Auditory transients do not affect visual sensitivity in discriminating between objective streaming and bouncing events

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

Jessica Ashton
School of Psychology, The University of Queensland, St. Lucia, Queensland, Australia

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

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

With few exceptions, the sound-induced bias toward bouncing characteristic of the stream/bounce effect has been demonstrated via subjective responses, leaving open the question whether perceptual factors, decisional factors, or some combination of the two underlie the illusion. We addressed this issue directly, using a novel stimulus and signal detection theory to independently characterize observers’ sensitivity ($d'$) and criterion ($c$) when discriminating between objective streaming and bouncing events in the presence or absence of a brief sound at the point of coincidence. We first confirmed that sound-induced motion reversals persist despite rendering the targets visually distinguishable by differences in texture density. Sound-induced bouncing persisted for targets differing by as many as nine just-noticeable-differences (JNDs). We then exploited this finding in our signal detection paradigm in which observers discriminated between objective streaming and bouncing events. We failed to find any difference in sensitivity ($d'$) between sound and no-sound conditions, but we did observe a significantly more liberal criterion ($c$) in the sound condition than the no-sound condition. The results suggest that the auditory-induced bias toward bouncing in this context is attributable to a sound-induced shift in criterion implicating decisional processes rather than perceptual processes determining responses to these displays.

Keywords: motion perception, stream/bounce, random element patterns, signal detection theory, multisensory perception


Introduction

Events in the world typically generate multiple forms of physical energy that, in turn, stimulate our different sense organs. When observing objects at a distance, vision and audition are the two primary senses stimulated, and the normal combination of these two inputs enriches the perceiver’s subjective experience. Indeed, filmmakers meticulously craft sound effects and surround sound environments to amplify the impact of the visual components. Another consequence of multisensory combination is that inputs in one modality can be artificially manipulated, leading to an erroneous interpretation of information in another modality. The McGurk effect (McGurk & MacDonald, 1976) is a classic example of experimentally manipulated visual inputs modulating auditory speech perception. Recently, however, striking visual distortions have been elicited from discrepant auditory information. For example, a single visual flash is perceived as two or more flashes when paired with up to four brief auditory clicks (Shams, Kamitani, & Shimojo, 2000).

The so-called stream/bounce effect (Sekuler, Sekuler, & Lau, 1997) has attracted a significant amount of attention in the multisensory research literature. One form of this illusion is a sound-induced switching of the predominant interpretation of an ambiguous motion display. In a typical display, two identical objects move from opposite sides of a 2D
display toward each other at a constant speed, meeting in the middle where they overlap (we will hereafter refer to this critical point of overlap as the point of coincidence). Subsequently, the objects continue their trajectories to the opposite sides of the display. Streaming is the dominant (>60% of trials) perception in displays containing just the moving targets and no additional transient. However, a brief tone presented at or close in time to the point of coincidence reverses this bias from streaming to bouncing. This auditory-induced bias toward bouncing has been subsequently replicated in a variety of contexts (e.g., Watanabe & Shimojo, 2001a, 2001b; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Remijn, Ito, & Nakajima, 2004; Kawachi & Gyoba, 2006; Grassi & Casco, 2009; Grove, Kawachi, & Sakurai, 2012), and generalized to visual (Burns & Zanker, 2000; Kawabe & Miura, 2006) and even tactile transients (Watanabe, 2001) temporally synchronous with or near the point of coincidence.

All of the studies that specifically investigated the influence of a brief transient on the resolution of motion sequences have employed identical visual targets. Other studies that have employed unambiguous displays did not employ transients at or near the point of coincidence (Metzger, 1953; Bertenthal, Banton, & Bradbury, 1993) or did not consider the effect of visual differences between the targets on the illusion (Kawabe & Miura, 2006).

