Interestingly, we also observed an increase in bounce responses in the no-sound condition at 0-ms offset for both StdSBD-nt and TstSBD-nt relative to the other temporal offsets. For all remaining temporal offset conditions, the proportion of bounce responses diminished relative to the 0-ms offset condition. Nevertheless, the proportion of bounce responses was higher in the sound condition than in the no-sound condition for temporal offsets between −60 ms and 240 ms.

These observations were supported by a two-way repeated measures ANOVA conducted on the proportion of bounces reported in each condition, with display-tone combination (StdSBD-t, TstSBD-t, StdSBD-nt, TstSBD-nt, and SnglTstSBD-t) and temporal offset as factors. We found a significant main effect for tone-display combination, $F(4, 44) = 25.06$, $p < 0.001$, and a significant main effect for temporal offset, $F(8, 88) = 19.50$, $p < 0.001$. The interaction effect between tone-display combination and temporal offset was also significant, $F(32, 352) = 6.22$, $p < 0.001$.

**Analysis of display type**

A posthoc analysis ($p < 0.05$, corrected for multiple comparisons by Ryan’s method; Ryan, 1960) compar-
ing the mean proportion of bounces in the display-tone combination conditions for each offset condition revealed that at -240 and -120 ms, the StdSBD-t condition generated higher bounce promotion than the others. At ±30 and ±60 ms, the StdSBD-t, TstSBD-t, and SnglTstSBD-t conditions yielded a higher proportion of bounce responses than those in the StdSBD-nt and TstSBD-nt conditions. At 0 ms, the StdSBD-t and TstSBD-t conditions yielded a higher proportion of bounce responses than those in the StdSBD-nt and TstSBD-nt conditions. At 120 ms, the StdSBD-t, TstSBD-t, SnglTstSBD-t, and TstSBD-nt conditions showed higher bounce proportion than the StdSBD-nt condition. The StdSBD-t condition also showed a higher bounce proportion than that in the TstSBD-nt condition. At 240 ms, the StdSBD-t and TstSBD-t conditions showed higher bounce proportion than the StdSBD-nt and TstSBD-nt conditions. Bounce proportion in the StdSBD-t condition was also higher than the TstSBD-t and SnglTstSBD-t conditions, and bounce proportion in the tSBD-t condition was higher than the SnglTstSBD-t condition.

In sum, a single tone inducing bouncing in multiple visual events is evident from the significantly higher proportion of bounce responses in both the StdSBD-t and TstSBD-t conditions than that in the StdSBD-nt and TstSBD-nt conditions from -60 to +240 ms (though at 120 ms, the difference between the TstSBD-t and -nt was not significant). Moreover, the proportion of bounce responses in the TstSBD-t condition is comparable to that in the SnglTstSBD-t condition for nearly all temporal offsets.

However, there are at least two explanations for these results. One is that as we suggested, a single tone affects multiple visual events. The other is that the tone-induced bouncing of one disk-pair itself biases the judgment of the other disk-pair toward bouncing. In order to test the latter possibility, we calculated the phi coefficient (a measure of the degree of correlation between the two stream/bounce judgments within a trial) for each temporal offset combined with the presence or absence of a tone from the pooled data from all participants. Phi coefficients were 0.302, 0.301, 0.583, 0.311, 0.948, 0.861, 0.733, 0.274, and 0.371 for each of the respective temporal offsets without a tone \( (p < 0.001) \). At \( -0.301, -0.195, 0.217, 0.311, 0.314, 0.341, 0.234, 0.156, \) and 0.219 for each of the respective temporal offsets with a tone \( (p < 0.001, \text{except} \ p < 0.01 \text{at 120 ms}) \). The within-trial correlations without a tone were very strong (Evans, 1996) for three of the temporal offsets. However, the correlations with a tone were weak at all temporal offsets. Moreover, the positive correlations without a tone were significantly stronger than those with a tone at -240, -120, -60, 0, 30, and 60 ms \( (p < 0.05) \). Although these findings do not exclude the possibility of the response bias within a trial, they show that the bounce judgment of one disk-pair is more weakly associated with the judgment of the other disk-pair for trials with a tone than for ones without a tone. Therefore, the results of our correlation analyses do not support the proposal that the tone-induced bouncing of one disk-pair biases the judgment of the other disk-pair toward bouncing. For further inspection of the results of this experiment, see Appendix A.

