Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

1-1-2005

Auditory - visual interactions in the perception of a ball's path

A J. Ecker

L M. Heller
laurieheller@cmu.edu

Follow this and additional works at: http://repository.cmu.edu/psychology

Published In
PERCEPTION , 34, 1, 59- 75.

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Perception, 2005, volume 34, pages 59–75

DOI:10.1068/p5368

Auditory – visual interactions in the perception of a ball's path

Adam J Ecker, Laurie M Heller¶
Department of Cognitive and Linguistic Sciences [¶ Department of Psychology], Box 1853, Brown University, Providence, RI 02912, USA; e-mail: Laurie_Heller@brown.edu
Received 10 May 2004, in revised form 3 August 2004; published online 12 January 2005

Abstract. We carried out two experiments to measure the combined perceptual effect of visual and auditory information on the perception of a moving object's trajectory. All visual stimuli consisted of a perspective rendering of a ball moving in a three-dimensional box. Each video was paired with one of three sound conditions: silence, the sound of a ball rolling, or the sound of a ball hitting the ground. We found that the sound condition influenced whether observers were more likely to perceive the ball as rolling back in depth on the floor of the box or jumping in the frontal plane. In a second experiment we found further evidence that the reported shift in path perception reflects perceptual experience rather than a deliberate decision process. Instead of directly judging the ball's path, observers judged the ball's speed. Speed is an indirect measure of the perceived path because, as a result of the geometry of the box and the viewing angle, a rolling ball would travel a greater distance than a jumping ball in the same time interval. Observers did judge a ball paired with a rolling sound as faster than a ball paired with a jumping sound. This auditory – visual interaction provides an example of a unitary percept arising from multisensory input.

1 Introduction

Events in the world often provide both auditory and visual information about what happened. For example, in the case of a car crash, both sight and sound can indicate in different ways that glass was broken and metal was crushed. A full account of perception needs to consider not only how each sense handles this information, but also how the senses integrate information that may be either redundant, complementary, or occasionally conflicting.

Perhaps the most often cited example of auditory – visual sensory integration is the McGurk effect (McGurk and MacDonald 1976). In this cross-modal effect, a video of a person articulating speech alters the perception of speech sounds. For example, when the acoustic stimulus "ba" is paired in time with a video of someone mouthing "ga", an observer typically hears "da". The perceiver is not aware of any conflict between audition and vision but rather experiences a unified percept that integrates the information from both senses.

The ventriloquist effect is an example of an auditory – visual interaction found in the spatial domain. It occurs when a sound and visual source come from different locations (for review see Bertelson and Aschersleben 1998). Under certain conditions, the sound is not perceived as originating from its actual point of production but instead is perceived at a location closer to the visual source.

Interpretations of both the McGurk effect and the ventriloquist effect have emphasized the influence of vision on audition. Several recent studies have demonstrated the reverse: that visual perception can be modulated by auditory cues. For example, sound influences visual localization when visual acuity is degraded (Hairston et al 2003a). Conversely, visual stimuli do not alter the localization of sounds that are easily localized by the auditory system, such as broad-band noise rather than high-frequency tones (Hairston et al 2003b). This is consistent with the more general theory that the

¶ Author to whom all correspondence should be addressed.

strength of cross-modal interactions is influenced by the reliability of the individual sense information (Young et al 1993; Backus and Banks 1999; Ernst and Banks 2002). Thus, when information from the visual system is perfect and does not appear to be influenced by audition, a ceiling effect may be in operation. It may be necessary to provide less than perfect visual information and/or more potent auditory information in order to reveal the role that auditory input plays in everyday perception.

There is some evidence that sound influences vision when the visual display is inherently ambiguous rather than degraded. Sekuler et al (1997) found that the addition of a brief transient tone presented near the time of collision between two moving discs increases the likelihood of observers seeing those discs as bouncing off of one another. Without this tone, observers are likely (but not 100% certain) to see the discs as passing through each other. The explanation for this result is that the collision of objects generally produces a distinctive sudden sound, and the transient tone conveys that cue. This is a compelling experimental demonstration of sound influencing the visual component of an audio – visual event.

In a subsequent study, Sekuler and Sekuler (1999) found that *any* transient temporally aligned with the would-be collision increased the likelihood of a bounce percept. This includes a pause, a flash of light on the screen, or a sudden disappearance of the discs. A study by Vroomen and de Gelder (2000) may reveal the mechanism of a transient's ability to induce a bouncing percept. They found that a transient tone 'freezes' the display of a rapidly alternating sequence of images so that the particular frame coincident with the sound appears to remain on the screen longer. In Sekuler and Sekuler's display, perhaps the transient sound causes a visual freeze percept, and the visual freeze at the point of disc overlap causes a collision percept. Thus, the auditory – visual interaction may be due to the ability of a transient to cause visual freezing rather than the information that the sound carries about the event (ie that there was an impact).

Recent investigations by Shams et al (2002) have shown that sound can also qualitatively alter the perception of an unambiguous visual stimulus. For example, a single visual flash in the visual periphery may look like multiple flashes when paired with multiple rapid audio beeps.

It is an open question whether sounds other than transients can affect visual perception of events. While single transients can indicate the precise timing of an event (usually an impact), much more detailed information can be conveyed through auditory means. The frequency spectrum and time-varying properties of a sound can communicate complex actions such as breaking, rubbing, or scraping, as well as conveying material properties such as hardness, size, shape, or density (Vanderveer 1979; Warren and Verbrugge 1984; Freed 1990; Li et al 1991; Gaver 1993; Lutfi and Oh 1997). For example, one object scraping across a rough surface involves rapid random impacts against the microstructure of the surface texture; a resultant band-limited noise conveys this scraping action (Gaver 1993). Here we explore whether a rich auditory stimulus conveying information about rolling surface contact would influence the perception of the motion of a ball, regardless of whether or not the stimulus contained a transient sound.

Given that transient sounds have been found to influence vision under some conditions, we ask whether other types of sounds are also effective and whether their effectiveness varies across a range of visual-cue strengths. That is, are there cases where auditory cues function on par with visual cues in the perception of events within a rich and complex visual and auditory environment?

Our paradigm was inspired by the visual displays of Kersten et al (1997) in which the perceived three-dimensional (3-D) trajectory of a ball depends on the placement of a cast shadow. A ball traveling a single path can be made to look like it is either rising in the frontal plane or rolling back in depth on the floor of the box by adding a cast shadow appropriate to such motion. Within this scene there are competing visual

cues to depth that allow for a shadow to be a disambiguating cue. In general, objects moving in depth on a ground plane towards the horizon appear higher in the visual field (Gibson 1950), so when the ball moves vertically on the screen this may indicate movement in depth. In contrast, the ball's constant size is a cue that competes against motion in depth and is consistent with the ball rising and falling in the frontal plane. The visual ambiguity in Kersten et al's paradigm was particularly appropriate for use with sound because the alternative events (rising or rolling) would produce two very different sounds. This provides the opportunity to test whether a similar visual shift can be accomplished with the addition of sound rather than a cast shadow. In experiment 1 we hypothesize that a sound that indicates rolling surface contact will cause the ball to appear to roll back in depth along the floor, whereas a sound that indicates a single impact with the ground will cause the ball to appear to jump into the air in the frontal plane. The power of these cues is investigated by pairing them with parametrically varied visual displays to cover a wide range of relative visual ambiguity (and hence presumed visual potency). In experiment 1a we used sound as a disambiguating cue to the ball's path. The lack of a size change in the ball weights the visual cues in favor of a jump interpretation, which better tests the effectiveness of the roll sound as an additional cue to depth. In experiment 1b we incorporated size changes to the ball consistent with it rolling back in depth. This brings in cues that favor a rolling perception, which better tests the effectiveness of the jump sound. An additional manipulation of the ambiguity of the ball's visual cues was achieved by varying the curvature of the ball's path. We predicted that sound would be a more effective cue with the more ambiguous visual stimuli.