Grove and Sakurai (2009) systematically manipulated the trajectories of the motion targets in conventional stream/bounce displays, vertically offsetting the horizontal trajectories of the two targets by varying amounts and recording the number of bounce responses in no-sound and sound trials. They replicated the noted bias toward bouncing in sound trials when the targets traced overlapping and collinear trajectories and found that this bias persisted for nonzero trajectory offsets in which the motion sequences unequivocally streamed. They observed a similar tendency when they displaced the trajectories of the targets stereoscopically in different depth planes separated by various depth intervals. These results demonstrate that the organizing strength of an auditory sound is more powerful than previously demonstrated, significantly influencing the interpretation of unambiguous visual displays.

The goal of this report is to exploit the demonstrated persistence of the stream/bounce effect in novel unambiguous motion sequences to conduct the first direct investigation of the relative contribution of perceptual and decisional factors underlying observers’ responses to these displays. Based on conventional behavioral data on the stream/bounce effect, an observer’s “bounce” response in the presence of an auditory transient can be attributed to perceptual processes, a decisional level of processing, or a combination of perceptual and decisional processes.

A standard approach to independently characterize perceptual and decisional processes in perception is signal detection theory (SDT) (Green & Swets, 1966; Macmillan & Creelman, 2005). This paradigm is appropriate for characterizing performance in conditions of uncertainty. One might present a weak signal or target event in just half of the randomly ordered trials in an experiment with the remaining trials containing nontarget events. Correctly indicating the target event occurred is termed a HIT. Failing to indicate a target event occurred when it was objectively presented is termed a MISS. Indicating a target event occurred when it was objectively absent is termed a FALSE ALARM, and correctly identifying the absence of a target event is a CORRECT REJECTION. Independent indices of the participant’s perceptual sensitivity (d’) and decisional criterion (c) are computed by comparing their transformed HIT and FALSE ALARM rates.

If sound is integrated with vision at a perceptual level, this should alter visual processing such that discriminating between objective streaming and bouncing events is degraded. In the terminology of signal detection theory, we should observe lower d’ values in the sound condition relative to the no-sound condition. On the other hand, if the sound at coincidence interacts with vision at a later stage (i.e., a decisional stage) governing the responses to stream/bounce displays, we predict changes in response criterion (c) would be observed such that observers would be more willing to report seeing a bounce in the presence of a sound than in visual-only displays. A strong prediction for the contribution of perceptual processes underlying this illusion is that sensitivity should be lower in the sound condition than the no-sound condition with no change in criterion between the two conditions. A strong prediction for a decisional account is that criterion should change as a function of sound condition with no change in sensitivity. A third possibility is that both perceptual and decisional factors underlie this phenomenon, so both sensitivity and criterion should shift as a function of sound condition.

In the two main experiments described here, we first expand on our previous findings to demonstrate that auditory-induced perceived motion reversals persist despite rendering the visual displays more consistent with streaming. We then exploit this result to investigate the effect of the presence or absence of a sound at coincidence on observers’ sensitivity and criterion, when discriminating between objective streaming and bouncing events.
Methods and stimuli

A total of 22 first-year psychology students at The University of Queensland participated. However, due to computer errors, three participants’ data had to be removed, leaving 19 participants. All participants were naive as to the hypothesis of the study.

Stimuli were generated and scripted using a Macintosh G4 computer (Operating system 9.2.2) running the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) for MATLAB. Stimuli were presented on a Samsung Syncmaster 20-inch CRT display with a refresh rate of 100 Hz. Motion sequences consisted of two dynamic random-element patches (20 rows × 20 columns) in which each element subtended 2.3 × 2.3 min arc at a viewing distance of 100 cm. The density of the patches was defined as the percentage of black elements in the square region consisting of 400 cells. For example, a 50% density patch had 200 black elements and 200 white elements randomly distributed throughout the 20 × 20 matrix, a 30% patch would have 120 black elements and 280 white elements, and so on. The distribution of black and white elements within each square patch was redrawn for each frame of the sequence. Therefore, any local motion vectors attributed to individual elements would be uncorrelated with the motion vector of the entire target patch. Texture densities in the two motion targets were presented in five conditions. In the first condition, the densities were identical (50% in both targets). In the remaining conditions, the targets differed by increments of 10% (50% in one patch, 40% in the other; 50% in one patch, 30% in the other; 50% in one patch, 20% in the other; 50% in one patch, 10% in the other).