### Analysis of temporal offset

A second posthoc analysis \( (p < 0.05, \text{corrected for multiple comparisons by Ryan’s method}) \) was conducted for each display-tone combination. We examined and found significant differences in the proportion of bounce responses between the 0-ms and the other temporal offsets: in the StdSBD-t condition, 0 ms > ±240, ±120, ±60, and ±30 ms; in the TstSBD-t condition, 0 ms > ±240, ±120, and 60 ms; in the SnglTstSBD-t condition, 0 ms > ±240 and ±120 ms; in the StdSBD-nt condition, 0 ms > ±240, ±120, ±60, and ±30 ms, and in the TstSBD-nt condition, 0 ms > ±240, ±120, ±60, and ±30 ms. Bouncing responses are maximal around the 0-ms temporal offset condition in all the display-tone combination conditions.

### Control experiment

If multiple visual events are influenced by a single auditory event, multiple bouncing may be obtained even in some visual configurations in which the coincidence points of the disk pairs are spatially offset. Here, we tested a visual configuration consisting of two disk pairs with horizontal trajectories vertically and horizontally offset from one another. The colors, designation of StdSBD and TstSBD, and audiovisual timings were all identical to Experiment 1. The following modification were made: The motion trajectories of both disk pairs were horizontal, vertically offset from each other by 1.06 deg of visual angle and one of the trajectories was leftwardly offset and the other rightwardly offset by 0.53 deg of visual angle (Figure 3a). A fixation dot also subtended 0.34 deg of visual angle. These trajectories avoided the complexity of the visual configuration in Experiment 1 and reduced the tracking difficulty induced by disks approaching in any spatial points except the coincidence points (Makovski & Jiang, 2009; Tombu & Seiffert, 2008). Moreover, using various motion trajectories may lead to the generalization of multiple bouncing effects.

Group mean data are shown in Figure 3b. The results show that audiovisual multiple bouncing is evident from the significantly higher proportion of bounce responses in both the StdSBD-t and TstSBD-t conditions than that in the StdSBD-nt and the
TstSBD-nt conditions from −60 to +120 ms ($p < 0.05$). Moreover, the proportions of bounce responses in the TstSBD-t condition were comparable to those in the SnglTstSBD-t condition for nearly all temporal offsets except −120 ms and −30 ms. These findings are consistent with those of Experiment 1. Therefore, we concluded that audiovisual multiple bouncing can be obtained even when the motion trajectories are parallel and spatially offset from one another.

We also calculated the phi coefficient for each temporal offset combined with the presence or absence of a tone. Phi coefficients were 0.714, 0.797, 0.947, 0.935, 0.956, 0.935, 0.834, 0.772, and 0.694 for each of the respective temporal offsets without a tone ($p < 0.001$), and −0.336, −0.419, 0.200, 0.670, 0.812, 0.684, 0.396, 0.150, and −0.160 for each of the respective temporal offsets with a tone ($p < 0.001$, except $p < 0.05$ at 120 ms and $p < 0.01$ at 240 ms). Consistent with the results of Experiment 1, the within-trial correlations without a tone were stronger than the ones with a tone at all temporal offsets ($p < 0.05$). These findings strengthen those from Experiment 1, demonstrating that response bias cannot explain the promotion of multiple bouncing for trials with a tone. For further inspection of the results of this experiment, see Appendix A.

To test whether Experiment 1 and the Control experiment generated similar results, we conducted a three-way mixed design ANOVA with the kind of experiment as a between-participant factor, display-tone combination and temporal offset as within-participant factors and the proportion of reported bounces as the repeated measure. We found no significant main effect for experiment type, $F(1, 14) = 0.25$, $p = 0.510$. We found a significant main effect for display-tone combination, $F(4, 56) = 48.90$, $p < 0.001$, and a significant main effect for temporal offset, $F(8, 112) = 9.29$, $p < 0.001$. The three-way interaction among experiment type, display-tone combination, and temporal offset was also significant, $F(32, 352) = 6.22$, $p < 0.001$. A posthoc analysis revealed that the bounce proportion of the StdSBD-nt at 0, 30, and 60 ms and TstSBD-nt at 0, −30, and ±60 ms were higher in Experiment 1 than in the Control Experiment ($p < 0.05$). We speculate that these effects depend on whether the coincidence points of the two disk pairs were spatiotemporally overlapping. For tone present.
trials, the proportion of bounce responses in the StdSBD-t and SnglTstSBD-t conditions were higher in the Control experiment than in Experiment 1 at -240 ms and 240 ms, respectively ($p < 0.05$). The proportion of bounce responses in the TstSBD-t condition was higher in the Experiment 1 than in the Control experiment at -60 ms. Although there were minor differences between the experiments, the general trends of the Experiment 1 closely matched those of the Control experiment.

