Visual determinants of a cross-modal illusion

James A. Armontrout, Michael Schutz, and Michael Kubovy
University of Virginia

Contrary to the predictions of established theory, visual information can influence the perceived duration of associated sounds (Schutz & Lipscomb, 2007). Here we deconstruct the visual component of the illusion, and show that (a) cross-modal influence depends on a visible impact event, after which (b) the illusion is controlled primarily by the duration of the post-impact motion. Other aspects of the visible motion after impact — distance traveled, velocity, acceleration, and jerk (the integral of acceleration) — play a minor role, if any.

Visual information can change the perceived duration of an auditory event (Schutz & Lipscomb, 2007). Participants see videos of a marimba player striking the instrument with either a long, flowing gesture that covered a large arc (labelled long), or with a short, choppy gesture that rebounded off the bar quickly and stopped (labelled short). Although notes produced with different gestures do not differ acoustically and the participants were asked to judge the duration of the auditory component alone (i.e., to ignore the gesture), duration ratings were longer when presented with the long, rather than the short gestures.

In the light of evidence (Walker & Scott, 1981) that vision does not influence auditory judgments of tone duration, this illusion is unexpected. It is an exception to the rule that the more accurate modality is not generally influenced by the less accurate one. In the context of temporal judgments audition is without doubt the winner. For example, estimates of flash timings are affected by temporally offset tones more-so than the estimates of tone timings are affected by temporally offset flashes (Fendrich & Corballis, 2001). Likewise, auditory flutter rate affects the perception of visual flicker rate, whereas the rate of visible flicker either fails to affect the perceived rate of concurrent auditory flutter (Shipley, 1964) or affects it minimally (Welch, DuttonHurt, & Warren, 1986).

In a previous paper, we (Schutz & Kubovy, in press-a) investigated the necessary acoustic characteristics of the stimulus for the illusion to occur. In this paper we determine some of its necessary visual characteristics.

Understanding the illusion

We have reason to believe that the illusion is triggered by perceived cross-modal causality. Before we present the evidence in support of our view, we must first rule out two incorrect explanations.

Ruling out a post-perceptual account. As Arieh and Marks (2008) have shown, certain patterns of cross-modal influence may be explained by decisional, rather than sensory shifts. In the Schutz-Lipscomb illusion, the slower gestures could have been associated in the participants' minds with longer durations, and would have produced the effect by a top-down process. To test this idea, Schutz and Kubovy (in press-a) conducted two experiments, in which they manipulated the causal relationship between the auditory and visual information.

In their first experiment they paired the same gestures with two classes of sounds: percussive and sustained (i.e., non-percussive). The non-percussive sounds, consisted of clarinet, french horn, white noise, and human voice. They also expanded their percussive sounds beyond the original marimba; they included a piano tone, which is also produced by an impact (although it involves a taut string, rather than a solid bar). Their participants were given the same instructions as in the original experiment: they were informed of audio-visual mismatch and asked to judge the duration of the auditory component alone. The gestures affected duration ratings of the percussive sounds (albeit the effect on the judged duration was lower for the piano than for the the marimba), but had no effect on the sustained ones.

In their second experiment, they manipulated the temporal synchrony between the gesture and sound: the onset of the sounds occurred either before the visible impact (audio-lead), after it (audio-lag), or simultaneously with it. They found an asymmetric visual influences — gestures affected perception in the audio-lag conditions (albeit to a lesser extent than when simultaneous), but not in the audio-lead conditions.

These results rule out the response bias account, since the long and the short gestures were equally suggestive of long and short durations under all conditions.

Optimal integration cannot explain the illusion. According to the theory of optimal integration, intermodal conflicts are resolved by giving more weight to the modality provides the more reliable information (Alais & Burr, 2004; Ernst

Research performed by JAA for a thesis in the Psychology Department Distinguished Majors Program (MK, advisor; William Epstein, reader). We thank S. Fitch of Mustard Seed Software for the outstanding programs we used to run the experiments. Correspondence should be addressed to MRS or MK, P.O.Box 400400, University of Virginia, Charlottesville, VA 22904–4400; e-mail: {schutz,kubovy}@virginia.edu. Supported by NIDCD (R01 DC 005636, MK, PI).

& Banks, 2002). For example, due to its superior spatial acuity, vision dominates spatial tasks such as the ventriloquism effect, in which speech appears to originate from the lips of a puppet (Jack & Thurlow, 1973), as well as its non-speech analogs (Jackson, 1953; Thomas, 1941; Witkin, Wapner, & Leventhal, 1952; Bertelson & Radeau, 1981; Bertelson, Vroomen, de Gelder, & Driver, 2000). Likewise, due to its superior temporal acuity, audition dominates in temporal tasks such as estimating tone duration (Walker & Scott, 1981), temporal order (Fendrich & Corballis, 2001), and visual flicker/auditory flutter rate (Shipley, 1964; Welch et al., 1986).

Furthermore, optimal integration correctly predicts that dominance patterns will be reversed as a result of ambiguity. For instance, when Wada, Kitagawa, and Noguchi (2003) paired fluttering tones with flickering lights visual influence on unambiguous sounds was minimal. However, when the quality of the auditory information was degraded, vision did have a significant influence. Similar effects have been reported by Battaglia, Jacobs, and Aslin (2003) as well as Alais and Burr (2004).