The random-element target patches were presented on a white background and initially separated by 8.3°. They traced a horizontal path along the middle of the display. A gray fixation point was displayed on the midline of the display 0.47° below the center of the patches’ path. The patches moved across the display with each target moving 13.8 min arc in each frame. Each motion sequence consisted of 37 frames with each frame lasting 30 ms. On frame 1, the targets appeared on opposite sides of the display. From frames 2 through 18, the targets proceeded toward each other. On frame 19, the two targets visually superimposed at the center of the display, the point of coincidence. On frames 20 through 37, the targets proceeded past one another and disappeared when each had reached the position originally taken by the other target. Therefore, the global targets moved continuously across the display (they objectively streamed) in all motion sequences in this experiment. In all experiments, targets travelled at 7.4°/s. Selected frames of the motion sequence depicting two targets of different texture densities are shown in Figure 1. For half of the motion sequences, an auditory tone (800 Hz, duration 8 ms) was presented at the beginning of frame 19, the point of coincidence. Sounds were presented via loudspeakers (Audio Excel MS-120) placed on either side of the display. The highest sound pressure was 67 dB SPL at the participant’s ear. Ambient noise was approximately 43 dB SPL. Random-element density and sound conditions were presented in random order, and the directions of the different density targets were counterbalanced, allowing the lower density targets to travel from left to right the same number of times as they travelled from right to left.

Observers were instructed to hold their gaze on the fixation dot for the duration of each motion sequence and attend to the motion targets. They initiated each trial by pressing the “Space” bar on a keyboard and indicated whether the targets appeared to stream past or bounce off of one another at the end of the sequence by pressing the “S” or “B” keys, respectively. Observers were instructed to base their judgments on the motion of the global form of the targets. No reference was made in the instructions to the motions of the individual elements comprising the targets. Given the nature of our stimuli, it is possible to imagine that individual elements within the targets collided or streamed. However, none of our participants responded based on this criterion nor did they indicate that they were aware of this possibility (see also Footnote 1). Observers completed 20 observations in each condition (five texture density pairings [50/50, 50/40, 50/30, 50/20, 50/10] × two sound conditions [no tone, tone]) for a total of 200 trials.

Results

We computed the proportion of trials each observer reported bouncing in each of the 10 conditions. These figures served as the units for our statistical analyses described in the following text. The group mean proportions of reported bounces as a function of
texture density difference between the targets in the presence and absence of an auditory tone at the point of coincidence are shown in Figure 2. It is clear from the figure that auditory-induced bounce resolutions dominate when the targets had identical texture density (0% texture density difference), replicating previous studies that have employed identical targets. As the texture density differences increased, the proportion of reported bounces decreased.2 Nevertheless, the proportion of trials in which bouncing was reported was higher in the sound condition than the no-sound condition at each texture density difference. Crucially for the current study, these results demonstrate that the auditory-induced bias toward bouncing persists despite modifications rendering the targets visually distinguishable resulting in the motion sequences unequivocally consistent with streaming, generalizing Grove and Sakurai’s (2009) observations to motion sequences with dynamic random-element stimuli.