2 Experiment 1a

2.1 *Methods*

2.1.1 *Subjects.* Twenty-three paid observers participated. All had normal or corrected-to-normal vision and no known hearing loss. One observer's data were removed because of a computer error, leaving twenty-two observers for analysis.

2.1.2 *Apparatus.* A PC with a Windows operating system controlled the presentation of stimuli via Media Player extensions on an NEC 38 cm color monitor with 640×480 pixel resolution. Stimuli appeared within a 320×240 pixel window. Each observer viewed the display from approximately 70 cm while seated in a sound-attenuating listening booth. Because a pilot study had indicated that the method of sound presentation—either headphones or speakers—did not affect the strength of the auditory cue, acoustic stimuli were presented diotically over Sennheiser HD600 headphones at a peak level of 56 dBA.

Displays: Video stimuli. All visual stimuli consisted of a ball moving in a 3-D 'box' adapted from the 'ball-in-a-box' paradigm (Kersten et al 1997). The scene contained a checkerboard floor with walls at the back, left, and right sides, and a ball that never changed in size as it moved within the box. Computer animation sequences were created in 3D Studio Max 5.1, rendered at a resolution of 320×240 pixels and displayed at 30 frames s^{-1}. All stimulus videos began with a stationary ball presented for 20 frames followed by either 40 or 80 frames of motion along the specified path (depending on path type; see below), and ended with the ball holding its resting position for 20 frames. The path was always symmetric about the center of the screen. All displays were viewed from a 22° elevation.

Displays: Auditory stimuli. Critically, each video was paired with one of three audio conditions: silence, roll-audio, and jump-audio. The sound for the roll-audio condition was recorded from a marble rolling across a rough plastic surface, impacting with a wood barrier, and rolling back in the opposite direction. The audio for the jump-audio

condition consisted of a recording of a basketball bounce. Digital audio recordings were performed with an Audio Technica 30 condenser microphone in an acoustic-foam-walled studio booth in order to reduce echoes. In the case of the rolling sound, the microphone was placed approximately 25 cm above the mid-point between the starting position of the marble and the wood barrier. This placement was a compromise between placing the microphone at the viewer's location and achieving an optimal signal-to-noise ratio throughout the recording. In the case of the jump-audio sound, the microphone was placed approximately 1 m away from the location of the ball's impact with the floor, with the microphone held approximately 25 cm above the floor. Sounds were integrated with the video displays by means of video editing software.

Displays: Path types and path curvature. For each audio condition, the ball traveled along one of four general path types: 'arc', 'timing', 'flanked', and 'no-transient'. Figure 1a shows the arc condition in which a ball started in the lower left-hand portion of the screen and moved at a constant horizontal velocity to the right while simultaneously moving upward until it reached the center of the box, at which point it traveled back down to the lower right and stopped (other path types will be described in detail later). From start to finish, the ball traveled 4 deg of visual angle in the horizontal axis and 2.1 deg in the vertical axis. Considering only visual cues, the arc path can be made to look more like a jump or more like a roll by adjusting the shape of the path. The curvature of the path could be parametrically varied between two endpoints: (i) perfectly parabolic, forming an inverted U (very jump-like), or (ii) linear segments, forming an inverted V, by moving straight to the back wall of the box and rebounding in the opposite direction (very roll-like). These two paths have the same starting point, mid-point, and endpoint. The jumping ball decelerated at a constant rate, reaching a vertical velocity of zero at the peak of the arc. The rolling ball kept a constant vertical velocity.

Path curvature was parametrically varied in seven steps by taking a weighted average of the parabolic and linear paths. These seven path curvatures are defined as the percentage of curvature consistent with a parabolic jump path: 0%, 20%, 33%, 50%,

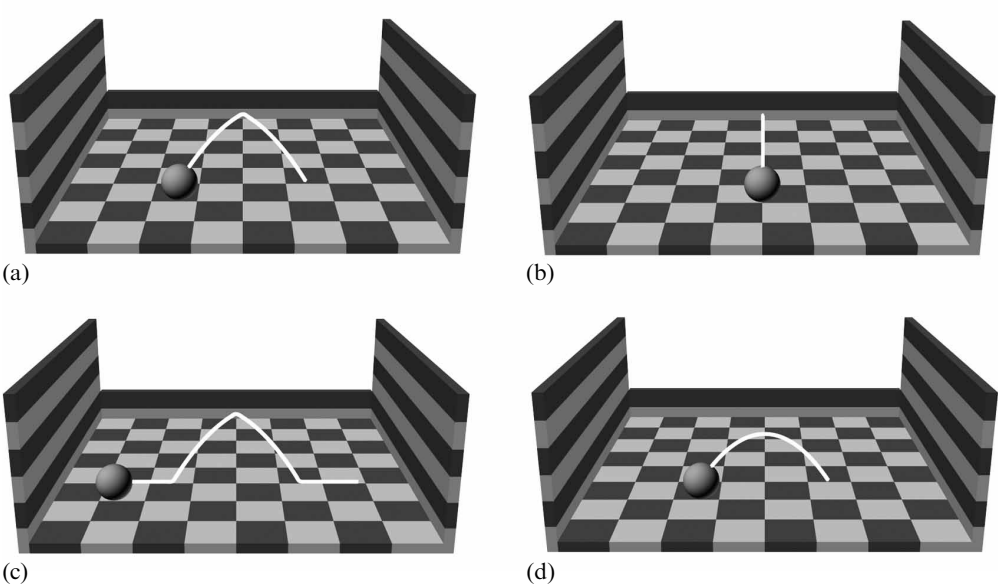


Figure 1. Illustration of the four path types: (a) arc, (b) timing, (c) flanked, and (d) no-transient. Example curvature is 50% in the first three paths; the parabola height is 60% in the no-transient path.

66%, 80%, and 100% (figure 2 illustrates 0%, 50%, and 100%). This means that 100% curvature is a perfectly parabolic path consistent with a jumping motion, and 0% represents two perfectly linear path segments consistent with a rolling motion in depth. At 33% curvature, for example, at any given video frame the ball is located one third of the way from the linear path towards the parabolic path.

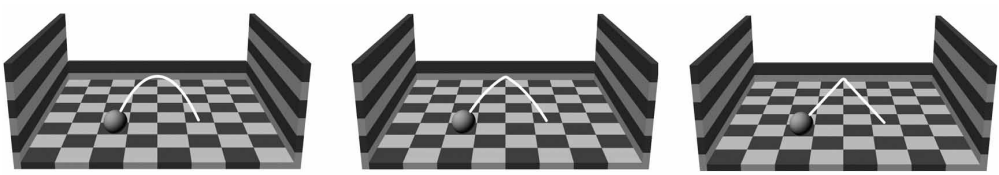


Figure 2. The effect of the curvature parameter: a parabolic-to-linear continuum illustrated by the two endpoints and mid-point of the seven steps. A path with 100% parabolic curvature appears the most like a jump (far left).