In our case, because percussive sounds decay gradually, their duration might be harder to perceive than the duration of non-percussive sounds. Observers might rely on the visual information for their judgments of auditory duration because — as the theory of optimal integration implies — the visual information was the more reliable of the two.

To show that this is not the case, Schutz and Kubovy (in press-a) estimated the uncertainty of sound duration by estimating the variability of duration ratings of percussive and non-percussive sounds when presented without the video. They compared these estimates with the magnitude of the illusion observed when these sounds were paired with impact gestures. If the optimal integration hypothesis were a correct account of these data, the relative variability of the duration ratings for percussive and non-percussive sounds in the absence of video, would predict the relative magnitude of the illusion for percussive and non-percussive sounds. It did not.

Causality and cross-modal integration. Kubovy and Schutz (in press) and Schutz and Kubovy (in press-a) proposed that causality triggers audio-visual integration. Indeed the gestures only integrated with (and therefore influenced the perception of) those sounds that they could have *caused*. Furthermore, the illusion did not only depend on the perception of a causal link, but it was also related to the degree of audio-visual plausibility. The illusion was largest when the causal link was strong: the marimba timbre in the first experiment and the synchrony condition in the second. It was moderate when the causal link was possible (though not as likely): the piano timbre in the first experiment and the audio-lag condition in the second. The illusion vanished when the gestures could not have caused the sounds: the non-percussive timbres in first experiment and the audio-lead in the second.

This idea has a precursor in the work of Sekuler, Sekuler, and Lau (1997). In an ambiguous visual display, two circles approach one another, overlap briefly, and then continue on their respective paths. The circles could be seen as either bouncing or passing through each other. The presentation of a tone at the moment of overlap increased the likelihood that the event was perceived as a bounce. Furthermore, research on the *unity assumption* (Welch, 1972; see also: Spence, 2007; Welch & Warren, 1980; Welch, 1999; Vatakis & Spence, 2007, 2008; Vroomen, 1999) and the *identity decision* (Bedford, 2001a, 2001b, 2004) also suggests that causality is important for cross-modal integration, which occurs only in the presence of cues that signal a commons source for the information in the two modalities. In a further development of this idea, Körding et al. (2007) formulated an ideal-observer model that infers whether two sensory cues originate from the same location and also estimates these locations.

Present studies

Here we ask, Which visible cues trigger the illusion? We conducted four experiments using different types of striking gestures. To make our visual stimuli readily manipulable, we created single-point versions of point light displays (Johansson, 1973), which have been used in studies of audio-visual interactions (Arrighi, Alais, & Burr, 2006; Petrini, Russell, & Pollick, 2008; Saygin, Driver, & de Sa, 2008).

Our visual stimulus consisted of a moving dot that either tracked the motion of the striking implement in the Schutz and Lipscomb videos, or was derived from it. We know (Schutz & Kubovy, in press-b) that such animations capture the aspects of the original motions that trigger the illusion.

In Experiment 1 we asked, Which aspects of the animation control the effect: pre-impact motion, post-impact motion, or some combination of the two. In Experiment 2 we asked, Which of the following elements of the animation are necessary for the illusion: (a) a change in the direction of motion at the moment the sound is heard, (b) an initial descending motion rather than an initial ascending motion (i.e., striking from above rather than striking from below), and (c) the horizontal component of the motion. In Experiment 3 we explored whether the illusion is affected by the dot's speed, the distance it travels, and the duration of its motion. In Experiment 4 examined the role of acceleration and its derivative ("jerk").

Experiment 1

We designed this experiment to determine which portion of the gesture (pre-impact or post-impact) is more important. Does the entire animation influence the perceived duration of the associated sound? Alternatively, might the portion of the animation that occurs before the moment of impact influence perceived duration, or the portion of the animation that occurs after the moment of impact? This question was first addressed by Schutz and Kubovy (in press-a) in a different manner, using two manipulations of the original videos showing the full striking gesture: (a) segment (pre-impact) — showing the gesture prior to the impact (freezing once the sound began), (b) segment (post-impact) — starting frozen on the frame depicting the moment of impact until the sound begins (then displaying the post-impact gesture along with the sound) (c) segment (both) — the original videos with the complete gesture. Their results showed that when watching half-gestures, the bulk of the visual influence could be attributed to the post-impact segment.

Since in the following experiments we used full gestures, features of which we planned to manipulate (such as velocity, distance traveled, and duration of movement), we needed to know which part or parts of the gesture to manipulate.

Method

Using GraphClick,[1] we recorded the successive positions of the mallet in the short and long conditions of Schutz and Lipscomb (when the marimbist was playing the lowest of the three notes). From these we generated two single-point point-light animations, short and long, and did not show the struck object. We used these animations to create four visual stimuli: long–long and short–short (original gestures), as well as long–short, and short–long (hybrid gestures). In the long–short animation we replaced the motion data for the post-impact portion of the short animation with the motion data for the post-impact portion of the long animation, and created the short–long analogously. In (Schutz & Kubovy, in press-b) we used the long–long and short–short animations to show that point-light stimuli have the same effect as the original videos.

