Visual determinants of a cross-modal illusion

JAMES A. ARMONTROUT
University of Virginia, Charlottesville, Virginia

MICHAEL SCHUTZ
McMaster University, Hamilton, Ontario, Canada

AND

MICHAEL KUBOVY
University of Virginia, Charlottesville, Virginia

Contrary to the predictions of established theory, Schutz and Lipscomb (2007) have shown that visual information can influence the perceived duration of concurrent sounds. In the present study, we deconstruct the visual component of their illusion, showing that (1) cross-modal influence depends on visible cues signaling an impact event (namely, a sudden change of direction concurrent with tone onset) and (2) the illusion is controlled primarily by the duration of post-impact motion. Other aspects of the post-impact motion—distance traveled, velocity, acceleration, and the rate of its change (i.e., its derivative, *jerk*)—play a minor role, if any. Together, these results demonstrate that visual event duration *can* influence the perception of auditory event duration, but only when stimulus cues are sufficient to give rise to the perception of a causal cross-modal relationship. This refined understanding of the illusion's visual aspects is helpful in comprehending why it contrasts so markedly with previous research on cross-modal integration, demonstrating that vision does not appreciably influence auditory judgments of event duration (Walker & Scott, 1981).

Schutz and Lipscomb (2007) reported a naturally occurring audio–visual illusion in which visual information changes the perceived duration of simultaneous auditory information. They demonstrated this by showing participants videos of a percussionist striking a marimba with either a long flowing gesture (labeled "long") that covered a large arc or with a short choppy gesture (labeled "short") that rebounded off of the bar and quickly stopped. Although the resultant sounds were acoustically indistinguishable and participants were asked to ignore visual information when judging tone duration, duration ratings were longer when presented with long rather than short gestures.

In light of evidence that vision does not influence auditory judgments of tone duration (Walker & Scott, 1981), this illusion is unexpected. It is an exception to the rule that, with respect to a given task, the modality offering less accurate information does not appreciably influence the modality offering more accurate information. For example, the superior temporal precision of the auditory system generally translates into auditory dominance for temporal tasks such as the judgment of tone duration. Likewise, estimates of flash timings are more affected by temporally offset tones than estimates of tone timings are affected by temporally offset flashes (Fendrich & Corballis, 2001); and 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).

Understanding the Illusion

We believe that the perception of a causal link between auditory and visual information is crucial to explaining why the illusion reported by Schutz and Lipscomb (2007) conflicts so strongly with previous work on sensory integration. However, before presenting evidence in support of this view, we will first discuss two alternative explanations that have been previously dismissed by Schutz and Kubovy (in press). We will close this section by explaining our reasons for proposing that causality plays an important role and by discussing links between this illusion and previous work on the unity assumption.

Post-perceptual processing cannot explain the illusion. As has been shown by Arieh and Marks (2008), certain patterns of cross-modal interactions may be explained by decisional changes, rather than by sensory shifts. Therefore, it is possible that longer gestures could have suggested longer durations, affecting ratings through a top-down process (i.e., a response bias), without any actual perceptual shift. To test this explanation, Schutz and Kubovy (in press) designed a series of experiments manipulating the causal relationship between the auditory and visual components of the stimuli.

M. Schutz, schutz@mcmaster.ca

In their first experiment, Schutz and Kubovy (in press) paired the impact gestures with two classes of sounds: percussive and non-percussive (i.e., sustained). The non-percussive sounds consisted of single tones produced by the clarinet, French horn, or human voice (singing), as well as white noise. The percussive sounds consisted of the original marimba tone as well as those produced by a piano (an impact event involving 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 were asked to judge the duration of the auditory component alone. Here, gestures affected duration ratings of both percussive sounds (albeit to a lesser degree for the piano than for the marimba) but had no effect on the non-percussive (sustained) ones.

In their second experiment, Schutz and Kubovy (in press) manipulated the temporal synchrony between the gesture and sound, such that tone onset occurred before the moment of visible impact (audio lead), after the moment of visible impact (audio lag), or simultaneously with the moment of impact (original videos). Here, the visual influence was asymmetric; gestures affected perception in the audio-lag condition (albeit to a lesser extent than in the simultaneous condition), but not in the audio-lead condition. This asymmetry with respect to audio lag and lead is consistent with the ecology of our environment, in which the speed of sound is significantly slower than the speed of light.

In their final experiment, Schutz and Kubovy (in press) addressed head-on the issue of a potential response bias by replacing some of the impact gestures with the written text "Long" and "Short." The text had no meaningful influence on duration ratings, demonstrating that the mere suggestion of long and short is insufficient to explain the original illusion. This is also consistent with a causal account; clearly the written text did not cause the tones in question, and therefore there is no reason for the two to be perceptually integrated.