**Experiment 2**

As we reviewed in the Introduction section, recent studies on an interaction between a single auditory event and multiple visual events, using perceived simultaneity as their dependent measure, reported that in the case where a tone is combined with (perceived to be synchronous with) a visual event, the tone tends not to be combined with an additional visual event (Roseboom et al., 2009; Roseboom et al., 2011; Van der Burg et al., 2013). As a phenomenon, the effect seems to be inconsistent with our report. The major difference between the present and previous studies seems to be the participant’s experimental task: stream/bounce judgment or simultaneity judgment. Here, we tested whether the presence of the StdSBD synchronous with the tone (for which we expect simultaneity responses) promotes the segregation of the TstSBD and tone (for which we do not expect simultaneity responses). We used identical stimuli to those in Experiment 1 but required participants to make simultaneity judgments between the sound and the coincidence points of the two disk pairs. We hypothesized that if the additional visual event (TstSBD) tends to be not combined with the tone, the proportion of simultaneity response in the TstSBD-t condition will be lower than that in the SnglTstSBD-t condition where a tone and an isolated visual event were presented.

**Methods**

**Participants**

Ten healthy adults participated. Eight participants were from Experiment 1 and one of the authors (YK). Except for YK, all participants were naive to the purpose of the experiment.

**Stimuli and procedure**

Stimuli were identical to Experiment 1. Conditions were nine temporal offsets relative to the coincidences of the disk pairs, and three display-tone combinations (StdSBD-t, TstSBD-t, and SnglTstSBD-t). We excluded the no-tone conditions from the experiment. Participants were asked to make two responses by pressing one of the assigned keys. One was whether the coincidence of the StdSBD (red or green disk pair) and tone are simultaneous or not. The other was whether the coincidence of the TstSBD (green or red disk pair) and tone are simultaneous or not. Within an experimental block, participants completed two trials for each of the temporal offsets (9) × disk color (2) × oblique trajectory (2) for the StdSBD-t and TstSBD-t conditions. Participants also completed two trials for each of the temporal offsets (9) × disk color (2) × oblique trajectory (2) for the SnglTstSBD-t condition. Each block consisted of 144 randomly ordered trials. Participants completed three blocks with a rest between blocks. In total, each of display-tone combination conditions (StdSBD-t, TstSBD-t, and SnglTstSBD-t) was repeated 24 times for each temporal offset condition.

**Results**

Group mean data are shown in Figure 4. We tabulated the proportion of simultaneity responses in each condition for each participant, and these were used as the units for statistical analyses. Most notably, the proportions of simultaneity responses in the TstSBD-t condition sharply decreased with larger temporal offsets in comparison to those in the SnglTstSBD-t condition. This tendency suggests that when participants were engaged in the simultaneity judgment task, not stream/bounce judgment task, an auditory event synchronous with a visual event might be segregated from the other visual event. The tendency is clearly different from that of the Experiment 1 and is consistent with the findings in the previous studies (Roseboom et al., 2009; Roseboom et al., 2011; Van der Burg et al., 2013).

These observations were supported by a two-way repeated measures ANOVA conducted on the proportion of simultaneity responses reported in each condition, with display-tone combination (StdSBD-t, TstSBD-t, and SnglTstSBD-t) and temporal offset as factors. Both main effects of tone-display combination and temporal offset were significant, $F(2, 18) = 105.88$, $p < 0.001$; $F(8, 72) = 97.44$, $p < 0.001$, respectively. The interaction effect between tone-display combination and temporal offset was also significant, $F(16, 144) = 33.75$, $p < 0.001$. A posthoc analysis of the mean proportion of simultaneity responses in the display-tone conditions for each offset revealed that at ±240, ±120, and 60 ms, the StdSBD-t condition generated a higher proportion of simultaneous re-
responses than the TstSBD-t. At ±240 and ±120 ms, the StdSBD-t generated higher proportion of simultaneous responses than the SnglTstSBD-t. Most importantly, at ±120 and ±60 ms, the TstSBD-t condition yielded a lower proportion of simultaneity responses than those in the SnglTstSBD-t condition. In sum, the finding indicates that when participants are engaged in a simultaneity judgment task, an interaction of an auditory event and multiple visual events does not occur.