In the roll-audio condition of the arc path type (figure 1a), for each video in the continuum, a roll sound was played throughout its duration of 1.33 s and the impact sound was aligned with the moment that the ball hit the back wall. In the jump-audio condition, there was no sound until an impact sound began at the frame in which the ball returned to its original vertical position. That is, when all path curvatures were paired with the roll sound they comprised the roll-audio condition, whereas when they were paired with the jump sound they formed the jump-audio condition.

In contrast to the arc path, the timing path (figure 1b) collapses across the horizontal dimension so that the ball simply moves up and down in a vertical line. Although the timing of the movement in the vertical axis is varied from parabolic to linear, the on-screen shape of the path does not change. This path type is intended to tease apart the parametric effects of varying the timing of motion along the path as opposed to varying the path shape and timing.

The flanked path (figure 1c) shows the ball rolling horizontally prior to and after the arc-shaped path. Thus in the jump-audio condition, the video begins with a rolling sound, has a gap of silence while the ball is in the air, has a bounce sound when the ball reaches the ground, and continues with a rolling sound. In the roll-audio condition there is a continuous roll sound with an impact coinciding with the ball hitting the back wall. The rationale for this path is that the rolling sounds prior to the silence during the jump-audio condition will allow the silence to signal that the ball has left the ground. The total duration of this video was 2.66 s on account of the extra rolling segments.

The no-transient path (figure 1d) consists of parabolic arcs of varying heights. At one extreme (100% curvature, parabolic motion), the no-transient path is identical to the arc path. At the other extreme, the ball would travel in a horizontal line and would obviously roll in the frontal plane. Parametric variation of the height of an arc included five parabolic arcs (a horizontal line was not included since it was unambiguously a roll). The jump sound was identical to that used in the jump-audio condition of the arc path type. In the roll-audio condition, the transient was digitally edited out of the roll sound, resulting in a continuous rolling sound throughout the video. The rationale for this manipulation is to test whether the roll sound can have an influence without the presence of the transient that occurs in the middle of the arc path. The reason for using parabolas exclusively (rather than mixing with a linear roll path) was so that the ball's smooth reversal of direction at the peak of the arc would appear plausible even without a collision (as indicated by a transient sound), perhaps due to spin on the ball or a tilted surface.

2.1.3 *Design.* The three audio conditions (roll-audio, jump-audio, silence) were crossed with three of the visual path types (arc, timing, flanked) and each of their seven parametric variations in curvature along the parabolic-to-linear continuum. The three audio conditions were also crossed with the no-transient path and each of its five parametric variations in parabola height. This resulted in 78 audio – video trials, presented to observers in random order.

2.1.4 *Procedure.* Observers were given written instructions to "make a judgment about a ball and the path it travels" and to "indicate what event you perceive". They were told that they would be asked if the ball rolled on the floor or jumped into the air. Anticipating the fact that there is always a rolling segment within the flanked video, observers were forewarned that if the ball changes direction they should make the roll/jump judgment when the ball "first changes its direction". Each video was repeated three times in sequence, separated by two black video frames. Observers were instructed to respond according to how the ball's path appeared at the last of the three repetitions. Observers used the mouse to select a button labeled "jump" or "roll", and then pressed another button to proceed to the next video. At the end of the experiment observers were given a follow-up questionnaire that asked if they saw the ball take any path other than the orthogonal roll and jump paths.

2.2 *Results*

The results from the arc, timing, and flanked paths (figure 3) are displayed as observers' average percentage of "roll" responses as a function of the parametric variation in curvature. The results from each path type are displayed separately in figures 3b – 3d and the average across all path types is displayed in figure 3a. The far left of the x axis represents a parabolic shape consistent with a jump (100% curvature) and the far right represents an inverted V shape with linear motion in depth consistent with a roll (0% curvature). The three lines connected by filled circles, open squares, and filled triangles show the results for roll-audio, silence, and jump-audio, respectively. The separation between the lines illustrates that more observers judge the ball as rolling when paired with a rolling sound than when the same ball is paired with a jumping sound. Averaging across the arc, timing, and flanked path types and all variations in curvature, rolls were seen 78% of the time in the roll-audio condition versus 20% of the time in the jump-audio condition (figure 3e). The results for the silent condition fall in-between the two sound conditions, with 38% of the silent videos seen as rolls. (Analysis of the no-transient path is presented separately below.)

An omnibus ANOVA with factors for sound (roll-audio, jump-audio, silence), path type (arc, timing, flanked), and curvature (percent parabolic versus linear) revealed a significant main effect of sound ($F_{2,42} = 65$, $p < 0.001$) and curvature ($F_{6,126} = 14$, $p < 0.001$) but not path type ($F < 1$), and a significant interaction between curvature and sound ($F_{12,252} = 3$, $p < 0.005$). Although both the jump and roll sounds had effects relative to silence (Scheffé planned comparisons, $F_{1,21} = 58$, $p < 0.001$ for roll-audio versus silence, $F_{1,21} = 39$, $p < 0.001$ for jump-audio versus silence), the effect of the roll sound was significantly larger than the effect of the jump sound ($F_{1,21} = 17$, $p < 0.01$ as a Scheffé a posteriori comparison between the magnitudes of roll-audio minus silence and jump-audio minus silence).

Because the three path types did not differ significantly in their tendency to elicit a roll or jump percept and there was no significant interaction with the effect of sound (figures 3b – 3d), subsequent analyses combined across these path types (figure 3a).

The effect of the curvature parameter was examined in more detail with statistical trend analyses. As path curvature varies from parabolic to linear, the percentage of "roll" responses increases linearly ($F_{1,126} = 80$, $p < 0.001$) with no significant quadratic or cubic trends. The functions for the roll and jump sound differ from each other by