Together, these three experiments provide compelling evidence against a response bias account of the illusion. The long and short gestures in the first two experiments were equally suggestive under all conditions, yet they selectively affected the perception of particular sounds—that is, those that they could have *caused*. Furthermore, written text did not affect duration ratings, demonstrating that the mere suggestion of long and short cannot explain the illusion. In light of these experiments, it is clear that the data reported by Schutz and Lipscomb (2007) cannot be dismissed as a response bias.

Optimal integration cannot explain the illusion. According to the theory of optimal integration, intermodal conflicts are resolved by giving more weight to the modality providing the more reliable information (Alais & Burr, 2004; Ernst & 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 nonspeech analogues (Bertelson & Radeau, 1981; Bertelson, Vroomen, de Gelder, & Driver, 2000; Jackson, 1953; Thomas, 1941; Witkin, Wapner, &

Leventhal, 1952). Likewise, due to its superior temporal acuity, audition dominates 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).

Optimal integration correctly predicts performance in a wide variety of tasks, including instances in which dominance patterns are reversed as a result of ambiguity. For example, 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 by Alais and Burr (2004). Therefore, there is reason to believe that the data presented by Schutz and Lipscomb (2007) might be explained by the theory of optimal integration. Because percussive (i.e., impact) sounds decay gradually, their duration might be harder to perceive than the duration of non-percussive sounds that have more clearly defined offsets. If that is so, observer–listeners might rely more on visual information than on audio information when judging the duration of impact sounds because—as predicted by the theory of optimal integration—in such cases visual information is more reliable than audio information.

To test this explanation of the illusion, Schutz and Kubovy (in press) examined the variability of the duration ratings for visually influenced percussive (e.g., marimba and piano) and non-visually-influenced sustained (e.g., clarinet and French horn) sounds when presented as audio alone. They then compared these evaluations with the magnitude of the illusion observed when these sounds were paired with the impact gestures. Contrary to the optimal integration hypothesis, the variability of duration ratings did not predict the relative magnitude of the illusion.

Causality and cross-modal integration. Because neither the response bias account nor the theory of optimal integration can explain the illusion, Schutz and Kubovy (in press) and Kubovy and Schutz (in press) have proposed that, in this context, the perception of a causal cross-modal link serves as a key trigger for audio–visual integration. In each case, the gestures integrated with (and therefore influenced the perception of) only those sounds that they could have *caused*. Furthermore, the strength of the illusion in these previous experiments was related to the degree of plausible causality. In other words, the illusion was largest when the causal link was strongest: the marimba timbre in the first experiment and the synchrony condition in the second. It was moderate when the causal link was possible (although less 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 the first experiment and the audio-lead condition in the second.

The role of causality in cross-modal integration is not without precedent. For example, Sekuler, Sekuler, and Lau (1997) devised an ambiguous visual display, depicting two identical circles approaching one another, overlapping briefly, and then continuing 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. Likewise, research on the unity assumption (Welch, 1972; see also Spence, 2007; Vatakis & Spence, 2007, 2008; Vroomen, 1999; Welch, 1999; Welch & Warren, 1980) and the identity decision (Bedford, 2001a, 2001b, 2004) also suggests that causality is important for cross-modal integration. In a further development of this idea, Körding et al. (2007) formulated an ideal-observer model from which it can be inferred whether two sensory cues originate from the same location; it also estimates these locations. We see the illusion reported by Schutz and Lipscomb (2007) as well as by subsequent investigations (Schutz & Kubovy, 2009, in press) as a continuation of this work and believe that causality serves as an important cue for triggering the assumption of unity.

Present Study

Because previous work examining the acoustic cues for integration was informative (Schutz & Kubovy, in press), here we designed four experiments to examine the necessary visual cues. Because this requires precise control over all aspects of the motion paths, we created single-point versions of point light displays (Johansson, 1973), which have previously proven useful in studies of audio–visual interactions (Arrighi, Alais, & Burr, 2006; Petrini, Russell, & Pollick, 2009; Saygin, Driver, & de Sa, 2008). Therefore, our visual stimuli consisted of a moving dot that either tracked the motion of the striking implement in the Schutz and Lipscomb (2007) videos or was derived from it. We know (Schutz & Kubovy, 2009) that such animations capture the salient aspects of the original motions triggering the illusion.

In Experiment 1, we explored 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: (1) a change in the direction of motion at the moment the sound is heard, (2) an initial descending motion rather than an initial ascending motion (i.e., striking from above rather than striking from below), and (3) the horizontal component of the motion. In Experiment 3, we explored whether the illusion is affected by the dot's velocity, the distance it travels, and the duration of its motion. In Experiment 4, we examined the roles of acceleration and of the rate of its change (i.e., its derivative, *jerk*).

