

Causality and Cross-Modal Integration

Michael Schutz and Michael Kubovy
University of Virginia

Schutz and Lipscomb (2007) reported an audiovisual illusion in which the length of the gesture used to produce a sound altered the perception of that sound's duration. This contradicts the widely accepted claim that the auditory system generally dominates temporal tasks because of its superior temporal acuity. Here, in the first of 4 experiments, we show that impact gestures influence duration ratings of percussive but not sustained sounds. In the 2nd, we show that the illusion is present even if the percussive sound occurs up to 700 ms after the visible impact, but disappears if the percussive sound precedes the visible impact. In the 3rd experiment, we show that only the motion after the visible impact influences perceived tone duration. The 4th experiment (replacing the impact gestures with the written text *long* and *short*) suggests that the phenomenon is not due to response bias. Given that visual influence in this paradigm is dependent on the presence of an ecologically plausible audiovisual relationship, we conclude that cross-modal causality plays a key role in governing the integration of sensory information.

Keywords: sensory integration, causality, auditory, visual, optimal integration

Schutz and Lipscomb (2007) discovered an illusion in which a visual event affects the perceived duration of an accompanying sound. They made several videotapes of a world-renowned performer (currently a professor of percussion at the Eastman School of Music) playing single notes on the marimba. The performer played each note using two types of gestures: We will call the gesture with which he tried to produce long notes the *long gesture* and the gesture with which he tried to produce short notes the *short gesture*. When Schutz and Lipscomb asked participants¹ to judge the durations of sounds produced with these two gestures in the absence of visual information, they judged the durations of the notes to be equal. However, when they heard the sounds while watching the video, they judged notes produced by long gestures to be longer than notes produced by short gestures, despite instructions to ignore visual information when judging duration. The authors concluded that although longer gestures do not produce tones with longer acoustic durations, they do create an illusion in which tones sound longer because of auditory–visual integration.

This finding is at odds with the consensus view that in temporal tasks, auditory influence on vision is strong (Shimojo et al., 2001; Shams, Kamitani, & Shimojo, 2002), whereas visual influence on audition is either weak (Welch, DuttonHurt, & Warren, 1986; J. T. Walker & Scott, 1981; Welch & Warren, 1980) or nonexistent (Shipley, 1964). The perception of event duration (which concerns us here) is no exception (J. T. Walker & Scott, 1981; Welch & Warren, 1986). Here, we show that this generalization does not hold when the relationship between the auditory and the visual information is causal (such as when a visible impact produces a percussive sound).

The consensus view is based on a large body of research that supports the *optimal integration hypothesis*, according to which intermodal conflicts are resolved by giving more weight to the modality providing the more reliable information (Alais & Burr, 2004; Ernst & Banks, 2002). When an audiovisual conflict is spatial, vision typically dominates because the spatial acuity of the visual system is better; when it is temporal, audition dominates because the temporal acuity of the auditory system is better. An alternative theory, known as the *modality appropriateness hypothesis* (Welch et al., 1986; Welch & Warren, 1980), instead posits that conflicts are resolved by always favoring the task-relevant modality—audition for temporal tasks and vision for spatial tasks. Essentially, this can be regarded as an over-generalization of optimal integration as it assumes (incorrectly) that modality acuity fully specifies information quality.

Because audition generally offers higher quality temporal information and vision higher quality spatial information, in most cases the two theories lead to similar predictions. However, they differ when information in the generally dominant modality is ambiguous, in

Michael Schutz and Michael Kubovy, Department of Psychology, University of Virginia.

Michael Schutz is currently at the School of the Arts at McMaster University in Hamilton, Ontario.

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC 005636 to Michael Kubovy, and was performed in partial fulfillment of the requirements for an MA degree. We thank S. Fitch of Mustard Seed Software for the outstanding programs we used to run the experiments, and J. Armontrout for his devoted help in both producing stimuli and running participants. J. Armontrout, S. Boker, W. Epstein, R. Keen, T. Salthouse, J. Shatin, M. Yu gave us helpful comments on versions of the article.

Correspondence should be addressed to either Michael Schutz, School of the Arts, McMaster University, 1280 Main St. W., Hamilton, Ontario L8S 4M2, or Michael Kubovy, Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904–4400. E-mail: schutz@mcmaster.ca or kubovy@virginia.edu

¹ All participants were trained musicians (music majors at Northwestern University). However, subsequent replications of the experiment using participants that were not selected on the basis of musical training have led to similar results.

which case optimal integration has proven a more accurate framework for understanding sensory conflicts. 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). Although optimal integration applies to a wide variety of sensory conflicts (Ernst & Banks, 2002), here we focus our review primarily on the relevant auditory–visual examples.