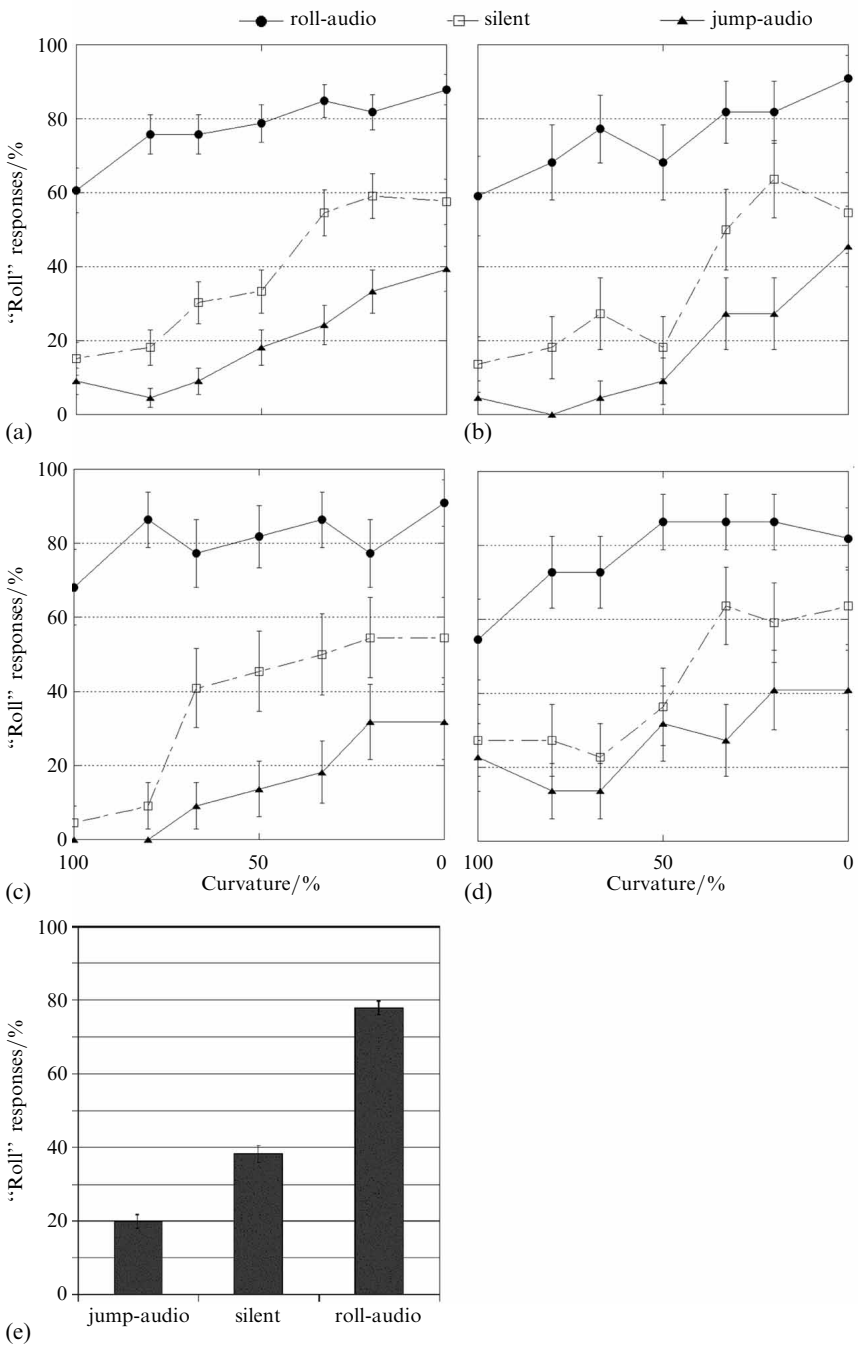


Figure 3. Results from experiment 1a in which size of the ball remained constant. Panels (b) – (d) display results for the arc path, the timing path, and the flanked path, respectively; panel (a) displays the average of these. The *y* axis shows the percentage of subjects who indicated that the ball appeared to roll back in depth as opposed to jumping in the frontal plane. The *x* axis shows the curvature of the path as indicated by the percentage that is contributed from a parabolic path. A 100% parabolic path is consistent with a jumping ball, whereas the 0% curvature is a linear path that is consistent with a ball rolling back in depth. The filled circles, open squares, and filled triangles correspond to the roll-audio, silent, and jump-audio conditions, respectively. The bars in panel (e) show the average effect of sound taken across all path curvatures and all path types. Error bars show standard error of the mean in this and all other figures.

roughly 50% of responses throughout the entire range of curvature [with slopes that differed marginally in a linear contrast ($F_{1,252} = 3$, $p = 0.065$)]. This means that sound had roughly the same magnitude of effect regardless of whether it was consonant with or in opposition to the favored interpretation of the visual display. Nonetheless it was predicted that there would be a greater effect of sound where the path curvature is most ambiguous (ie 50% curvature). Although such an effect was not immediately apparent in the graph, in addition to the linear component there was a small but statistically significant quadratic component in a contrast between the jump and roll functions ($F_{1,252} = 5$, $p < 0.05$).

The no-transient path type. The percentage of "roll" responses obtained with the no-transient path is shown in figure 4. An ANOVA that included factors of sound and parabola height (analogous to curvature in the other path types) showed that sound had a significant effect ($F_{2,42} = 43$, $p < 0.001$) as well as parabola height ($F_{4,84} = 3$, $p < 0.05$). Observers were more likely to perceive the ball as either jumping or rolling depending on the sound type. The difference in effect of roll-audio versus jump-audio was 57% on average across all curvature variations. It is possible to directly compare one common data point in the roll-audio condition between the no-transient and arc paths: when the curvature was 100% these two conditions had identical visual stimuli and the audio differed only in the removal of the transient at the peak of the arc. At this data point, the no-transient path has 50% roll responses and the arc path has 59% roll responses, but this difference was not significant ($F < 1$). Note that the results at 100% curvature for the jump-audio and silent conditions were already shown in figure 3b (arc path) and were repeated in figure 4 because the stimuli are identical.

Phenomenology. A single 'catch trial' at the end of the experiment contained a cast shadow consistent with a jump path and auditory stimuli consistent with a roll path. The ball took a fully parabolic arc path. Twenty out of twenty-two observers responded that this ball was jumping, which indicates that they did not feel compelled to respond according to the sound when the visual cues were in strong opposition. The data show that two of the twenty-two observers responded that all trials contained a 'jumping' ball regardless of sound and curvature cues. Their data were included in the analysis. In the post-experiment questionnaire, all observers reported that they saw the ball either jump in the frontoparallel plane or roll to the back wall along the floor, but never saw a path in-between.

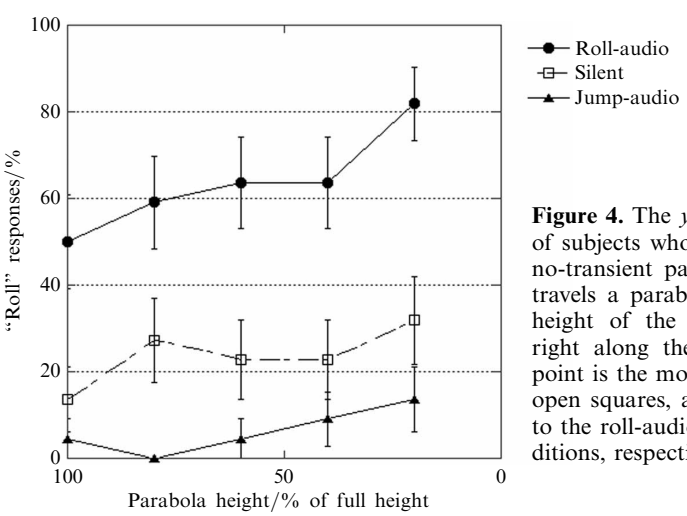


Figure 4. The y axis indicates the percentage of subjects who see the ball as rolling in the no-transient path of experiment 1a. The ball travels a parabolic path in all cases but the height of the path decreases from left to right along the x axis so that the leftmost point is the most jump-like. The filled circles, open squares, and filled triangles correspond to the roll-audio, silent, and jump-audio conditions, respectively.

3 Experiment 1b

In experiment 1b the size of the ball changed in a manner consistent with it moving in depth and therefore provided a visual cue consistent with the rolling interpretation. The degree of size change varied as well. The aim of this manipulation was to see whether it was possible to obtain an effect of sound when the visual display contained more cues in favor of a roll than in experiment 1a.

3.1 *Methods*

Experiment 1b was conducted as a continuation of experiment 1a after a 2 min pause. No additional instructions were given. The same twenty-two observers participated.