EXPERIMENT 1

We designed this experiment to determine which portion of the gesture (pre-impact or post-impact) is more important when full gestures are viewed. This question was first addressed in part by Schutz and Kubovy (in press) differently, by dividing the original videos into two segments. The first (pre-impact) showed only the gesture before the moment of impact, freezing once the sound began. The second (post-impact) started frozen on the frame depicting the moment of impact, then displayed the post-impact gesture beginning at the onset of the sound. Their results demonstrate that when half gestures are viewed, the influence of the pre-impact segment is trivial relative to the influence of the post-impact segment.

Experiment 1 was designed to extend that work by using abstract representations of *full* gestures, allowing for more sophisticated manipulations than previously were available with video recordings. This provides an important first step for the subsequent experiments, which were designed to explore the independent contributions of acceleration, velocity, distance traveled, and motion duration within the (presumably) more salient post-impact segment of the gesture.

Method

Using GraphClick (www.arizona-software.ch/graphclick), we recorded the successive positions of the mallet in the short and long conditions of Schutz and Lipscomb (2007, when the marimbist was playing the lowest of the three notes). From these we generated 2 single-point animations, short and long (none of the animations contained any representation of the struck object). We used these animations to create 4 visual stimuli: long–long and short–short (original gestures) and long–short and short–long (hybrid gestures). In the long–short animation, we paired the motion data from the pre-impact portion of the long with the post-impact portion of the short; we created the short–long animation analogously. The long–long and short–short animations were identical to the long and short single-dot animations used by Schutz and Kubovy (2009). We used six marimba tones: 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 4 animations with the six tones, we created 24 animations.

None of these audio–visual pairings appeared implausible, not even the extreme case of pairing the long–long gesture with the shortest sound or the short–short gesture with the longest sound. Because gesture length has no effect on acoustic note duration (Schutz & Lipscomb, 2007), it is not implausible that the two might disagree. In fact, such disagreement would be inevitable when a long gesture is used to strike an object that produces short sounds or when a short gesture is used to strike an object that produces long sounds.

Twenty-eight University of Virginia undergraduates participated in exchange for credit in an introductory psychology course. The experiment took place in a quiet room; we used an Apple Macintosh G4 computer running custom-designed software.[1] Stimuli were presented on a ViewSonic E790b monitor (resolution = 1,280 × 1,024 pixels; refresh rate = 85 Hz) and Sennheiser HD580 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 15-trial 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 *independently of the visual information with which it was paired*. Although they were never told that the stimuli were derived from impact gestures, from conversations with participants in a pilot experiment, we learned that most interpreted the motions as depicting some type of impact event.

After each animation, participants rated sound duration using an unmarked 101-point slider (displayed on screen), with endpoints labeled "Short" and "Long." To ensure that they were attending to the visual information, as in previous studies 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 labeled "Low agreement" and "High agreement." Rosenblum and colleagues (Rosenblum & Fowler, 1991; Saldaña & Rosenblum, 1993) have shown that this secondary task regarding audio–visual
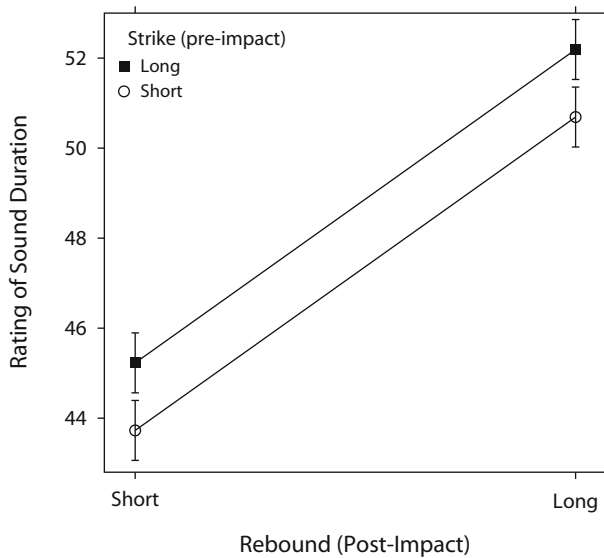
Figure 1. Experiment 1. Although the pre-impact *strike* portion of the gesture had a small influence, the majority of the illusion was driven by the post-impact *rebound* motion. The error bars are least-significant difference bars: If they do not overlap, the observations are different, with a *p* value ≤ .05.

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 will not discuss them further in this article.

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 & de 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, 2008; Maxwell & Delaney, 2004, Pt. IV). For each set of data, we estimate effects by using a *minimal adequate* (or reduced) model, which (1) is simpler than the *maximal model* (which contains all factors, interactions, and covariates that might be of any interest), (2) does not have less explanatory power than the maximal model, and (3) 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 ($CI_{95\%}$), 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 Lof-

tus (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 American Psychological Association Task Force on Statistical Inference (Wilkinson & the Task Force on Statistical Inference, 1999).