To formally explore these findings, we conducted a two (sound [absent, present]) × five (texture density difference [0%, 10%, 20%, 30%, 40%]) within participants’ analysis of variance (ANOVA) with proportion of trials for which bounces were reported as the repeated measure and found a significant effect for sound, $F_{1,18} = 24.9, p < 0.01$; a significant effect for texture density difference, $F_{4,72} = 35.4, p < 0.01$, indicating that the proportion of bounces decreased as texture density differences increased. A significant interaction between sound and texture, $F_{4,72} = 7.6, p < 0.01$, was found, reflecting the reduction in the effectiveness of the sound to promote bouncing above the no-sound condition as the targets became more visually distinct. This argues against a simple sound-induced additive bias, as pointed out by one reviewer. Paired comparisons (using the Bonferroni correction for multiple $t$-tests, $z = 0.01$) between the sound and no-sound conditions at each level of texture density differences revealed that the proportions of bounce responses were nevertheless significantly higher in the sound condition than the no-sound condition at all texture density differences (0% difference: no sound [$M = 0.34; SD = 0.34$]; sound [$M = 0.76; SD = 0.28$], $t_{18} = 4.05, p < 0.01$; 10% difference: no sound [$M = 0.2; SD = 0.24$]; sound [$M = 0.63; SD = 0.29$], $t_{18} = 4.21, p < 0.01$; 20% difference: no sound [$M = 0.08; SD = 0.2$]; sound [$M = 0.36; SD = 0.3$], $t_{18} = 3.27, p < 0.01$; 30% difference: no sound [$M = 0.05; SD = 0.17$]; sound [$M = 0.26; SD = 0.28$], $t_{18} = 3.73, p < 0.01$; 40% difference: no sound [$M = 0.05; SD = 0.2$]; sound [$M = 0.17; SD = 0.25$], $t_{18} = 3.02, p < 0.01$).

**Control experiment**

One possible explanation for the persistence of the stream/bounce effect in spite of progressive texture differences between the targets is the participants did
not detect the texture density differences. If so, then the motion sequences would have been effectively ambiguous in all conditions of Experiment 1. Therefore, we measured texture density discrimination thresholds (JNDs) for each of our observers. We employed 11 texture density differences in steps of 2% using the method of constant stimuli. Expressing the motion targets as pairs with the first number indicating the density of the rightward-moving target and the second number representing the density of the leftward-moving target, the stimulus set was comprised of the following pairs: 40, 50; 42, 50; 44, 50; 46, 50; 48, 50; 50, 50; 50, 48; 50, 46; 50, 44; 50, 42; and 50, 40. The 11 density pairs were each presented 20 times in random order for a total 220 trials. No sounds were presented in this experiment. The observers’ task was to hold their gaze on the fixation point while observing the 37-frame motion sequence and then identify the direction in which the denser target was moving. We tabulated individual responses indicating that the leftward-moving target was denser than the rightward-moving target as a function of the relative element densities and fit cumulative Gaussian curves to the data (Wichmann & Hill, 2001). Each observer’s texture density discrimination thresholds was determined to be half the percentage texture density difference that was bounded by 25% and 75% “leftward-moving target is denser” responses. On average, the JND for discriminating texture density differences was 4.5%, much smaller than the differences at which bouncing was the dominant response in sound trials. A typical psychometric function and individual JNDs are illustrated in Figures 3a and 3b, respectively. Observers were clearly able to discriminate all the nonzero texture density differences employed in Experiment 1. Therefore, the persistent dominance of auditory-induced bounce responses can be attributed to an organizing influence of the auditory signal and not insensitivity to the texture differences.

### Experiment 2

Having established in the previous experiments that auditory-induced bouncing persists for targets rendered visibly distinguishable by texture density differences, we next exploited these findings to explore the perceptual and decisional components underlying the resolution of these displays using signal detection theory.

### Methods and stimuli

Twelve new observers participated. The apparatus and general characteristics of the motion sequences were similar to Experiment 1. However, in half of the motion sequences, the targets objectively streamed. That is, the targets moved continuously across the display throughout the motion sequence. In the other half of the motion sequences, the targets objectively bounced. That is, the targets reversed their motion after coinciding in the center and returned to their original positions. Additionally, on half of the streaming trials and half the bouncing trials, a sound was presented at the point of coincidence. The remaining trials were visual only. Three sets of motion sequences were generated. In the first, the motion targets differed in texture density by 5%; in the second, they differed by 10%; and in the third, they differed by 15%. This manipulation enabled us to explore the effect of the presence or absence of sound at coincidence on the discrimination task at different levels of difficulty. We reasoned that, if discrimination was too difficult, we would observe a floor effect with no room for a degradation in \( d' \) values. Making the task too easy would result in a ceiling effect. The data from Experiment 1 and extensive pilot testing guided our choice of these three texture density differences. In all texture density conditions, the higher density was 50%. As in Experiment 1, the directions of the different density targets were counterbalanced, allowing the lower density targets to travel from left to right the same number of times as they travelled from right to left.