In experiment 1b, the 0% curvature path included a physically appropriate size change for the change in depth that would occur if the ball rolled to the back wall. This constitutes a strong depth cue, as the radius of the ball shrinks to about 77% of its original size at the apex of the inverted V of the path along the floor. In addition, the ball's velocity was kept nearly constant in 3-D so that, in its perspective projection onto 2-D, its vertical velocity decreased towards the apex. The intermediate paths were weighted averages of both the curvature and ball size in these two endpoints (in proportions consistent with experiment 1a). There was no 100% curvature condition because it would not have had a size change. The arc, timing, and flanked paths were used.

3.2 *Results*

On averaging across all three path types, 82% of balls were seen as "rolls" in the roll-audio condition and only 41% of balls were seen as "rolls" in the jump-audio condition (that is, 59% of balls were seen as "jumps" in the jump-audio condition). 55% of the silent videos were seen as "rolls". Figure 5 shows the average proportion of "roll" responses.

An omnibus ANOVA with factors for sound, curvature, and path type revealed a significant main effect of sound ($F_{2,42} = 31$, $p < 0.001$), curvature ($F_{5,105} = 55$, $p < 0.001$), and path type ($F_{4,42} = 4$, $p < 0.029$), and an interaction between curvature and sound ($F_{10,210} = 6$, $p < 0.001$).

Comparison of experiments 1a and 1b. An ANOVA comparing experiments 1a and 1b revealed that a ball that changed in size produced significantly more "roll" responses ($F_{1,21} = 12$, $p < 0.005$), showed a smaller effect of sound ($F_{2,42} = 6$, $p < 0.01$), and showed a greater effect of path curvature ($F_{5,105} = 9$, $p < 0.001$) than a ball of constant size. These effects appear to reflect the greater tendency to respond "roll" in the jump-audio and silent conditions, specifically when the path curvature is lower and thus the size change is greatest (ie visual cues are more consistent with a roll).

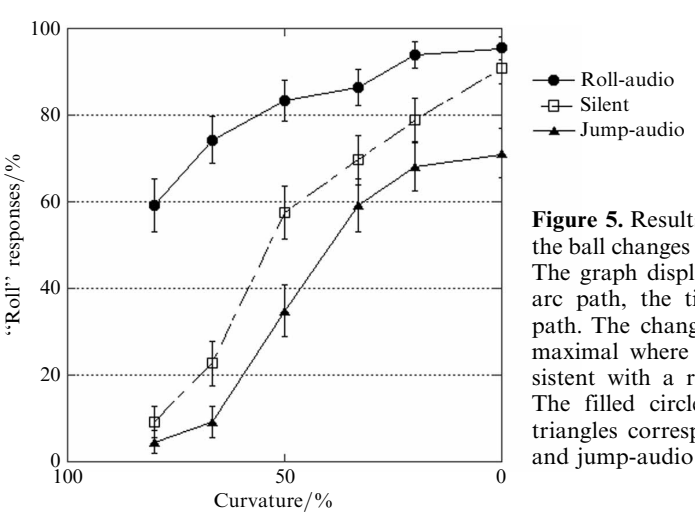


Figure 5. Results from experiment 1b in which the ball changes size as it moves back in depth. The graph displays the average results of the arc path, the timing path, and the flanked path. The change of the size of the ball was maximal where its curvature was most consistent with a roll (0% parabolic curvature). The filled circles, open squares, and filled triangles correspond to the roll-audio, silent, and jump-audio conditions.

Phenomenology. Five of the twenty-two observers in the size-change experiment reported seeing the ball move both back in depth and also above the ground on "a few", "some", or "three" trials. Four of the observers said they had responded that such an in-between path was a "jump", and one observer called such a path a "roll".

3.3 *Discussion*

In experiment 1a visual cues (path shape and ball size change) were manipulated relative to auditory path cues and the combined perceptual effect was measured with a single metric. In much the same way that Kersten et al (1997) showed that a shadow cue could disambiguate the 3-D path of a ball, so too can the appropriate audio cue. We postulate that sound interacts with visual cues in this manner because the ultimate perception of an event is influenced by all available sensory cues.

The most striking result of experiment 1 is the large size of the effect of sound, regardless of the ambiguity of the visual display. On average, observers' responses agree with the sound cue on 79% of trials. Even with a path curvature in which the ball looks like it is jumping 85% of the time in silence, a rolling sound causes 65% of the observers to see a rolling ball. Thus, sound is effective even when the silent visual display strongly favors one percept.

The addition of a size-change cue and vertical velocity consistent with motion in depth increased the tendency to see the ball as rolling in depth, especially for the path shapes most consistent with a roll. This outcome is to be expected because the addition of a size change is a visual depth cue that indicates rolling as opposed to jumping. Nonetheless, the auditory cues still altered perception in both directions relative to silence, which shows that the basic effect is robust across differing displays and visual cues.

We now consider the implications of each of the four different path types. Unlike previous studies in which auditory information interacts with visual cues, these results do not rely on a transient sound. Results from the no-transient path show that, even when the ball does not apparently collide with the back wall to produce a transient sound, the perceptual effect of the rolling sound clearly remains. It is the acoustic *content* of the audio cue that alters motion perception, in that the meaning of the environmental sound is incorporated into the final perception of the scene.

Because the results from the timing path were similar to the results from the other paths, it is evident that observers were sensitive to the temporal information provided in the ball's vertical motion. Because the flanking path did not produce substantially different responses from the arc path, it is evident that the rolling sounds flanking the silent portion of the jump did not increase the tendency to see a jumping ball. It is possible that the arc path was as effective as the flanked path at producing a jump percept because the videos repeated three times on each trial. Thus the collision sound at the end of the first viewing may give meaning to the silence at the beginning of the subsequent viewings with regard to the ball's path. This gave the opportunity for a change in the perception of the balls' path by the third viewing.

The parametric variation of the visual cues provided us with the opportunity to measure the effect of auditory cues against a range of visual-cue ambiguity and certainty. We find it remarkable that sound has a sizeable effect over most of the range. Thus, it appears that sound is able to either oppose or reinforce the predominant visual interpretation of an event. Note that, if this experiment had only paired the sounds with the visual displays containing the strongest cues towards rolling (the linear path shape with a size change), one might have erroneously concluded that auditory perception has little or no effect.

Although the data seem to indicate that observers' *perception* of the ball as either rolling in depth or jumping in the frontal plane was influenced by auditory information, let us consider whether the results of experiment 1 represent a genuine perceptual

phenomenon, or a post-perceptual decision (ie a conscious strategy). The fact that observers are willing to call an event a jump when the audio cue indicates a roll (or vice versa) shows that they are not exclusively using one cue or the other as a result of task demands. The fact that the number of "roll" responses increases in the expected direction with changes in path curvature regardless of audio condition indicates that observers are basing their responses on a combination of the audio and visual cues. Although these data are consistent with a perceptual phenomenon, they do not decisively answer the question whether it is perceptual or post-perceptual because observers were directly asked what path the ball took. Experiment 2 was designed to address this question by having observers judge a secondary aspect of the display.

4 Experiment 2

In experiment 2, observers judged the speed of the ball rather than directly indicating its path. In these displays the length of the rolling path is longer than the length of the jumping path. Because the ball travels these paths in the same period of time, a rolling ball would be moving faster in 3-D space than a jumping ball. If a ball is judged as faster when paired with a rolling sound than when that same video is paired with a jumping sound, we can infer that the observer indeed perceives a difference in the paths of the two otherwise identical ball motions.