The post-impact rebound portion of the gesture has a greater effect than the pre-impact strike portion of the gesture. As Figure 1 shows, the pre-impact and 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 points (± 0.6, $CI_{95\%} = [5.7, 8.2]$, $p \approx 0$), whereas the effect of the strike portion of the gesture was only 1.5 points (± 0.6, $CI_{95\%} = [0.2, 2.8]$, $p = .02$).

Magnitude of the illusion. The magnitude of the combined effect of the strike and the rebound in this experiment was 8.5 points (± 0.9, $CI_{95\%} = [6.6, 10.3]$, $p \approx 0$). This is about half as large as the effects previously observed when using video recordings (Schutz & Kubovy, in press; Schutz & Lipscomb, 2007), as opposed to the animations used in these experiments. It is possible that this reduction reflects the reduced degree of realism inherent in abstract stimuli, an issue that will be addressed in future studies.

The effect of sound duration. Figure 2 shows the rated durations of the six sounds. The visual influence on these sounds (Figure 1) is indicative of 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 durations would slide the pattern of Figure 1 up and down the *y*-axis.
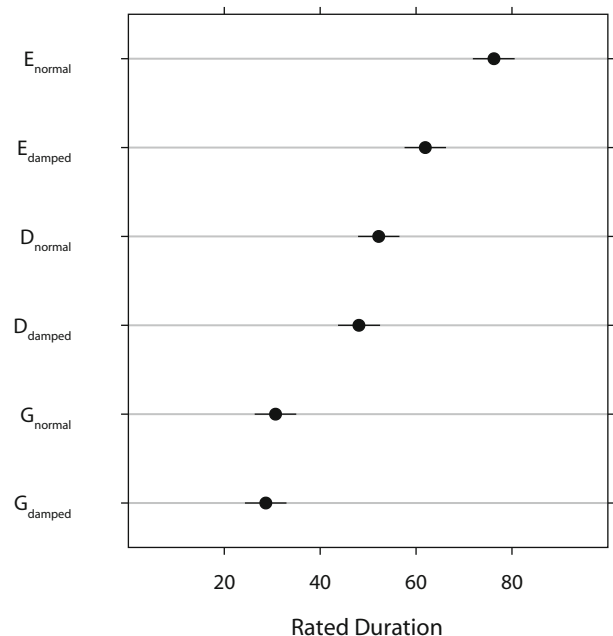

Figure 2. Experiment 1. The perceived durations of the six sounds used in this study. The effects summarized in Figure 1 were additive with the effect of these perceived durations.

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

EXPERIMENT 2

In this experiment, we addressed four questions: (1) Is horizontal motion of the dot required for the illusion? (2) Does the absence of horizontal motion reduce the illusion? (3) Is a reversal in the direction of visible motion at the moment of impact (i.e., at the onset of the percussive sound) necessary? (4) Is the orientation of the striking motion important; that is, would an up–down gesture yield similar results?

Method

We modified the short and the long stimuli 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 mirrored the normal rebound (with slight smoothing to avoid artifacts). As in the first experiment, the struck bar was not represented because previous research showed that it is not a requirement for the integration of percussive sounds and impact gestures (Schutz & Kubovy, 2009).

From these animations, we derived 4 inverted stimuli, in which the direction of motion was reversed (i.e., the up–down motion mimicked striking an object from below). The sounds were the three natural marimba tones used in Experiment 1. By combining these eight animations with these three sounds we created 24 stimuli. Forty-five University of Virginia undergraduates participated for credit in an introductory psychology course. Each animation was presented twice in random order for a total of 48 trials, preceded by a warm-up period. The procedure was otherwise the same as in Experiment 1.

Results and Discussion

The effect of gesture depends on whether the dot rebounds after impact. As Figure 3A shows, when the dot rebounds at the moment the sound is heard, the gesture has a 5.4-point effect (± 0.8, $\text{CI}_{95\%} = [3.8, 7.1]$, $p \approx 0$). In contrast, as Figure 3B shows, when the dot does not rebound at the moment the sound is heard, the gesture has only a marginal 1.5-point effect (± 0.8, $\text{CI}_{95\%} = [-0.1, 3.1]$, $p = .07$). Furthermore, although inverted (up–down) motions are rated as being 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, as compared with an effect of 8.5 points in Experiment 1, in which the motion of the dot had a horizontal component. This 3.1-point difference (± 1.2, $\text{CI}_{95\%} = [0.7, 5.5]$) is not large, but it is statistically significant. Thus the horizontal motion of the dot may contribute to the illusion. This is consistent with our observation in the first experiment that reductions to the degree of gesture realism might reduce the magnitude of 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 discuss them further.
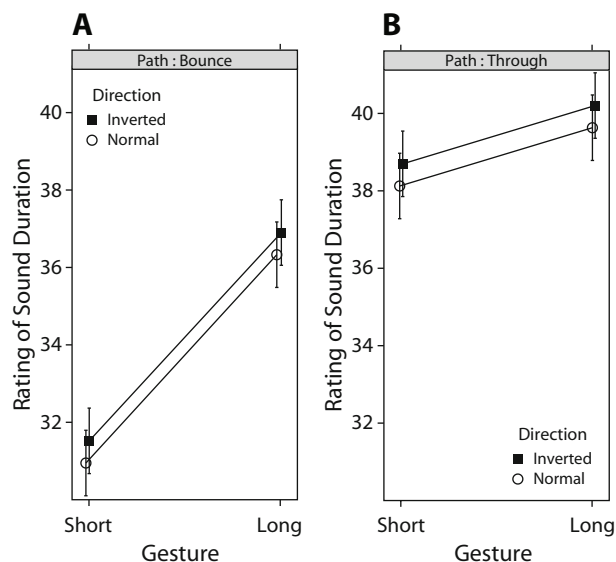