Participants were told that half of the motion sequences objectively streamed and the other half objectively bounced. Their task was to indicate which event occurred on each trial. Participants fixated centrally for the duration of each motion sequence. Initiation of the motion sequences and participants’
responses were triggered/recorded via the same buttons as Experiment 1. In total, each observer completed 720 trials: three texture density differences (5%, 10%, and 15%) × two objective motion sequences (streaming, bouncing) × two sound conditions (absent, present) × 60 observations each. Streaming and bouncing motion sequences, sound conditions, and the targets’ motion directions were all randomized within a block. Texture density differences were blocked with order randomly determined for each participant.

Results

We arbitrarily coded objective bounces as the “target event” to be detected. Therefore, a HIT entailed responding “bounce” to an objectively bouncing motion sequence. A MISS entailed responding “stream” to an objectively bouncing sequence. A FALSE ALARM entailed responding “bounce” to an objectively streaming sequence. A CORRECT REJECTION entailed responding “stream” to an objectively streaming sequence. For each observer, we calculated sensitivity as $d' = z(\text{HITS}) - z(\text{FALSE ALARMS})$, where $z$ is the inverse cumulative normal. We calculated criterion as $c = -0.5(z(\text{HITS}) + z(\text{FALSE ALARMS}))$. These individual scores were tabulated in both the sound and no-sound conditions and served as the units of our statistical analysis described in the following text. Group mean HIT and FALSE ALARM rates are plotted against reference isosensitivity curves for $d' = 0$, 0.5, and 2 in Figure 4. Inspection of the figure reveals that, overall, discrimination performance improved as texture density differences increased but with no differences in sensitivity between the no-sound and sound conditions at any of the three levels. Mean hit and false alarm rates in the respective no-sound and sound conditions fall very close to the same isosensitivity curves. On the other hand, mean hit and false alarm rates in the no-sound condition are shifted along the same isosensitivity curve relative to the sound condition at each texture density difference, indicating shifts in criterion. Corresponding mean sensitivity ($d'$) and criterion ($c$) values are shown in Figures 5a and 5b.

The previous observations are supported by formal statistical analyses. First, we conducted a two (sound [absent, present]) × three (texture density difference)

Figure 3. (a) A typical ogive in the texture discrimination experiment. Dashed lines indicate 25% and 75% responses. (b) Individual JNDs (% density difference) for discriminating texture density differences between the motion targets. The right-most bar represents the group mean ± 1 SEM.

Figure 4. Group mean ($n = 12$) HIT and FALSE ALARM rates. Open symbols represent trials when no sound was presented at coincidence. Filled symbols represent conditions when a sound was presented at coincidence. Squares represent 5% element density difference between targets; circles represent 10% density differences; and diamonds represent 15% density differences. For reference, dashed isosensitivity curves are drawn for $d' = 0$, $d' = 0.5$, and $d' = 2$, respectively. Error bars = ± 1 SEM.
[5%, 10%, 15%]) within participants ANOVA with sensitivity ($d'$) as the repeated measure. We found no effect for sound, $F_{1,11} = 4.6, p > 0.05$, on observers’ ability to distinguish streaming events from bouncing events. The effect of texture density difference was significant, $F_{2,22} = 37.2, p < 0.01$, indicating that discrimination performance increased as texture density differences increased. Observers’ overall ability to discriminate between streaming and bouncing events improved as the target density difference increased. On average, there was a five-fold increase in $d'$ values from texture density differences of 5% to 15%. There was no significant interaction between sound and texture density conditions, $F_{2,22} = 1.9, p > 0.05$. Therefore, we found no evidence for an auditory-induced change in sensitivity when discriminating between objectively streaming and bouncing motion sequences.

We next conducted a two (sound [absent, present]) × three (texture density difference [5%, 10%, 15%]) within participants ANOVA with criterion ($c$) as the repeated measure. We found a significant effect for sound, $F_{1,11} = 27.5, p < 0.01$, on observers’ criterion when distinguishing streaming events from bouncing events. The effect of texture density difference was not significant, $F_{2,22} = 0.9, p > 0.05$ nor was there a significant interaction between sound and texture density conditions, $F_{2,22} = 2.8, p > 0.05$.