We used six marimba notes: a damped (short duration) and natural (longer duration) tone from three pitch levels: E1 (~82 Hz), D4 (~587 Hz), and G5 (~1568 Hz). By combining the four animations with the six tones, we created twenty-four animations.

Twenty-eight University of Virginia undergraduates participated in exchange for credit in an introductory psychology course. The experiment took place in a quiet room using an Apple Macintosh G4 computer running custom-designed software.[2] Stimuli were presented on a ViewSonic E790B monitor (resolution: 1280×1024; refresh rate: 85 Hz) and Sennheiser HD 580 Precision headphones. Participants were allowed to adjust loudness during the warm-up period.

We randomized the order of the animations (independently for each participant), and presented each five times for a total of 120 trials. These trials were preceded by a warm-up period containing samples randomly drawn from the 24 stimuli used in the actual experiment (ratings from the warm-up period were not analyzed). Participants were told that some of the stimuli contained mismatched auditory and visual components, and were asked to judge the duration of the tone *independent of the visual information with which it was paired*. They did not receive any further information regarding the visual component of the stimuli. However, from conversations with participants in a pilot experiment we learned that they all saw the movements as impact events.

After each animation, participants rated sound duration using an unmarked 101-point slider (displayed on-screen after each stimuli), with endpoints labeled "Short" and "Long." To ensure that they were attending to the visual information, they were also required to rate the degree to which the auditory and visual components of the stimulus agreed, using a second on-screen slider with endpoints "Low agreement" and "High agreement." Rosenblum and colleagues (Rosenblum & Fowler, 1991; Saldaña & Rosenblum, 1993) have shown that this secondary task regarding audio-visual agreement does not impair ability to attend to other aspects of the auditory stimuli. Since the purpose of these ratings was only to draw the participants' attention to the visual component, we do not discuss them further here.

Results and Discussion

Data analyses. Our conclusions are based on linear mixed-effects models (also known as multilevel analyses or hierarchical linear models) estimated by restricted maximum likelihood (REML), using the function `lmer` (Bates & Sarkar, 2007), running on R (Ihaka & Gentleman, 1996). Several textbooks (Baayen, 2008; Kreft & Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999) present mixed-effects analyses, which have considerable advantages over traditional so-called repeated-measures analyses based on quasi-F tests, by-subjects analyses, combined by-subjects and by-items analyses, and random regression (Baayen, Davidson, & Bates, in press and Maxwell & Delaney, 2004, Part IV). For each set of data, we obtain estimates of effects from a *minimal adequate* (or reduced) model, which is (a) is simpler than the *maximal model* (which contains all factors, interactions and covariates that might be of any interest), (b) does not have less explanatory power than the maximal model, (c) has no submodel that is deemed adequate. The minimal adequate model is obtained from the maximal model by a process of term deletion (also known as backward selection; for an introduction, see Crawley, 2007, pp. 323–329). We report each result in terms of an effect (and its standard error, SE, in parentheses), from which a Cohen effect size, d, can be obtained by dividing the effect by its SE). To these we add a 95% confidence interval (henceforth CI), as well as a p-value for a test of the null hypothesis that the effect in question is 0. By presenting the correct error bars for mixed models we follow the recommendations of Loftus (2002, with appropriate allowance for the differences in statistical techniques); and by minimizing the role of null-hypothesis statistical tests, we implement the recommendations of the APA Task Force on Statistical Inference (Wilkinson, 1999).

The rebound portion of the gesture has a greater effect than the pre-impact portion of the gesture. As Figure 1 shows, the pre-impact and the post-impact portions of the gesture (called the *strike* and the *rebound*) have additive effects on the perceived duration of the sound. The effect of the rebound portion of the movement was 7.0 $(\pm 0.6, 95\%\text{CI} : [5.7, 8.2], p \approx 0)$ points, whereas the effect of the strike portion of the gesture was only 1.5 $(\pm 0.6, 95\%\text{CI} : [0.2, 2.8], p =$

[1] `http://www.arizona-software.ch/graphclick/`

[2] Simeon Fitch of Mustard Seed Software (`http://www.mseedsoft.com`)

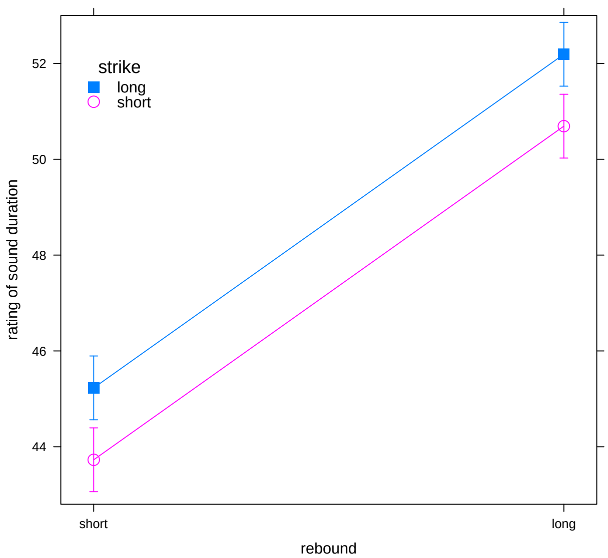


Figure 1: Experiment 1. The effect of the gesture is mainly due to the post-impact movement, the *rebound*. The pre-impact movement, the *strike*, has a small additive effect as well. The error bars are *least-significant difference* (LSD) bars: if they do not overlap, the observations are different with a p-value ≤ 0.05