Figure 3. Experiment 2. (A) When the dot rebounds at the moment the sound is heard, the gesture affects the sound's perceived duration. (B) When the dot does not rebound and continues to move in the same direction at the moment the sound is heard, the effect of the gesture is marginal. Although inverted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

Conclusions. The answers to our four questions indicate that (1) horizontal motion is not a requirement for the illusion, however (2) it may strengthen it; that (3) the change of direction at the moment the sound begins is crucial; and that (4) orientation does not affect the illusion.

EXPERIMENT 3

In this experiment, we explored the relative contributions of post-impact velocity, distance, and duration. Because these are interdependent, we created three groups of gestures, manipulating two variables within each group while holding the third constant. This allowed us to examine each variable independently by measuring the strength of the illusion in its absence (i.e., if the illusion was weaker when X was held constant but Y and Z varied, we can conclude that X plays a meaningful role).

Table 1
Design of Experiment 3: Rebound Velocity,
Distance, and Duration

| | Rebound | | |
Parameter	Velocity (cm/sec)	Distance (cm)	Duration (sec)
Velocity	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

A   **Distance Is Held Constant**

B   **Velocity Is Held Constant**
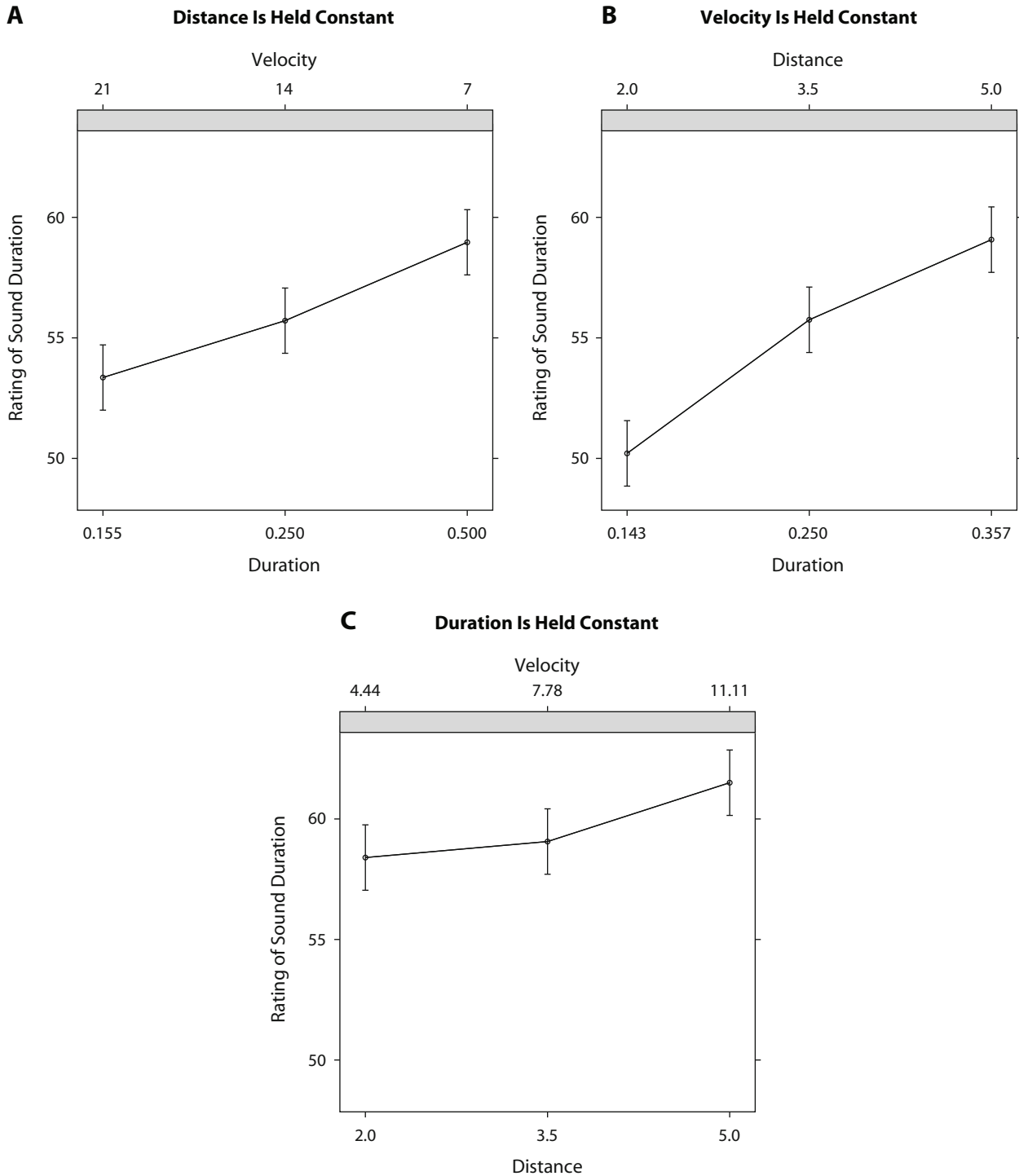

C   **Duration Is Held Constant**

Figure 4. Experiment 3. Effect of holding parameters constant (i.e., removing their influence) on the strength of the illusion. The reduction in illusion strength when holding distance (A) and velocity (B) constant is considerably less than that of holding duration (C) constant. Bars indicate least-significant difference error.

Method

From the vertical component of the *long* animation used in Experiment 2, we created nine point-light animations (three groups with three gestures each) with identical pre-impact motions, as summarized in Table 1. In the uniform-velocity condition,[2] we varied the distance and duration of the dots' motion, while holding their ratio (velocity) constant. In the uniform-distance condition, we varied

motion duration and velocity, while holding the distance traveled constant. In the uniform-duration condition, we varied distance and velocity, while holding the duration of motion constant. None of the animations contained horizontal motion.

In an effort to more carefully control the auditory stimuli, in this experiment, we used tones with exponentially decaying (i.e., percussive) envelopes of varying length. The tones consisted of short,

Table 2
Illusion Strength When Single Parameters Are Held Constant,
Expressed As a Point Difference on the Rating Scale, a 95%