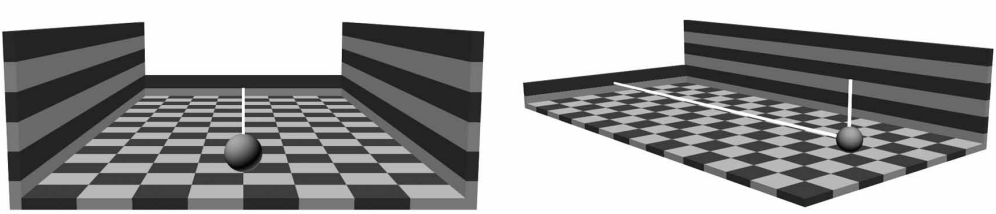
Observers were presented with a pair of videos and asked to select the one containing the faster-moving ball "as it would be moving in 3-dimensional space". In order to prevent response bias, we included many additional trials other than ones in which rolling balls were the fastest so that a balanced number of trials also had faster jumping balls. The experiment also included trials in which the ball's path was unambiguously defined by a moving shadow instead of a sound as a control condition.

4.1 *Methods*

4.1.1 *Subjects.* Fourteen volunteers participated for course credit at Brown University. Data from one observer were removed because of experimenter error. None of these observers participated in experiment 1.

4.1.2 *Apparatus.* The apparatus was the same as in experiment 1.

4.1.3 *Displays.* Video stimuli: New computer animation sequences were created for experiment 2. To exaggerate the speed difference between the jump and roll paths, the roll path was lengthened by increasing the depth of the box by a factor of three. The viewing angle was adjusted to a 10° elevation so that the path and distance that the ball moves *on-screen* is identical to the on-screen motion of experiment 1. Figure 6a shows the scene as viewed by observers and figure 6b illustrates the difference between the jump and roll paths by rotating the scene to one side.



(a) (b)

Figure 6. The box in experiment 2 was extended in depth to exaggerate the differences in speed between a roll path and a jump path. The on-screen motion of the ball was identical to the timing path in experiment 1a and the ball size remained constant. Panel (a) shows the scene as viewed by observers and panel (b) illustrates the difference between the jump and roll paths by rotating the scene to one side.

The video in this experiment contained a ball moving vertically along the timing path. This path was chosen because it simplifies the motion down to one dimension. The parametric weights of the parabolic-to-linear curvature continuum were set at 50%. As in experiment 1a, the ball remained the same size throughout its motion. The vertical velocity of the ball decreased slightly along its path towards its apex since it was an equal mixture of the constant speed of the linear path and the deceleration of the parabolic path.

This visual display was made into two different videos by adding the sounds from either jump-audio or roll-audio conditions of experiment 1 to define one of two paths. At the same on-screen speed, balls perceived as rolling are predicted to be seen as faster than balls perceived as jumping because of the difference in path length. It was important to create additional trials other than ones in which rolling balls were the fastest to prevent a response bias. This was accomplished by varying the overall speed of the videos so that sometimes the jumping balls had the fastest 2-D on-screen speed. Each ball could travel at one of three on-screen speeds. The medium speed was the same speed as in experiment 1, the slow speed was 15% slower, and the fast speed was 15% faster.[1] This resulted in jumping balls being objectively fastest on one-third of the trials, equal on one-third, and slowest on one-third.

To match the rolling audio from experiment 1 to the timing of these new videos, the compress/expand speech algorithm in Sonic Foundry's Sound Forge 6.0 (http://mediasoftware.sonypictures.com/) was used. This maintained the pitch of the rolling sound regardless of the video condition, but it lengthened or shortened the sound to be the same duration as the visual display. The collision sound that occurred when the ball hit the back wall was not altered and was inserted at the appropriate time in the rolling sound by digital sound editing.

In addition to using auditory cues to define the ball's path, a complementary set of videos was constructed with visual cues in the form of a cast shadow. A jumping path was defined by a cast shadow that started out directly underneath the ball and remained stationary while the ball traveled up and down. A rolling path was defined by a cast shadow that traveled underneath the ball throughout its motion in the display. The two possible shadow cues and the three different display speeds yielded six additional video conditions. All videos containing shadows were silent.

4.1.4 *Design.* All permutations of path cue (jump/roll), cue type (audio/shadow), and display speed (slow/medium/fast) resulted in twelve videos. Each trial presented two videos successively for comparison, in random counterbalanced order, requiring 144 trials to be viewed once by each observer. Trials that included shadows were included for comparison, because cast shadow is a well-established cue to path. We were primarily interested in speed judgments on *trials in which one ball appears to jump and the other ball appears to roll, especially when viewed at the same on-screen speed.* These trials we shall call "critical trials". The fact that balls with the same on-screen speed were being compared in some trials was obscured by the fact that the speed discrimination was reasonably difficult in all trials (achieved by limiting the range over which the speed varied to $\pm 15\%$).

4.1.5 *Procedure.* Observers made a forced choice between two alternating videos. To aid with response selection, the first video always contained a blue ball and the second contained a green ball. The videos were separated by a 67 ms pause created by two black frames and repeated in the same order, such that observers saw a blue, a green, a blue, and a green ball. After viewing these videos the observers indicated which

[1] The 15% speed difference was chosen because pilot results determined that a 25% speed differential was trivially easy to discriminate whereas a 10% differential was quite difficult.

color of ball moved fastest relative to the other by using a computer mouse to select a button labeled "Blue" or "Green". The next trial began when the observer indicated readiness.

Instructions. Observers were instructed to judge the ball's average speed "as it would be moving in 3-dimensional space. The ball may slow down when it changes directions and that's okay—we mean the speed when it's moving". The instructions avoided the terms "path", "jump", or "roll".

The instructions also stated: "Pay attention to everything you see and hear, and make your judgments based on your *perceptions.* Don't try to 'out-think' the experiment: the 'right' answer is what you perceive". These instructions were included in response to pilot testing in which observers revealed that they had actively attempted to ignore the sound. They had mistakenly thought their task was to judge the on-screen speed as accurately as possible and felt that the sound interfered with this task. The final version of the instructions eliminated this problem for all but one subject.

Pretest report and post-test questionnaire. Before the experimental trials began, observers saw a demonstration of two silent videos, one of which contained a shadow cue for jumping, and the other of which contained a shadow cue for rolling. Observers were asked to "describe the motion of the ball in this video". This pretest was designed to identify those observers who might not perceive the ball's intended path of motion in depth. This pretest was undertaken because we learned from experiment 1 that some individuals never see the ball as rolling back in depth, as evidenced by the fact that they labeled every video/audio combination as a "jump". Observers who do not experience the ball moving in depth would not provide any data to confirm or deny the hypothesis that a difference in perceived speed might result from a difference in path perception. The pretest provided an objective way to separate out these individuals because it did not rely on examining their data a posteriori nor did it use auditory stimuli. After completion of the experiment, observers were verbally asked a systematic series of questions about their perceptions.

Replication of experiment 1. After all videos were viewed and judged, a brief replication of experiment 1 was conducted. The aim was to find out whether the results of experiment 1 would replicate despite the deeper box of experiment 2, varying display speeds, and different methodology of alternating videos. Observers saw the subset of trials from experiment 2 in which one ball was paired with a roll sound and the other a jump sound at the same on-screen speed, and they were now asked whether the first ball rolled or jumped. The videos were viewed at three different display speeds on separate trials, played in random order. After this replication, another 6 trials consisted of a replication with the same single-interval methodology as in experiment 1.