**Discussion**

The present study clearly showed that a single auditory tone synchronous with the point of coincidence in one SBD modifies participants’ responses to that display and an additional SBD when the trajectories intersect one another (Experiment 1) or when they are parallel and vertically and horizontally separated (Control experiment). The tone promoted multiple bouncing in conditions where the coincidence in the StdSBD led the other coincidence in the TstSBD by between 30–240 ms (120 ms in the Control experiment) and lagged the other one by between 30–60 ms. The bounce-promoting effect of the tone synchronous with the StdSBD coincidence extends both forward and backward in time on the resolution of the TstSBD. It is noteworthy that the proportion of bounce responses and the temporal window within which bouncing was promoted were comparable in the TstSBD-t and SnglTstSBD-t conditions. This finding indicates that a tone exerts a similar modulatory effect on audiovisual event judgments on at least one additional visual event.

One account of the stream/bounce effect is that a transient event may disrupt attentive tracking of multiple moving disks, inducing a bouncing percept when disk pairs coincide (Watanabe & Shimojo, 1998). According to this account, a brief tone at the coincidence of the StdSBD disrupts attention to the disk motion and, consequently, spatiotemporal motion integration. This disruption hinders the establishment of correspondence of objects in a straight-motion path and the disks are perceived to bounce (Watanabe, 2001; Kawabe & Miura, 2006; Kawachi & Gyoba, 2013; Roudaia, Sekuler, Bennett, & Sekuler, 2013). Extending this attentive tracking account to the displays used in the present paper, a brief auditory tone might distract tracking of an additional visual event that is spatiotemporally proximate to the tone, promoting bouncing. Even though the StdSBD is physically simultaneous with a tone and the coincidence and the tone is apparently integrated, our results show that a tone that is spatiotemporally proximate to the coincidence of the TstSBD exerts an influence on the percept of the TstSBD, possibly by distracting the (attentive) tracking of visual motion. Moreover, a sensory transient due to a sudden physical change is considered to draw our attention to a location of interest (Kanai & Verstraten, 2004) although Dufour, Touzalin, Moessinger, Brochard, and Després (2008) reported audiovisual bouncing effect in the absence of conscious perception of an auditory transient. Nevertheless, the balance of...
evidence to date on the stream/bounce effect favors the attentional distraction account.

However, this account, on its own, does not predict the noted asymmetrical window of influence of an auditory tone in the TstSBD-t condition. How might we explain the temporal asymmetry of the influence of the tone on visual events? Further consideration about the time it takes to deploy attention across modalities may be illuminating. Attentional accounts of these audiovisual interactions (Watanabe & Shimojo, 1998) posit that the temporal asymmetry observed is due to the sluggishness of shifts in attention between modalities. Consider an observer attentively tracking the disks of a SBD. If a tone occurs before coincidence of the SBD, then the observer can deploy attention to the tone with some time left to determine how it relates to the coincidence. Alternatively, if the tone sounds too long after coincidence, then by the time attention has been deployed to the sound, the disks are too far apart to be interpreted as bouncing. This argument applies to the conventional stream/bounce display and can be applied to our multiple SBDs. The additional finding reported here is that the tone in our displays is not exclusively bound to the synchronous visual coincidence but exerts influence on an additional motion sequence occurring subsequently and, to a lesser extent, one occurring previously. Thus, the sound before the to-be-modulated event modulates the perception of multiple visual events over a wider range of temporal offsets than a sound occurring after the to-be-modulated event (Watanabe, 2001).

An alternative explanation for the asymmetrical audiovisual integration window may be related to the competition between unimodal and crossmodal processing. Here, unimodal processing corresponds to perceptual grouping based on the law of smooth continuation, and crossmodal processing corresponds to the auditory signal modulating the percept of the continuous motion displays. Previous studies have reported that unimodal processing, such as perceptual grouping in a single modality, overrides crossmodal processing (Kawachi & Gyoba, 2006, 2013; Keetels, Stekelenburg, & Vroomen, 2007; Sanabria, Soto-Faraco, Chan, & Spence, 2005; Watanabe & Shimojo, 2001). The present results suggest that a tone synchronous with the preceding StdSBD coincidence overrides a possible contribution of perceptual grouping cues such as the law of smooth continuation (Grassi & Casco, 2010; Metzger, 1937) that could contribute to the resolution of the subsequent TstSBD. When the tone/StdSBD pairing follows the TstSBD, the perceptual ambiguity in the leading TstSBD could be resolved based on the law of smooth continuation. The presentation of the tone at coinci-