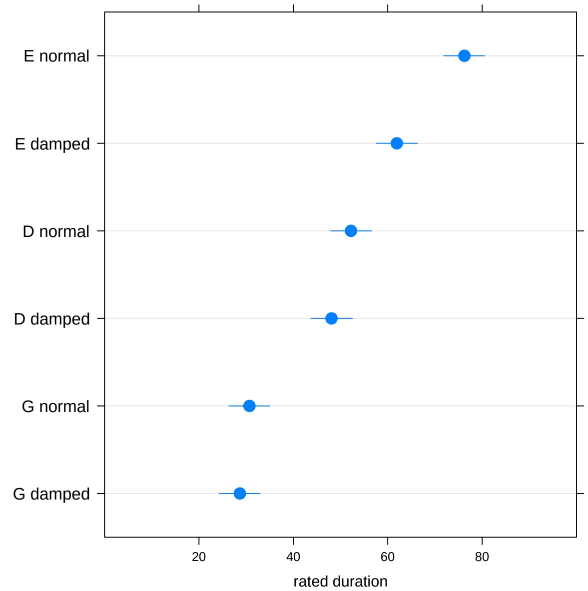


Figure 2: Experiment 1. The perceived durations of the six sounds used in this study. The effects summarized in Figure 1 were additive with the effect of these perceived durations.

0.02) points.

The illusion may be weaker with a point-light mallet-head. The magnitude of the combined effect of the strike and the rebound in this experiment was 8.5 (± 0.9, 95%CI : [6.6, 10.3], $p \approx 0$) points. This is about half as large as the effects previously observed with using actual videos of a percussionist striking the marimba (Schutz & Lipscomb, 2007; Schutz & Kubovy, in press-a). It may reflect the reduced realism of the animations.

The effect of sound duration. Figure 2 shows the rated durations of the six sounds we used. The results summarized in Figure 1, which are additive with these perceived durations (whose ratings range from 43 to 53) are the results that might be obtained with a sound whose perceived duration was between our D damped (mean rating 43.8) and our E normal (mean rating 57.7). Sounds of different perceived duration would slide the pattern of Figure 1 up and down the y-axis.

Conclusion. The illusion is largely a function of the post-impact portion of the gesture.

Experiment 2

In this experiment we asked four questions: (a) Is horizontal motion of the dot required for the illusion? (b) Does the absence of horizontal motion reduce the illusion? (c) Is a reversal in the direction of visible motion at the moment of impact (at the onset of the percussive sound) necessary? (d) Is the orientation of the striking motion important — would an up-down gesture yield similar results?

Method

We modified the short and the long animations of Experiment 1 by removing the horizontal component of the dot's motion. From these two animations we derived two others in which the dot continued moving downward after the moment of impact following a path that was the mirror image of the normal rebound (with slight smoothing to avoid artifacts). Although the struck bar was never shown in this or previous experiments, the nature of the event (i.e., "impact" vs. "through-the-bar") was discernible from the motion path alone (Schutz & Kubovy, in press-b).

From these animations we derived four inverted stimuli in which the direction of motion was reversed. The up-down motion looked like an object being struck from below. The sounds were the three natural marimba tones used in Experiment 1. By combining the eight animations with these three sounds we created twenty-four stimuli.

Forty-five University of Virginia undergraduates participated for credit in an introductory psychology course.

The animations were presented two times each in random order for a total of forty-eight trials, preceded by a warm-up period. The procedure was otherwise the same as in Experiment 1.

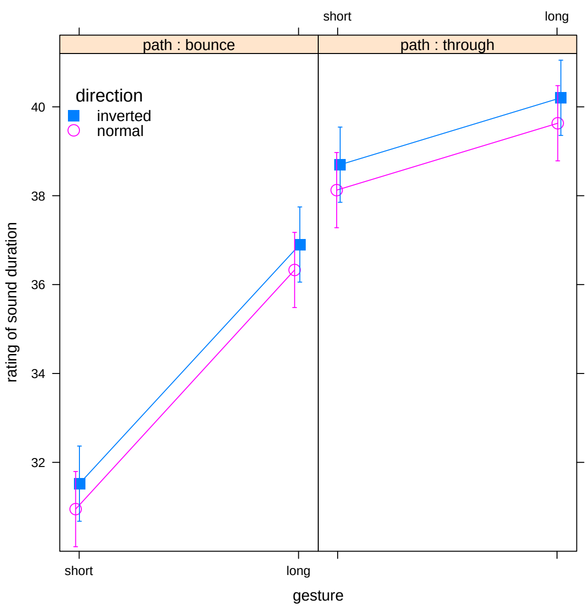


Figure 3: Experiment 2. When the dot rebounds at the moment the sound is heard, the gesture affects the sound's perceived duration (left panel). When the dot does not rebound, and continues to move in the same direction at the moment the sound is heard, the effect of the gesture is marginal (right panel). Although inverted motions (that begin by going up) appear slightly longer, this difference is not significant. (LSD error bars.)