4.2 *Results*

Data from all thirteen observers showed a significant effect of sound in the predicted direction (see below) but primary analyses discarded data from three of the observers. Two observers failed the pretest because they never saw the ball move back in depth when accompanied by an appropriate shadow cue. One observer did not follow instructions: she actively ignored the audio and counted to herself to time the events. Note that, while experiment 1 addressed the entire sample population, experiment 2 addressed the hypothesis that if an observer sees a difference in path then he/she should see a difference in speed. The magnitude of this effect is presumably most accurately represented by the use of data from the ten observers who saw path differences and followed instructions.

Figure 7 represents results of the subset of 'critical trials' that compare videos in which the on-screen speed is identical and only the sound or shadow path cue differs. These results are from the final group of ten observers. The vertical axis shows the

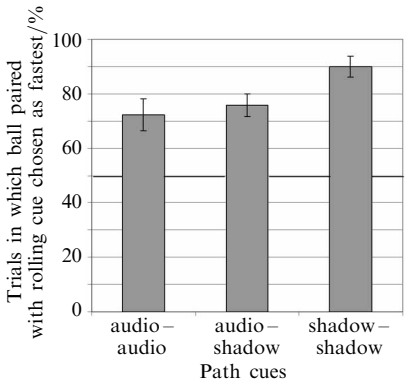


Figure 7. The main result of experiment 2 is from the subset of 'critical trials' where the on-screen speed of the two balls is identical but the path cues they are paired with differ (ie one ball's sound or shadow cue indicates a jump and the other ball's sound or shadow cue indicates a roll). The y axis shows the percentage of subjects who pick the ball with the role cue as faster than the ball with the jumping cue. Separated along the x axis are the three different combinations of path cues that could occur within two intervals of a critical trial. Chance performance would be at 50% as indicated by the thickened line.

percentage of observers who picked the ball with the roll cue (either audio or shadow) as faster than the ball with the jump cue (either audio or shadow). Separated along the horizontal axis are the three different combinations of path cues that can occur within a critical trial. In the most important condition, a ball accompanied by a roll sound can be compared to a ball accompanied by a jump sound ('audio–audio'). Observers chose the ball with the roll sound as faster than the ball with the jump sound, *at the same on-screen speed*, 72% of the time in the audio–audio condition. In the other conditions, a roll sound can be compared to a jump shadow or a roll shadow can be compared to a jump sound ('audio–shadow'), or a roll shadow can be compared to a jump shadow ('shadow–shadow'). The ball with the roll cue was chosen as faster 76% and 90% of the time in audio–shadow and shadow–shadow conditions, respectively.

Considering only the subset of trials in which a roll cue was compared to a jump cue, an ANOVA coded for type of path cue (audio–audio, audio–shadow, shadow–shadow), order of presentation within a trial (the jump was either first or second), and speed of video (fast, medium, and slow). The proportion of trials in which a roll was seen as faster than a jump was significantly greater than a chance level of 50% ($F_{1,9} = 41$, $p < 0.001$) and the type of path cue had a significant effect ($F_{2,18} = 8$, $p < 0.005$). No other main effects or interactions involving within-trial order of presentation and within-trial speed of video were significant. Examining the most important condition separately, the mean response of 72% rolls in the audio–audio condition was significantly above a chance level of 50% ($F_{1,9} = 11$, $p < 0.01$). [A parallel analysis that included all thirteen original observers also found the main effect of sound to be significant ($F_{1,12} = 5$, $p < 0.05$).] The shadow–shadow condition had significantly more rolls chosen as faster than did either the audio–audio condition ($t_9 = 3.2$, $p < 0.05$, a posteriori) or the audio–shadow condition ($t_9 = 4.3$, $p < 0.01$, a posteriori).

The critical trials showed that sound had an effect on speed judgments when on-screen speed was identical. However, retinal velocity is a powerful cue to speed perception (McKee and Welch 1989), and it was of interest to know whether an effect of sound on speed judgments would still be measurable in conditions where retinal speed differences were available as a basis for judgment. Figure 8 shows that observers were generally accurate in judging retinal speed differences of 15% (either slow versus medium, or medium versus fast), correctly choosing the faster ball 85% of the time when the audio path cue was the same across both intervals (82.5% for roll versus roll and 87.5% for jump versus jump). When the audio cue differed across intervals, responses were more accurate when the roll sound was paired with the fastest ball (95%) and were less accurate when the roll sound was paired with the slowest ball (77.5%). Compared to baseline performance, the ball with the slowest retinal speed was more likely to be judged as faster if paired with a rolling sound and more likely

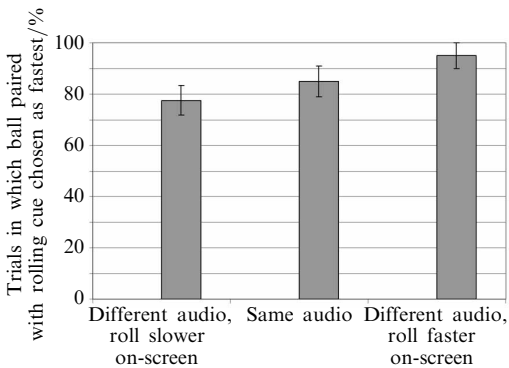


Figure 8. Results from a subset of the non-critical audio – audio trials from experiment 2 in which the on-screen speed of one interval was 15% faster than that of the other interval. The *y* axis shows the percentage of times that subjects chose the ball with the fastest on-screen speed as the fastest-moving ball. The middle bar labeled 'same audio' indicates baseline ability to discriminate the on-screen speed difference of 15% when the two balls have the same sound. The leftmost bar indicates that the tendency to pick the fastest on-screen speed is reduced when a roll sound accompanies the slower of the two on-screen videos. The rightmost bar indicates that the tendency to pick the fastest on-screen speed is enhanced when a roll sound accompanies the faster of the two on-screen videos.

to be judged as slower if paired with a jumping sound ($F_{1,9} = 6$, $p < 0.05$). Thus the effect of sound on speed perception is shown to occur even when there is other speed information available in the forced-choice task.

Finally, the short replication of conditions similar to experiment 1a was successful. The effect of sound was a significant indicator of path perception in an ANOVA with factors of sound, speed, and trial structure (two-interval forced-choice or single-interval) ($F_{1,9} = 24$, $p < 0.001$). On average, observers saw 73.3% of roll-audio trials as rolls and 84.2% of the jump-audio trials as jumps. No other factors or interactions were significant. The results of the replication did not differ significantly from those of experiment 1a (in an ANOVA with factors of experiment 1 versus 2 and sound type, $F < 1$). Of note, only one observer guessed our hypothesis that sound affects the perception of path and the resultant change in path affects the judgment of speed.

4.3 *Discussion*

In experiment 2 the same question was investigated as experiment 1 without explicitly mentioning sound or the path of the ball. Observers perceive a speed difference depending on whether a ball is paired with a rolling sound cue or a jumping sound cue. The speed difference favors the rolling ball, which travels faster in 3-D space than the jumping ball. Unlike experiment 1 where, conceivably, observers might explicitly consider audio cues in their judgment of the ball path, in experiment 2 the explicit motivation is to make an accurate speed judgment. Most observers reported that they simply experienced a speed difference, rather than trying to calculate it from path differences, and only one observer guessed the hypothesis. Therefore it seems unlikely that speed judgments were the result of a chain of explicit inferences (from sound to path, and from path length to speed). Even if our claim that observers are judging speed is arguable—that is, supposing observers say the rolling ball is faster because they perceive a longer path in the same time period rather than perceiving speed per se—this inference would nonetheless require the observers to treat the paths differently. This suggests that the paths are treated differently on the basis of sound and, because this happens without the explicit motivation to judge the path as a roll or a jump, it supplies converging evidence that sound genuinely affects the *perception* of path.