dence in the subsequent StdSBD could trigger a re-examination of the resolution of perceptual ambiguity in the preceding TstSBD, but it would have to occur before the grouping processes are complete. Temporal offsets for which bouncing was promoted (30–60 ms before the subsequent StdSBD and a tone) might be soon enough to override unimodal grouping processes related to the preceding TstSBD before they are complete. This analysis opens up another possible line of investigation in which transients such as tones or visual flashes at or near the point of coincidence, which would suggest bouncing, are pitted against perceptual grouping cues, such as proximity, size, shape, color, and smooth continuation, which could be rendered consistent with streaming. A variety of cue combinations could be investigated to determine the relative strength of organizing cues in a similar approach to Grove and Sakurai (2009), who showed that auditory-induced bouncing persisted even when the trajectories of the individual disks were spatially displaced, tipping the probability of the resolution towards streaming (see also Grove, Ashton, Kawachi, & Sakurai, 2012 for a similar effect when the targets were rendered distinguishable via texture density differences).

One important consideration is that the temporal properties of the audiovisual interactions as described above are specific to the types of stimuli and experimental tasks employed. Generally, the temporal window of crossmodal integration is broader for pairing of complex stimuli such as lip movements and voices than for geometric shapes and pure tones or white noise (Spence & Squire, 2003). Additionally, the temporal window is variable across different kinds of crossmodal illusions involving relatively simple stimuli. For example, the sound-induced flash illusion is effective with only about 100-ms separation between flashes and beep sounds (Shams, Kamitani, & Shimojo, 2000, 2002) but sound-induced bouncing is obtained over a range of about 200 ms between the coincidence of objects and a collision sound (Fujisaki et al., 2003; Grove et al., 2012; Watanabe, 2001). Differences of stimuli induce a different state of expectation, such as the unity assumption in which observers assume that two sensory signals from different modalities come from the same multisensory event (Vatakis & Spence, 2008). Moreover, various differences in experimental tasks employed in the crossmodal phenomena discussed here may vary in the attentional demands required of participants (e.g., Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006) and tend to shift participants’ response criteria based on incidental features of the question being asked (Yarrow, Jahn, Durant, & Arnold, 2011).
The different pattern of results between Experiments 1 and 2 clearly show the different effects of the stream/bounce and simultaneity judgment tasks on an interaction between an auditory event and multiple visual events. Notably, in Experiment 2, the proportion of simultaneity responses in the TstSBD-t condition tends to be lower than that in the SnglTstSBD-t at ±120 and ±60 ms. The tendency is clearly different from that of the Experiment 1. If the proportion of bounce responses in the SnglTstSBD-t indicates the (conventional) temporal window of audiovisual interaction, the lower tendency in the TstSBD-t will indicate that a tone combined with a visual event (StdSBD) is unlikely to be combined with an additional visual event (TstSBD). The simultaneity judgment task is different from the stream/bounce perception task (e.g., Freeman et al., 2013). In the former, participants focus on which of two visual events is simultaneous with an auditory event, but in the latter the task does not explicitly ask participants about whether an auditory event was simultaneous with a visual event or not (although the stream/bounce judgment is dependent upon audiovisual simultaneity; Fujisaki et al., 2003; Grove et al., 2012; Watanabe, 2001). Thus, this task seems to be relatively free from participants focusing on a one auditory event to one visual event combination. Future research should investigate whether these variations in stimuli and measurements are associated with varying crossmodal interaction and its temporal window in more detail.