Results and Discussion

The effect of gesture depends on whether the dot rebounds after impact. As the left-hand panel of Figure 3 shows, when the dot rebounds at the moment the sound is heard, the gesture has a 5.4 (± 0.8, 95%CI : $[3.8, 7.1]$, $p \approx 0$) point effect. In contrast, as the right-hand panel of Figure 3 shows, when the dot does not rebound at the moment the sound is heard, the gesture only has a marginal 1.5 (± 0.8, 95%CI : $[-0.1, 3.1]$, $p = 0.07$) point effect. Furthermore, although inverted (up-down) motions are rated longer than normal motions, the difference between them is minuscule.

The illusion does occur without horizontal motion, but it may be weaker. The results just summarized show that the effect of gesture occurs when horizontal motion is removed. However, the magnitude of the effect of gesture in this experiment was only 5.4 points compared to an effect of 8.5 points in Experiment 1, in which the motion of the dot had a horizontal component. This 3.1 (± 1.2, 95%CI : $[0.7, 5.5]$) point difference is not large, but it is statistically significant. Thus the horizontal motion of the dot may contribute to the illusion.

Other findings. As in Experiment 1 the three pitches we used were perceived to have different durations (with ratings ranging from 31 to 68). We also observed a small increase in average ratings over blocks. Neither effect interacted with the findings just discussed; we will not explore them further.

Conclusions. The answers to our four questions indicate that (a) horizontal motion is not a requirement for the illusion, however (b) it may strengthen it. (c) The change of direction at the moment the sound begins is crucial. (d) Orientation does not affect the illusion.

Experiment 3

Here we determined the relative contributions of post-impact speed, distance, and duration. Since these variables are inter-dependent, we manipulated two at a time and held the third constant. By comparing the relative strength of the illusion in each condition, we estimated the relative contribution of each parameter to the strength of the illusion. If the illusion is weaker when X is held constant (but Y and Z vary) between trios of long, medium, and short gestures, then X affects the illusion.

Method

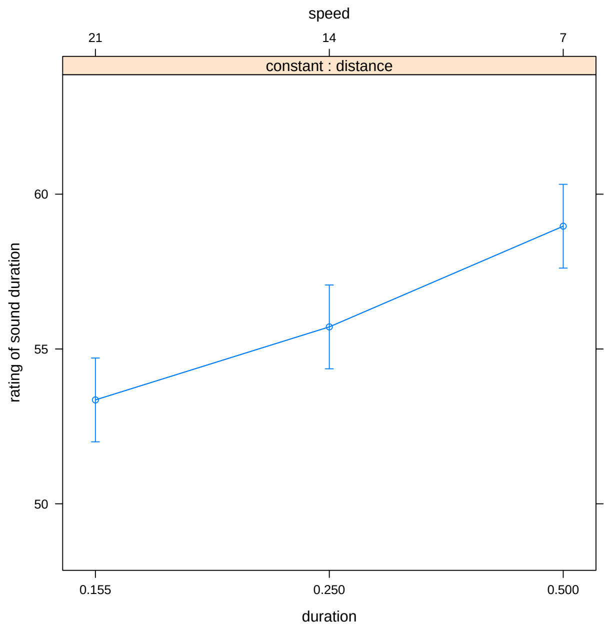
From the long animation used in Experiment 2 we created nine point-light animations with identical pre-impact motions, manipulating the the time, velocity, and distance of the post-impact rebound (summarized in Table 1). In the constant speed condition[3] we varied the distance and duration of the dot's motion while holding their ratio constant. In the constant distance condition we varied the duration of the dot's motion, and thus varied its speed. In the constant duration condition we varied the distance covered by the dot, and thus varied its speed. In an effort to more carefully control the auditory stimuli, here we used six percussive-envelope pure tones. The tones consisted of short, medium, and long (400, 850, and 1300 ms) versions with a low (A3: 220 Hz) or a high (A4: 440 Hz) pitch. They sounded unambiguously percussive. We combined the nine animations with the six sounds to create fifty-four stimuli.

Twenty-two University of Virginia undergraduates participated. They were recruited by fliers and word of mouth. They were presented with each stimuli three times for a total of 162 trials, plus 18 audio alone trials for a total of 180 trials.

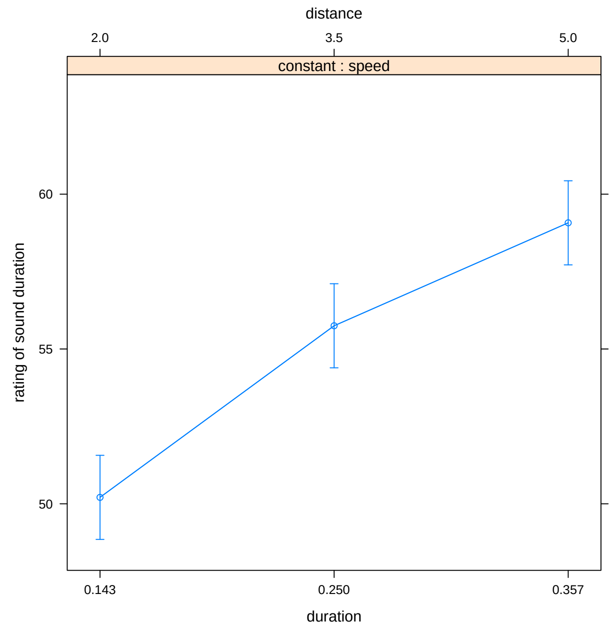
Results and Discussion

The illusion is driven by visual event-duration. We summarize the results of Experiment 3 in Figure 4 and in Table 2, showing the relative strength of each of the three tested post-impact parameters — velocity, distance, and duration. The results indicate that the illusion is weakest when duration is held constant. In other words, it is the duration of the post-impact motion that contributes most strongly to the illusion.