To further explore the main effect of sound on response criterion, we conducted paired comparisons (using the Bonferroni correction for multiple $t$-tests, $z = 0.017$) of mean criterion values between no-sound and sound conditions at each level of texture density difference. We found significant differences in criterion ($c$) between the no-sound and sound conditions at each level of texture density difference (5%: no sound [$M = 0.85; SD = 0.62$]; sound [$M = -0.5; SD = 0.51$], $t_{11} = 4.6, p = 0.001$; 10%: no sound [$M = 0.65; SD = 0.45$]; sound [$M = -0.12; SD = 0.47$], $t_{11} = 3.7, p = 0.003$; 15%: no sound [$M = 0.64; SD = 0.35$]; sound [$M = -0.17; SD = 0.5$], $t_{11} = 3.7, p = 0.003$) indicating that the presence of a sound at coincidence resulted in a more liberal criterion to report bounces at every level of texture density difference between the motion targets.

We were unable to find evidence of auditory influence on observers’ ability to discriminate between objective streaming and bouncing events as mean $d'$ values were not statistically different between sound conditions at any level of texture density difference. On the other hand we did observe a significant change in criterion with observers demonstrating a more liberal criterion (more willing to give a “bounce” response) in
the sound condition than in the no-sound condition at all levels of texture density difference.

One reviewer commented that, if the sound only exerts an influence on observers’ criterion without interacting with sensitivity, the shift in criterion should be observed when sensitivity is close to zero or when it is near maximum. We tested this prediction on an additional eight participants. The experimental details were the same as those previously discussed except we employed targets that differed in texture density by 2%, 10%, or 45%. Pilot testing indicated that discriminations when the targets differed by 2% were very difficult, yielding $d'$ values close to zero, while discriminations when the targets differed by 45% were very easy, yielding $d'$ values near ceiling. We repeated measurements of the 10% difference condition for comparison.

The results of these measurements are shown in Figures 5c and 5d. Inspection of Figure 5c reveals that $d'$ values increased more than 28-fold as texture density differences increased from 2% and 45%. However, there were no systematic differences in sensitivity between no-sound and sound conditions at any of the density differences. On the other hand, inspection of Figure 5d shows that criterion values differ markedly between the no-sound and sound conditions at each of the density differences. Participants exhibited a more liberal criterion, as evidenced by more negative $c$ values, in the sound condition versus the no-sound condition at all levels of texture density difference. These observations are supported by formal statistical analyses. A two (no sound, sound) × three (2%, 10%, and 45% texture density differences) ANOVA with $d'$ as the repeated measure revealed a significant effect for texture density, $F_{2,14} = 91.5, p < 0.01$, but no significant effect of sound, $F_{1,7} = 3.4, p > 0.1$, and no significant interaction between texture density difference and sound conditions, $F_{2,14} = 2.1, p > 0.1$. A second two (no sound, sound) × three (2%, 10%, and 45% texture density differences) ANOVA with criterion ($c$) as the repeated measure revealed no significant effect for texture density difference, $F_{2,14} = 1, p > 0.3$, but a significant effect for sound, $F_{1,7} = 19.5, p < 0.01$, and a significant interaction between texture density difference and sound, $F_{2,14} = 5.3, p < 0.05$.

To further explore the main effect of sound on response criterion, we conducted paired comparisons (using the Bonferroni correction for multiple $t$-tests, $z = 0.017$) of mean criterion values between no-sound and sound conditions at each level of texture density difference. We found significant differences in criterion ($c$) between the no-sound and sound conditions at each level of texture density difference (2%: no sound [$M = 1.03; SD = 0.91$]; sound [$M = -0.78; SD = 0.65$], $t_7 = 3.5$, $p = 0.01$; 10%: no sound [$M = 0.9; SD = 0.53$]; sound [$M = -0.51; SD = 0.56$], $t_7 = 3.9$, $p = 0.006$; 45%: no sound [$M = 0.5; SD = 0.3$]; sound [$M = 0.04; SD = 0.15$], $t_{11} = 4.0, p = 0.005$). The interaction between texture density difference and sound indicates that the magnitude of the shift in criterion reduced, though remained statistically significant, as the texture density difference increased. Criterion shifts should reduce as performance approaches ceiling and few if any FALSE ALARMS or MISSES are recorded. Indeed, when sensitivity is at ceiling, criterion is zero.