In the Control experiment, although the proportions of bounce responses in the TstSBD-t condition were higher than those of the TstSBD-nt and StdSBD-nt condition from −60 to +120 ms, we observed a drop in the proportion of bounce responses in the TstSBD-t condition relative to the SnglTstSBD-t condition at −120 and −60 ms (marginally significant at −60 ms). This is similar to the drop in the proportions of simultaneity responses in the TstSBD-t condition relative to the SnglTstSBD-t condition at −120 ms and −60 ms in Experiment 2. These findings suggest that while an auditory tone promoted bouncing responses in both StdSBD and TstSBD for temporal offsets of −60 to 120 ms, the bounce proportion in the TstSBD-t may be also affected by the segregation of the TstSBD and tone. Thus, it is possible that these findings indicate two types of interactions between an auditory event and multiple visual events. One is a cooperative crossmodal interaction that facilitates the same percept for the two visual events, as indicated by audiovisual multiple bouncing. The other is a competitive interaction that suppresses the same percept for the two events. Future and more detailed investigations will be needed to confirm the existence of the two possible types of interactions and to understand what determines the way audiovisual signals interact.

In sum, we focused on the temporal properties of crossmodal interactions utilizing a novel multiple SBD, to test if a single auditory event can influence the perceptual resolution of multiple visual events. The results clearly showed that a brief tone modulates the perception of a coincident visual event, and an additional visual event occurring from 60 ms before up to more than 120 ms after its occurrence when it is synchronized with one of the visual events. The temporal order and separation of audio and visual stimuli modulates the way the perceptual system resolves at least two ambiguous visual events. These findings contribute to our understanding of a fundamental problem for a perceptual system that is required to process information in cluttered multisensory environments where there are many candidate visual events that can be paired with a single auditory event.

Keywords: crossmodal interaction, multiple stream/bounce display, temporal window

Acknowledgments

Grant-in-Aid of the Japan Society for the Promotion of Science (JSPS) for Specially Promoted Research (no. 19001004) and for Scientific Research (no. 25285202) to KS and for Scientific Research (no. 25530173) to YK; Australian Research Council International Fellowship (no. LX0989320) to PMG.

Commercial relationships: none.
Corresponding author: Yousuke Kawachi.
Email: yousuke.kawachi@gmail.com.
Address: Kansei Fukushi Research Institute, Tohoku Fukushi University, Sendai, Japan.

Footnotes

1In order to keep this condition similar in every possible respect to the multi-SBDs, we simply rendered the StdSBD invisible. This made the display identical to conventional SBDs but matched all the critical timings across all our displays.

2A particularly striking feature of the present data is that participants predominantly reported bouncing in both the StdSBD and TstSBD at 0-ms offset condition without a tone. This bouncing is contrary to the existing literature where the noted perceptual bias in transient free displays is towards streaming. One
possibility is that motion sequences in this study can prompt detectors sensitive to complex motion such as expanding-contracting motion (Morrone, Burr, & Vaina, 1995; Regan & Beverley, 1978; Sekuler, 1992). The specialized detectors may integrate motion signals of different directions from different disk locations, generating the bounce percept similar to the percept of motion in depth, as in a looming display. Another possibility is the limited attentive tracking of objects when they are close to the point of coincidence. At the point of coincidence, participants may have difficulty in distinguishing the four disks from each other. This loss of discriminability could make attentive tracking of objects difficult (Makovski & Jiang, 2009). As a result, the perceptual representation of objects at coincidence may interfere with each other and be suppressed. Thus, observers may perceive the objects tracing back through the previous trajectories. As expected from either of the two possibilities above, nonzero offsets between the coincidences will disrupt the stimulus configuration and its consistency with motion in depth and increase the discriminability of the objects. Hence, for nonzero offsets, the proportion of bouncing responses is drastically diminished in the no-tone condition. Investigations to understand this unimodal bouncing effect are currently underway in our laboratories.

References


Appendix A

Response type proportion in Experiment 1 and Control experiment

In Figure A1, we plotted the proportions of all types of responses for further inspection of the results in Experiment 1 and the Control experiment to test the possibility that our results can be explained by a response bias in which bounce response for the StdSBD synchronous with the tone biases response for the TstSBD toward bouncing. For tone-present trials in Experiment 1 and the
Control experiment (Figures A1b, d), the proportion of the bouncing-bouncing (BO-BO) responses were generally higher than for tone-absent trials (Figures A1a, c). The response bias account predicts that the proportion of the BO-BO responses should not change as a function of temporal offset. However, Figure A1b and d show a clear dependency of the rate of the BO-BO responses on temporal offsets. Therefore, we conclude that the present results cannot be explained only by a response bias.

Figure A1. Mean response type proportion. ST and BO correspond to streaming response and bouncing response. (a) Without a tone in Experiment 1. (b) With a tone in Experiment 1. (c) Without a tone in the Control experiment. (d) With a tone in the Control experiment.