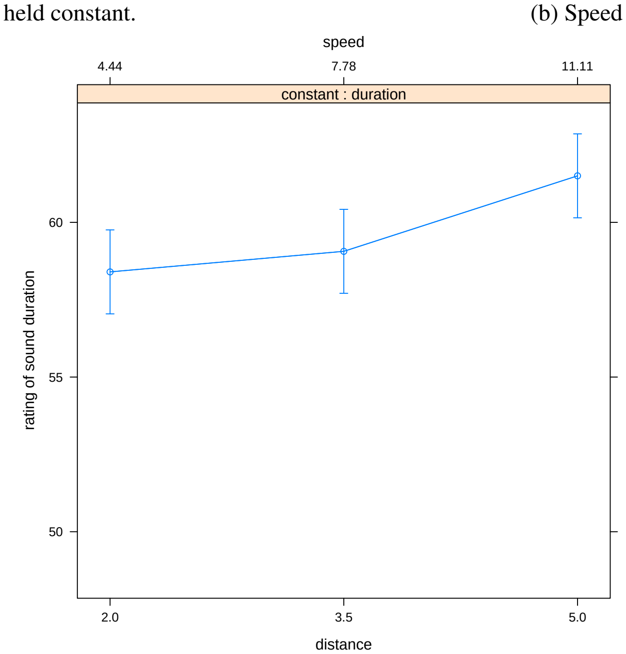
[3] We held parameters were held constant to within the degree of control available within our software.



(a) Distance is held constant.



(b) Speed is held constant.



(c) Duration is held constant.

Figure 4: Effect of holding parameters constant (i.e., removing their influence) on the strength of the illusion. The reduction in illusion strength from holding distance (left panel) and speed (middle panel) constant is considerably less than that of holding duration constant (right panel). Bars indicate LSD error bars.

Table 1: Design of Experiment 3.

	rebound		
parameter	speed (cm/sec)	distance (cm)	duration (sec)
speed	19.76	7.06	0.357
	19.76	4.94	0.250
	19.74	2.82	0.143
distance	9.88	4.94	0.500
	19.76	4.94	0.250
	31.86	4.94	0.155
duration	15.68	7.06	0.450
	10.98	4.94	0.450
	6.27	2.82	0.450

Table 2: Illusion strength when holding single parameters constant, expressed as points difference, a 95%CI, and Cohen's δ

		95% CI		
held constant	strength	lower	upper	δ
distance	5.6	2.9	8.3	2.1
speed	8.9	6.2	11.5	3.4
duration	3.1	0.5	5.8	1.2

The absence of horizontal motion and acceleration do not affect the illusion. In Experiment 2 — in which we had removed the horizontal motion — the illusion was smaller than in Experiment 1. Here we have evidence that this may not be the case: in the constant speed condition, the effect of gesture was $8.9\,(\pm1.4, 95\%\text{CI} : [6.2, 11.5], p \approx 0)$ points about the same of the effect in Experiment 1, which was $8.5\,(\pm0.9, 95\%\text{CI} : [6.6, 10.3], p \approx 0)$ points.

The use of sine-wave sounds with percussive envelopes does not weaken the illusion. In Experiment 1 we used recorded marimba sounds whereas the sounds used here were synthesized. The lack of difference between the results of these two experiment suggests that we could have used *any* percussive sound in our experiments. This is consistent with previous results indicating that impact gestures influence perception of piano (which are produced by a hammer striking a taught piano string) but not sustained (e.g. clarinet, french horn) tones (Schutz & Kubovy, in press-a).

Other effects. As in the preceding experiments, the perceived duration of the six sounds we used was quite varied (with mean ratings ranging from 10 to 82). Nevertheless the effects of these differences on the ratings were additive with the principal effects just summarized, and may be ignored.

Conclusions. Of the three examined parameters (velocity, distance, and duration of the post-impact gesture), duration contributes most strongly to the illusion. Additionally, these results suggest that removing horizontal motion may not reduce the illusion.

Experiment 4

Although Experiment 3 suggests that post-impact duration drives the illusion, it used simplified motion paths without the acceleration (the second derivative of displacement with respect to time) and jerk (the third derivative) present in the original gestures. It is possible that acceleration or jerk may play a role in the perception of animacy gender, and emotion (Pollick, Paterson, Bruderlin, & Sanford, 2001; Pollick, Lestou, Ryu, & Cho, 2002; Tremoulet & Feldman, 2000), and might therefore play a role in the illusion. To explore this possibility, we took the original gestures and successively removed parameters of the motion, creating three new conditions: (a) original motions (containing both acceleration and jerk), (b) no-jerk motions (using constant acceleration), and (c) no-acceleration motions (using constant velocity).