It could perhaps be argued that some spurious property of the rolling sound influences speed perception, independently of path perception. However, such a coincidence could not explain all of our findings. First, this would not explain the results of the shadow – shadow trials in which rolling balls were overwhelmingly judged as faster in the absence of any auditory information. Second, it would not explain the direction of the effect: as predicted, the roll sound increased the judged speed of the ball. If anything, the existing evidence on the perception of sound would have predicted the opposite effect. The filled-duration illusion (Thomas and Brown 1974) shows that time intervals filled with sounds appear longer than empty ones. The operation of this illusion would result in the time interval of the video containing the rolling sound appearing to be longer than the one containing the jump sound, and predicting that the fastest ball would be the one paired with a *jumping* sound. Thus, parsimony favors the conclusion that the effect of sound on speed, like the effect of shadow on speed, is mediated by changes in the ball's path.

Reinforcing this argument, an effect of sound on speed was found even when the rolling ball was moving slower on-screen than the jumping ball. If sound did not affect path perception, responses to these unequal-speed trials should strictly follow the on-screen speed of the ball.[2] The effect of sound on speed is robust and not specifically the result of a forced choice between two otherwise identical stimuli.

There was a stronger effect when the paths of both balls were defined by a shadow as opposed to sound. This was expected because the shadow has a reliable effect on depth perception in similar displays (Kersten et al 1997). The trials that include cast shadows give an indication of the ceiling effect that could be obtained by providing one of the best possible cues in this paradigm. In comparison, the smaller effect of sound on speed can be understood with respect to the reliability of its effect on path: not every video paired with a roll sound will be perceived as having a rolling ball. In the replication of experiment 1, the path consistent with the sound cue was reported, on average, 79% of the time. Therefore the best possible outcome, if speed judgments were perfectly correlated with path judgments, would be that a roll was judged as faster than a jump on 79% (rather than 100%) of the critical trials. In this context, the fact that roll sounds were judged as faster 72% of the time in experiment 2 is in remarkable agreement with our hypothesis. Taken together, experiments 1 and 2 provide converging evidence that the perception of a ball's motion is influenced by a combination of auditory and visual cues.

5 General conclusions

Perception of the world is influenced by input from multiple senses. Events in the world have both auditory and visual components, and what is needed is a better understanding of how such information is combined, especially where sound is integral to the event. Presumably, this combination occurs at a level at which information is integrated from different senses to produce a unified percept of an event, but prior to conscious inference. We speculate that this interaction occurs at a cortical site that supports multimodal perceptual processing of objects and motion, such as the superior temporal sulcus (Beauchamp et al 2004).

The results of this study are consistent with the hypothesis that depth perception, and hence path perception, is accomplished through the interaction of multiple cues.

[2] In the small subset of trials in which the on-screen speed differed and both intervals contained roll sounds, it is possible that the timing of the audio contained a cue to speed differences between fast, medium, and slow rolls, but these were not the "critical trials" comparing jumps to rolls. Because the judgment of relative speed was no less accurate when both intervals contained jump sounds, there is no evidence to support this potential concern.

Not only does this interaction occur among the various visual cues in the display (perspective, relative size, texture, shadows, etc), but *also* between visual and auditory cues. Even an event that seems as intuitively 'visual' as the perceived path of a ball can be modulated by auditory cues. Thus, a complete account of event perception, even for nominally visual events, needs to consider auditory information as part of the system.

Acknowledgments. Dan Kersten helped us adapt his ball-in-the-box demonstrations for our use. Michael Tarr and William Warren Jr provided helpful comments on earlier versions of this manuscript. Student funding was provided by an NSF IGERT Training Grant No DGE-9870676.

References

Backus B T, Banks M S, 1999 "Estimator reliability and distance scaling in stereoscopic slant perception" *Perception* **28** 217 – 242

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

Bertelson P, Aschersleben G, 1998 "Automatic visual bias of perceived auditory location" *Psychonomic Bulletin and Review* **5** 482 – 489

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

Freed D J, 1990 "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events" *Journal of the Acoustical Society of America A* **87** 311 – 322

Gaver W, 1993 "What in the world do we hear? An ecological approach to auditory event perception" *Ecological Psychology* **5** 1 – 29

Gibson J J, 1950 *The Perception of the Visual World* (Boston, MA: Houghton Mifflin)

Hairston W D, Laurienti P J, Burdette J H, Wallace M T, 2003a "Multisensory enhancement of localization under conditions of induced myopia" *Experimental Brain Research* **152** 404 – 408

Hairston W D, Wallace M T, Vaughan J W, Stein B E, Norris J L, Schrillo J A, 2003b "Visual localization ability influences cross-modal bias" *Journal of Cognitive Neuroscience* **15** 20 – 29

Kersten D, Mamassian P, Knill D, 1997 "Moving cast shadows induce apparent motion in depth" *Perception* **26** 171 – 192

Li X, Logan R J, Pastore R E, 1991 "Perception of acoustic source characteristics: walking sounds" *Journal of the Acoustical Society of America A* **90** 3036 – 3049

Lutfi R A, Oh E L, 1997 "Auditory discrimination of material changes in a struck-clamped bar" *Journal of the Acoustical Society of America A* **102** 3647 – 3656

McGurk H, MacDonald J, 1976 "Hearing lips and seeing voices" *Nature* **264** 746 – 748

McKee S, Welch L, 1989 "Is there a constancy for velocity?" *Vision Research* **29** 553 – 561

Sekuler A B, Sekuler R, 1999 "Collisions between moving visual targets: What controls alternative ways of seeing an ambiguous display?" *Perception* **28** 415 – 432

Sekuler R, Sekuler A B, Lau R, 1997 "Sound alters visual motion perception" *Nature* **385** 308

Shams L, Kamitani Y, Shimojo S, 2002 "Visual illusions induced by sound" *Cognitive Brain Research* **14** 147 – 152

Thomas E C, Brown I, 1974 "Time perception and the filled-duration illusion" *Perception & Psychophysics* **16** 449 – 458

Vanderveer N J, 1979 *Ecological Acoustics: Human Perception of Environmental Sounds* Thesis No. 40/09B, 4543, Cornell University, Ithaca, NY

Vroomen J, Gelder B de, 2000 "Sound enhances visual perception: Cross-modal effect of auditory organization on vision" *Journal of Experimental Psychology: Human Perception and Performance* **26** 1583 – 1590

Warren W H, Verbrugge R R, 1984 "Auditory perception of breaking and bouncing events: A case study in ecological acoustics" *Journal of Experimental Psychology: Human Perception and Performance* **10** 704 – 712

Young M J, Landy M S, Maloney L T, 1993 "A perturbation analysis of depth perception from combinations of texture and motion cues" *Vision Research* **33** 2685 – 2696

© 2005 a Pion publication

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 34 2005

www.perceptionweb.com

Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.