In summary, the results of both sets of measurements in Experiment 2 clearly indicate a lack of influence of sound on sensitivity ($d'$) but a strong influence of sound on criterion ($c$) when discriminating between objective streaming and bouncing motion sequences, regardless of texture density difference.

**Discussion**

This report is the first to characterize perceptual and decisional factors underlying responses to audio/visual stream bounce displays using SDT. We failed to find an influence of auditory stimulation on sensitivity in this context. Mean $d'$ values were nearly identical in the sound and no-sound conditions at each level of texture density difference. On the other hand, we did observe a significant shift in criterion ($c$) toward a greater willingness to respond “bounce” in sound trials than in no-sound trials at all levels of texture density difference. The results clearly demonstrate a strong decisional component but failed to reveal a perceptual component when discriminating between objective streaming and bouncing motion sequences in the presence or absence of a transient sound at coincidence.

The persistence of reported bouncing on sound trials in spite of progressively greater texture density differences observed in Experiment 1 is surprising in light of the modality appropriateness hypothesis (e.g., Welch & Warren, 1980; Soto-Faraco, Spence, Lloyd, & Kingstone, 2004; Heron, Whitaker, & McGraw, 2004; Zhou, Wong, & Sekuler, 2007), which posits that the resolution of a multisensory display should be determined by information to the sensory modality that provides the most accurate estimate of the attribute. Vision is assumed to provide more accurate information about the location and motion of objects than audition and therefore should determine the resolution of our unambiguous displays. This clearly did not happen. Sound exerted a strong influence on observers’ responses that went well beyond simply altering ambiguous visual motion, biasing the resolution of motion displays in which the targets differed in element density by as much as 40%, a difference of approximately nine JNDs as measured in our supplementary experiment.
The results from our SDT studies in Experiment 2 are inconsistent with the data reported by Sanabria, Correa, Lupiánez, and Spence (2004). These authors inferred from their reaction time data that they measured a behavioral component of the stream/bounce effect that was less prone to higher order task-demands and possibly tapped into perceptual rather than decisional processes. Their data support the idea that there is a perceptual component underlying the mechanism responsible for the illusion. Our direct measurement of perceptual and decisional components in a discrimination task, not unlike the task employed by Sanabria et al., revealed no change in sensitivity as a function of sound which does not support the existence of a perceptual component, but does show a decisional component.

Although we did not find evidence for the influence of sound on the perceptual component of the discrimination task in our experiments, Sanabria, Spence, and Soto-Faraco (2007) reported independent contributions of perceptual and decisional processes, using SDT, in the detection and discrimination of auditory motion when accompanying visual motion was either in the same direction or the opposite direction. They reported that the audiovisual interactions were such that the uninformative visual motion information “captured” auditory motion, thus degrading detection and discrimination performance. We failed to find a similar effect of a transient auditory signal on visual discrimination performance. This could be due to an asymmetry in audiovisual interactions, noted by Sanabria et al., in motion discrimination tasks where vision exerts a greater influence on audition than the other way around. The transient sound in our displays, while powerful enough to bias responses in unambiguous displays, does not affect the perceptual processes underlying discrimination between our targets.

To our knowledge, no other studies have explicitly characterized the perceptual and decisional processes underlying observers’ responses to stream/bounce displays. However, very recently Grassi and Casco (2012) characterized observers’ ability to discriminate motion sequences in which identical disks completely overlapped at coincidence from sequences in which the disks partially overlapped in the presence or absence of a sound at coincidence. They reported differences in sensitivity ($d'$) between no-sound and sound conditions with better sensitivity in no-sound conditions. They also reported differences in criterion ($c$) with sound. A major difference between Grassi and Casco’s study and the present study is that their measurements speak to influences of sound on perceptual judgments of overlap at the point of coincidence, not on observers’ ability to distinguish between objective streaming and bouncing events. Their use of identical targets makes such discriminations impossible.