Method

As in Experiment 3, we created nine artificial motion paths with identical pre-impact motions. In the uniform speed condition we used long and short animations in which the dot rebounded from the point of impact to its highest point at a uniform speed and then stopped. In the uniform deceleration condition we used long and short animations in which the dot rebounded from the point of impact to its highest point with a constant deceleration so that it gradually slowed to a stop. In the third condition we used the long and short animations from Experiment 2. Because these animations were derived from video of the performer they included jerk. The durations and extent of all the long rebounds were equal to each other, which was also true of the short animations. We used six auditory stimuli: 220 and 440 Hz versions of a 400, 850, or 1300 ms percussive envelope pure tone. Combining these stimuli yielded thirty six animations.

We recruited thirty-five participants from the University of Virginia and the Charlottesville area; we paid them $8 for a session lasting approximately 15 min.

In all other respects, the procedure was the same as in the preceding experiments.

Results and Discussion

The magnitude of the effect on perceived tone duration ratings in the three conditions was identical, suggesting that acceleration and jerk of the motion have no effect on the illusion, and that it is only the duration of the motion that matters. The effect of the duration of the dot rebound was $9.4\,(\pm0.7, 95\%\text{CI} : [8.1, 10.7], p \approx 0)$ points. In contrast the three motion conditions had no effect (and did not interact with duration: acceleration condition was a paltry $1.1\,(\pm0.8, 95\%\text{CI} : [-0.5, 2.8], p = 0.2)$ points higher than the marimbist condition and a minuscule $0.5\,(\pm0.8, 95\%\text{CI} : [-1.2, 2.0], p = 0.6)$ points lower than the speed condition.

We saw no evidence of a role for acceleration or jerk in the illusion. However, by removing the horizontal component of the motion, we may have rendered jerk harder to perceive.

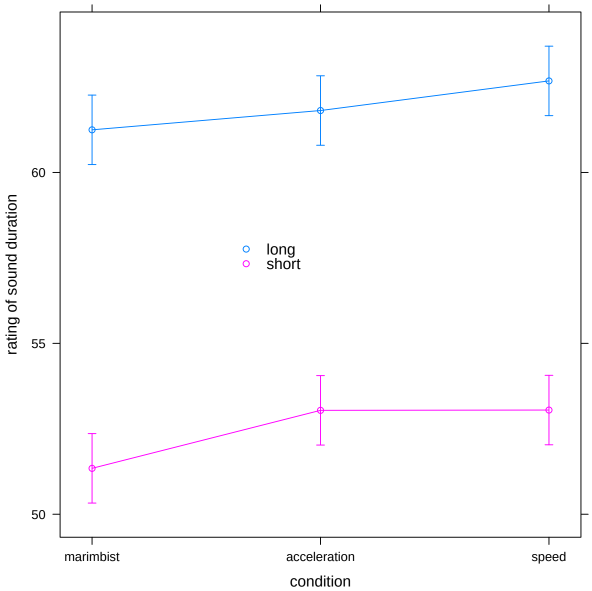


Figure 5: Experiment 4. No difference in the effect of motion duration in the experimental conditions: (a) In the marimbist condition, the dot tracked the original motion of the mallet in the video. (b) In the acceleration condition all derivatives higher than the second (including jerk), were removed, leaving uniform deceleration (while retaining variations in speed). (c) In the speed condition the velocity was uniform. (LSD error bars.)

General discussion

In four experiments we showed that the duration of motion in the post-impact gesture is primarily responsible for the visual influence on auditory judgments of tone duration. In Experiment 1 we showed that by and large the post-impact portion of the visual stimulus causes the effect. From Experiment 2 we learned that the illusion requires a rebound suggestive of an impact event. Experiments 3 and 4 revealed that duration is the most important feature of the post-impact motion, rather than distance covered, acceleration, or jerk.

These experiments show that the illusion is robust in the face of several kinds of stimulus-impoverishment: (a) reduction of the performer to the motion of a dot tracking the motion of the mallet head; (b) inversion of the motion; (c) removal of the motion's horizontal component; (d) the replacement of a recorded percussive auditory event with a synthesized percussive-sounding event; (e) the removal of all but the dot's motion-duration.

In addition to its implications for sensory integration, this work is a contribution to our understanding of the role of visual information in music perception (Thompson, Graham, & Russo, 2005; Schutz, 2008). Audiences can perceive a performer's emotional intentions (Dahl & Friberg, 2007) and the size of sung intervals (Thompson & Russo, 2007) even in the absence of auditory information. Furthermore, dance can convey certain structural and expressive aspects of music (Krumhansl & Schenck, 1997). Vision can influence the perception of pitch (Thompson et al., 2005; Gillespie, 1997), loudness (Rosenblum & Fowler, 1991) and timbre (Saldaña & Rosenblum, 1993). It also alters the perception of affect (Thompson, Russo, & Quinto, 2008), expressivity (Davidson, 1993, 1994), audience interest (Broughton & Stevens, in press), phrasing (Vines, Krumhansl, Wanderley, & Levitin, 2006), and performance quality (Wapnick, Darrow, Mazza, & Dalrymple, 1997; Wapnick, Mazza, & Darrow, 1998), as well as improve the comprehension of sung lyrics (Hidalgo-Barnes & Massaro, 2007). Because of the wide-spread nature of its influence, further understanding the aspects of individual gestures such as those used by percussionists to control perceived note duration will be useful to performers and audiences alike.