Confidence Interval (CI$_{95\%}$), and Cohen's d

Parameter Held Constant	Strength	CI$_{95\%}$		Cohen's d
		Lower	Upper	
Distance	5.6	2.9	8.3	2.1
Velocity	8.9	6.2	11.5	3.4
Duration	3.1	0.5	5.8	1.2

medium, and long (400-, 850-, and 1,300-msec) versions with a low (A3, 220 Hz) or a high (A4, 440 Hz) pure tone, which sounded unambiguously percussive. We combined the nine animations with the six sounds to create 54 stimuli.

We presented each audio–visual block three times and a single audio-alone block containing three presentations of each sound, for a total of 162 audio–visual and 18 audio-alone trials. The order of blocks, as well as the order of stimuli within each block, was randomized for each participant. Twenty-two University of Virginia undergraduates recruited by fliers and word of mouth participated in the experiment.

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.

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 uniform-velocity condition, the effect of gesture was 8.9 points ($\pm$1.4, CI$_{95\%}$ = [6.2, 11.5], $p \approx 0$), about the same as the effect in Experiment 1, which was 8.5 points ($\pm$0.9, CI$_{95\%}$ = [6.6, 10.3], $p \approx 0$).

The use of pure tones 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 experiments suggests that we could have used *any* percussive sound in our experiments. This is consistent with previous results indicating that impact gestures influence the perception of piano tones (produced by a hammer striking a taught string) but not sustained tones, such as those produced by a clarinet or French horn (Schutz & Kubovy, in press).

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 can be ignored.

Conclusions. Although velocity and distance contribute to the illusion, the contribution of duration is clearly

the strongest. Additionally, the fact that the illusion replicates with exponentially decaying pure tones suggests that this effect may generalize to all impact sounds.

EXPERIMENT 4

Although Experiment 3 did suggest that post-impact duration drives the illusion, it used simplified motion paths without the acceleration (second derivative of displacement with respect to time) and jerk (third derivative) present in the original gestures. Because acceleration and jerk may play a role in the perception of qualities such as animacy, gender, and emotion (Pollick, Lestou, Ryu, & Cho, 2002; Pollick, Paterson, Bruderlin, & Sanford, 2001; Tremoulet & Feldman, 2000), it is possible that their omission might be problematic. To explore this possibility, we took the original gestures and successively removed parameters of the motion, creating three new conditions: (1) marimbist (original motions containing both acceleration and jerk), (2) uniform acceleration (no jerk), and (3) uniform velocity (no acceleration, no jerk).