Another study by Zhou et al. (2007) is relevant. These authors investigated the efficacy of two auditory cues (timing relative to the point of coincidence, auditory intensity) and two visual cues (luminance change at coincidence, pause in motion at coincidence) as well as their combinations in influencing the resolution of ambiguous stream bounce displays. After equating the physical attributes of each cue such that, individually, they had the same influence on the resolution of the stream bounce display, they explored the influence of combinations on the resolution of the motion sequences. Their results indicate that the cues combine in a nonadditive fashion, inconsistent with an account based on probability summation, which they argue would be support for a decisional process in resolving the presence of the transients in the motion sequences. Rather, a variable weight model best accounts for their data, suggesting that the various cues have different potencies depending on what other cues they are combined with as well as individual observer differences. This is an interesting result demonstrating that various bounce inducing cues within the same modality combine at a perceptual level and then influence observers’ responses to an ambiguous stream/bounce display. However, their data do not speak to the multisensory integration of the individual or combined transients with the visual motion sequence. On the other hand, Experiment 2 of the present report speaks directly to the issue of multisensory integration between the auditory transient and the visual motion sequence. Our independent characterization of sensitivity ($d'$) and criterion ($c$) when discriminating between objective streaming and bouncing events suggest that the behavioral data consisting of reported streams and bounces is best characterized as a shift in response criterion more consistent with a decisional process rather than a change in sensitivity that would be associated with low-level perceptual integration between the auditory transient and key epochs of the visual motion sequence.

We are not aware of any other studies reporting data suggesting that only decisional factors are responsible for other audiovisual phenomena. However, there have been studies using signal detection showing the opposite finding—that audiovisual interactions are perceptually based with no decisional processes involved. For example, in their investigation of Shams et al.’s (2000) sound-induced flash illusion, Watkins, Shams, Tanaka, Haynes, and Rees (2006) reported significant changes in sensitivity ($d'$) with no accompanying change in criterion when discriminating between one or two physical flashes on a display in silence or accompanied by multiple clicks. However, McCormick and Mamassian (2008) later reported evidence showing that both perceptual and decisional processes were involved in that illusion. The results of these studies...
and the present study show that SDT is a useful approach to untangle sensory and decisional processes responding to multisensory displays.

Acknowledgments

Financial support was provided by a Japanese Grant-in-Aid of Japan Society for the Promotion of Science (JSPS) for Specially Promoted Research (no. 19001004) to Kenzo Sakurai and an Australian Research Council International Fellowship (LX0989320) to Philip M. Grove.

Commercial relationships: none.
Corresponding author: Philip M. Grove.
Email: p.grove@psy.uq.edu.au.
Address: School of Psychology, The University of Queensland, St. Lucia, Queensland, Australia.

Footnotes

1We use the terms “objectively streamed” and “objectively bounced” to describe the motions of the global targets, distinguishable by differences in element density. It is possible that, as one reviewer suggested, “accidental balanced motion signals between a pair of dots...could give room for perceiving bouncing in objective streaming events.” However, our random element patterns were refreshed on every frame of the sequence, generating local motion vectors that were uncorrelated with the motion vector of the global form making such accidental coincidences unlikely.

2Very few bounces were reported in the no-sound condition at texture density differences of 30% or 40%. We inspected the individual data from the 30% difference/no-sound condition. This revealed three observers reporting one bounce each and one observer reporting three bounces out of 20 trials. One unusual observer reported 15 bounces out of 20 trials in that condition. This same observer was the only one to report any bounces in the 40% difference/no-sound condition (18 of 20 trials). This observer was notably biased toward reporting bounces (at least 75%) in all conditions. Nevertheless, we chose to retain that participant’s data as they did not qualitatively change the outcome of our analysis. All remaining 18 observers reported zero bouncing in that condition.

References