This work also provides tools for research on the role of visual information in the perception of sounds produced by percussive instruments (Broughton & Stevens, in press; Dahl, 2004; Dahl & Friberg, 2007).

The robustness of the illusion suggests that these kinds of effects are not based on musical conventions. It is thus possible that applied research on this topic may lead to knowledge that music educators and students might reliably use to improve musical performances.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.

Arieh, Y., & Marks, L. E. (2008). Cross-modal interaction between vision and hearing: A speed-accuracy analysis. *Perception & Psychophysics*, *70*, 412–421.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, *6*, 260–268.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (in press). *Mixed-effects modeling with crossed random effects for subjects and items.* (Corrected proof, accessed June 5, 2008, `http://dx.doi.org/10.1016/j.jml.2007.12.005`)

Bates, D., & Sarkar, D. (2007). *`lme4`: Linear mixed-effects models using S4 classes.* (R package version 0.9975-13)

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Optical Society of America*, *20*, 1391–1397.

Bedford, F. L. (2001a). Object identity theory and the nature of general laws. *Cahiers de Psychologie Cognitive—Current Psychology of Cognition*, *20*, 277–293.

Bedford, F. L. (2001b). Towards a general law of numerical/object identity. *Cahiers de Psychologie Cognitive—Current Psychology of Cognition*, *20*, 113–175.

Bedford, F. L. (2004). Analysis of a constraint on perception, cognition, and development: One object, one place, one time. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 907–912.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*, 578–584.

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

Broughton, M., & Stevens, C. (in press). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*.

Crawley, M. J. (2007). *The R book*. New York, NY: Wiley.

Dahl, S. (2004). Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica united with Acustica*, *90*, 762–776.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*(5), 433-454.

Davidson, J. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*.

Davidson, J. (1994). Which areas of a pianist's body convey information about expressive intention to an audience? *Journal of Human Movement Studies*, *26*, 279–301.

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

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*, 719–725.

Gillespie, R. (1997). Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations. *Journal of Research in Music Education*, *45*, 212–220.

Hidalgo-Barnes, M., & Massaro, D. W. (2007). Read my lips: an animated face helps communicate musical lyrics. *Psychomusicology*, *19*, 3-12.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*, 299–314.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills*, *37*, 967–979.

Jackson, C. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*, 52–65.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *2*, 201–211.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLos ONE*, *2*, 1–10.

Kreft, I., & Leeuw, J. de. (1998). *Introducing multilevel modeling*. London, UK: Sage.

Krumhansl, C., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? a perceptual experiment on balanchines choreography of mozarts divertimento no. 15. *Musicae Scientae*, *1*, 63–85.

Kubovy, M., & Schutz, M. (in press). Audio-visual objects. *Philosophy & Psychologyy*, *1: Objects and Sound Perception*, ??–??

Loftus, G. R. (2002). Analysis, interpretation, and visual presentation of experimental data. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 4: Methodology in experimental psychology, pp. 339–390). New York, NY, USA: John Wiley & Sons.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ, USA: Erlbaum.

Petrini, K., Russell, M., & Pollick, F. (2008). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, *In Press, Corrected Proof*, –.

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S.-B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, *42*, 2345–2355.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*, B51–B61.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London, UK: Sage.

Rosenblum, L. D., & Fowler, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 976–985.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, *54*, 406–416.

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception. *Psychological Science*, *19*, 469–475.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review*, *3*, 83–108.

Schutz, M., & Kubovy, M. (in press-a). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance*.

Schutz, M., & Kubovy, M. (in press-b). Deconstructing the Schutz–Lipscomb illusion: Exploring visual influences on the perception of tone duration. *Canadian Acoustics*.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*, 888–897.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, *145*, 1328–1330.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustic Science & Technology*, *28*, 61–71.

Thomas, G. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, *28*, 163–175.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performace: Visual influences on perception and experience. *Semiotica*, *156*, 203–227.

Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, *18*, 756–757.

Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition & Emotion*, *22*, 1457–1470.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*, 943–951.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*, 744–756.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*, 12–23.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J.

(2006). Cross–modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113.

Vroomen, J. (1999). Ventriloquism and the nature of the unity decision. Commentary on Welch. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 389–393). Amsterdam: Elsevier.

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *Psychophysiology*, *50*, 117–124.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1327–1339.

Wapnick, J., Darrow, A.-A., Mazza, J. K., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, *45*, 4270-479.

Wapnick, J., Mazza, J. K., & Darrow, A.-A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education*, *46*, 510–521.

Welch, R. B. (1972). The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception & Psychophysics*, *12*, 453–456.

Welch, R. B. (1999). Meaning, attention, and the 'unity assumption' in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam, Netherlands: North-Holland/Elsevier.

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, *39*, 294–300.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638–667.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604.

Witkin, H. A., Wapner, S., & Leventhal, T. (1952). Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology*, *43*, 58–67.