Method

As in Experiment 3, we created nine artificial motion paths (three groups of three gestures) with identical pre-impact motions. In the *marimbist* condition, the post-impact motion was simply the vertical component of the long and short animations used in Experi-
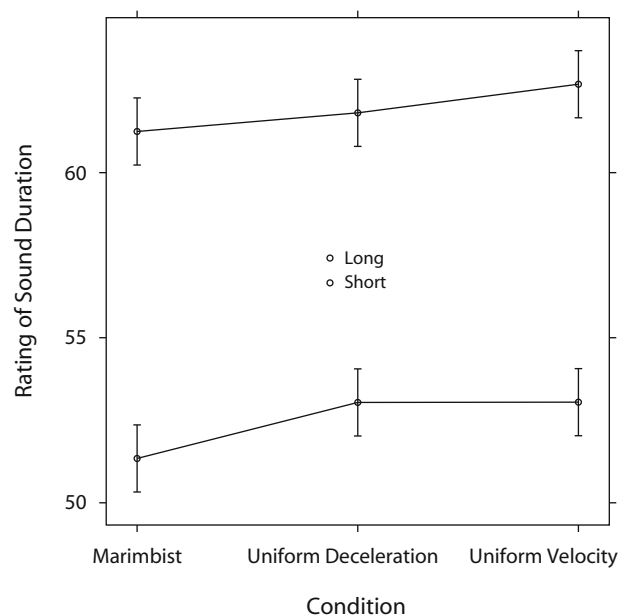

Figure 5. Experiment 4. There was no difference in the effect of motion duration in the experimental conditions. (A) In the marimbist condition, the dot tracked the original motion of the mallet in the video. (B) In the second condition, all derivatives higher than the second (including jerk) were removed, leaving uniform deceleration (while retaining variations in velocity). (C) In the final condition, the velocity was uniform. Bars indicate least-significant difference error.

ments 2 and 3. Because these animations were derived from videos of human motions, they included jerk, deceleration, and velocity. For the *uniform-deceleration* condition, we generated long and short animations in which the dot rebounded with a uniform deceleration, gradually slowing to a stop. For the *uniform-velocity* condition, we generated long and short animations in which the dot rebounded from the point of impact at a uniform velocity. The durations and extent of all the long rebounds were equal, which was true also of the short animations. As in Experiment 3, none of the animations contained horizontal motion. We used six auditory stimuli: 220- and 440-Hz versions of a 400-, 850-, or 1,300-msec percussive envelope pure tone. Combining these stimuli yielded 36 animations.

We presented each audio–visual block twice, in addition to a single audio-alone block containing two presentations of each sound, for a total of 72 audio–visual and 12 audio-alone trials. The order of blocks, as well as the order of stimuli within each block, was randomized for each participant. We recruited 35 participants from the University of Virginia and the Charlottesville area and 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 did not differ significantly across the three conditions (Figure 5). This suggests that acceleration and jerk do not contribute meaningfully and that post-impact duration is truly the most important factor contributing to the illusion. The effect of rebound duration was 9.4 points (± 0.7, $CI_{95\%} = [8.1, 10.7]$, $p \approx 0$). In contrast, the other two variables had minimal effects and did not interact. The illusion magnitude in the uniform acceleration condition was a paltry 1.1 points (± 0.8, $CI_{95\%} = [-0.5, 2.8]$, $p = .2$) lower than in the marimbist condition and a minuscule 0.5 points (± 0.8, $CI_{95\%} = [-1.2, 2.0]$, $p = .6$) lower than in the uniform-velocity condition. Therefore, we saw no evidence of a role for acceleration or jerk. However, it is possible that by removing the horizontal component of the motion, we may have rendered jerk harder to perceive.

GENERAL DISCUSSION

Experiment 1 showed that visual influence is governed primarily by the post-impact portion of the gesture, and Experiments 3 and 4 that post-impact duration (rather than distance covered, velocity, acceleration, or jerk) is the driving force behind this effect. However, as shown in Experiment 2, this influence is conditioned upon the perception of a causal cross-modal link: Vision's influence disappeared when the gesture appeared to move "through-the-bar" (and could not have caused the accompanying sound). Therefore, we conclude that within this paradigm, visual influence is (1) contingent upon the perception of a causal cross-modal relationship and (2) largely a function of post-impact motion duration. As such, these experiments build upon the work of Schutz and Kubovy (in press) by demonstrating that the reason for the conflict between the illusion documented by Schutz and Lipscomb (2007) and previous work on audio–visual integration stems from the clear causal cross-modal link inherent in impact events. More broadly speaking, we view this work as a continuation of research on the role of causality in sensory integration

(Bedford, 2001a, 2001b; Spence, 2007; Vatakis & Spence, 2007, 2008; Vroomen, 1999; Welch, 1972).

In addition, these experiments demonstrate that the illusion is robust in the face of several kinds of stimulus impoverishment: (1) reduction of the performer's image to the motion of a dot tracking the motion of the mallet head, (2) inversion of the motion, (3) removal of the motion's horizontal component, (4) the replacement of a recorded percussive auditory event with a synthesized percussive-sounding event, and (5) the removal of all but the dot's motion duration.

In addition to its implications for sensory integration, this work represents a contribution to our understanding of music perception, a domain in which visual information is now regarded as playing an important role (for reviews, see Schutz, 2008, and Thompson, Graham, & Russo, 2005). Audiences can visually ascertain a performer's emotional intentions (Dahl & Friberg, 2007) and the relative size of sung intervals (Thompson & Russo, 2007), as well as extract certain structural and expressive aspects of a musical composition from an accompanying choreographed dance (Krumhansl & Schenck, 1997). Furthermore, vision can influence the perception of pitch (Gillespie, 1997; Thompson et al., 2005), loudness (Rosenblum & Fowler, 1991), and timbre (Saldaña & Rosenblum, 1993), as well as affect (Thompson, Russo, & Quinto, 2008), expressivity (Davidson, 1993, 1994), audience interest (Broughton & Stevens, 2009), phrasing (Vines, Krumhansl, Wanderley, & Levitin, 2006), performance quality (Wapnick, Darrow, Kovacs, & Dalrymple, 1997; Wapnick, Mazza, & Darrow, 1998), and lyric comprehension (Hidalgo-Barnes & Massaro, 2007). Because of the widespread nature of vision's role in music, a further understanding of the perceptual consequences of musicians' physical gestures will be useful to performers and audiences alike. It is thus possible that applied research on this topic may prove useful to music educators and to students, as well as to professional musicians interested in improving the quality of their interactions with audiences.

However, the presence of the illusion even under our manipulations suggests that these kinds of effects are not based on musical conventions or training. They establish its robust nature by demonstrating persistence even in the face of severe reductions in the degree of visual (single-dot animations vs. videos) and auditory (pure tone sounds vs. audio recordings of marimba tones) information. Because this illusion contrasts strongly with previous research on audio–visual integration, the abstract representations of the original striking gestures discussed here will be helpful in furthering our understanding of how sensory information is integrated across modalities.

completed while M.S. was in the graduate program at the University of Virginia. Correspondence should be addressed to M. Schutz at McMaster University School of the Arts, Togo Salmon Hall, 1280 Main Street West, Hamilton, ON, L8S 4L8 Canada (e-mail: schutz@mcmaster.ca) or M. Kubovy at P.O. Box 400400, University of Virginia, Charlottesville, VA 22904-4400 (e-mail: kubovy@virginia.edu).

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 using R*. Cambridge: Cambridge University Press.

BAAYEN, R. H., DAVIDSON, D. J., & BATES, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory & Language*, **59**, 390-412.

BATES, D. M., & 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. *Journal of the Optical Society of America*, **20**, 1391-1397.

BEDFORD, F. L. (2001a). Object identity theory and the nature of general laws: Commentary reply. *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 & 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. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, **37**, 137-153.

CRAWLEY, M. J. (2007). *The R book*. Chichester, U.K.: Wiley.

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

DAVIDSON, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, **21**, 101-113.

DAVIDSON, J. W. (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 & 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 & 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*, **14**, 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. doi:10.1371/journal.pone.0000943

KREFT, I. G. G., & DE LEEUW, J. (1998). *Introducing multilevel modeling*. London: Sage.

KRUMHANSL, C., & SCHENCK, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae*, **1**, 63-85.

KUBOVY, M., & SCHUTZ, M. (in press). Audio–visual objects. *Review of Philosophy & Psychology*.

LOFTUS, G. R. (2002). Analysis, interpretation, and visual presentation of experimental data. In H. Pashler (Series Ed.) & J. Wixted (Vol. Ed.), *Stevens' handbook of experimental psychology: Vol. 4. Methodology in experimental psychology* (pp. 339-390). New York: Wiley.

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

PETRINI, K., RUSSELL, M., & POLLICK, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, **110**, 432-439.

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: 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 & 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: Judgments of audiovisual temporal relations are enhanced for upright walkers. *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. (2009). Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Canadian Acoustics*, **37**, 23-28.

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

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: Sage.

SPENCE, C. (2007). Audiovisual multisensory integration. *Acoustic Science & Technology*, **28**, 61-70.

THOMAS, G. J. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, **28**, 163-177.

THOMPSON, W. F., GRAHAM, P., & RUSSO, F. A. (2005). Seeing music performance: 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 performances. *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. *International Journal of 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 & Performance*, **7**, 1327-1339.

WAPNICK, J., DARROW, A.-A., KOVACS, J., & DALRYMPLE, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, **45**, 470-479.

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: 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., & THE TASK FORCE ON STATISTICAL INFERENCE, AMERICAN PSYCHOLOGICAL ASSOCIATION (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.

NOTES

1. Designed by Simeon Fitch of Mustard Seed Software (www.mseedsoft.com).

2. Parameters were held constant within the degree of control available in our software.