The superior spatial acuity of vision accounts for its dominance in 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 analogs (Bertelson & Radeau, 1981; Bertelson, Vroomen, de Gelder, & Driver, 2000; Jackson, 1953; Thomas, 1941; Witkin, Wapner, & Leventhal, 1952). Likewise, the superior temporal acuity of audition can account for a wide variety of cases in which audition influences vision, but not vice versa. With respect to the Schutz–Lipscomb illusion, the most relevant finding is that visual information does not alter auditory judgments of tone duration, whereas auditory information significantly influences visual judgments of flash duration (J. T. Walker & Scott, 1981). Other instances of auditory dominance in temporal tasks include (a) perception of the number of visual flashes is affected by the number of concurrently presented tones (Shams et al., 2002), (b) the perceived rate of visual flicker is affected by the rate of concurrent auditory flutter (Shipley, 1964; Welch et al., 1986), (c) 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), and finally (d) hearing temporally discrepant auditory and visual stimuli affects the subsequent visual perception of temporal rate (Recanzone, 2003).

Causality and Unity

The conception of optimal integration in audiovisual integration that we have just reviewed cannot tell the whole story, however, because integrating across these modalities makes sense only when the perceptual system has evidence that the sights and the sounds originated from a common event. When we walk down the street, we might hear dogs barking, vehicles rumbling, and people talking. We might also see the dogs, the vehicles, and the people that make these sounds. Obviously, the perceptual system does not arbitrarily integrate sights and sounds on the basis of information quality. It only binds those that specify common events: the dog and its bark; the vehicle and its rumble; the people and their voices.

Similarly, if you listen to background music while reading this article, your comprehension might be affected by the distraction, or from leakage between modalities. But this is not an example of cross-modal integration because the perceptual system is not trying to identify the common source of the music and the text.

This requirement has been called the *unity assumption* (Welch, 1972; see also Spence, 2007; Vatakis & Spence, 2008; Vroomen, 1999; Welch, 1999; Welch & Warren, 1980) as well as the *identity decision* (Bedford, 2001a, 2001b, 2004): Whenever the congruence of two sensory inputs is great, they will be perceived to have

been caused by a single distal event, and therefore cross-modal binding will more readily occur. We believe that cross-modal systems seek identity cues to support the unity assumption, which is a necessary condition for cross-modal integration.

Proponents of the theory of optimal integration as applied to visual–haptic tasks have understood the importance of the unity assumption. In such contexts, cross-modal influences are known to be enhanced by spatial coincidence (Congedo, Lécuyer, & Gentaz, 2003; Gepshtein, Burge, Ernst, & Banks, 2005), increased ecological validity (Guest & Spence, 2003), and congruence with respect to spatial encoding strategies (Newport, Rabb, & Jackson, 2002). In addition, they are stronger when participants believe that information from the two modalities specifies a common source (Helbig & Ernst, 2007; Miller, 1972).

The importance of spatial agreement in audiovisual tasks suggests that here too the unity assumption plays a role. For example, visual influence on auditory localization judgments is weaker when the location of visual information is noticeably displaced from the apparent location of the auditory source (Jackson, 1953; Körding et al., 2007; Roach, Heron, & McGraw, 2006). Likewise, events are more likely to be perceived as simultaneous when the auditory and visual sources are spatially congruent than when they are spatially discordant (Zampini, Guest, Shore, & Spence, 2005). Although such examples suggest that spatial agreement is necessary to trigger the audiovisual unity decision, the question remains whether it is sufficient. In other words, what cues are known to lead to the perception of audiovisual unity?

At least four types have been noted:

1. **Voice–face gender congruence.** Gender congruence is a clue for integrating faces and voices. When judging the temporal order of the individual components of face–voice pairings, performance is worse when the genders of faces and voices agree. Vatakis and Spence (2007) explain that weakened performance indicates integration, which inhibits comparative analysis of the individual components. This identity cue does not seem to be triggered under all conditions. For example, the McGurk effect (in which visible lip movements alter a listener's perception of spoken syllables; McGurk & MacDonald, 1976) is insensitive to such discrepancies (Green, Kuhl, Meltzoff, & Stevens, 1991; S. Walker, Bruce, & O'Malley, 1995), as long as the lip movements and the syllables are fairly well synchronized.
2. **Synchrony.** In the McGurk effect, when the two sources are synchronized to within 180 ms (Munhall, Gribble, Sacco, & Ward, 1996), the two sources appear to trigger an identity decision, and the phenomenon is robust: (a) It is unaffected by manipulations of word meaning or sentence context (Sams, Manninen, & Surakka, 1998), (b) it is insensitive to discrepancy between the gender of the face and the voice (Green et al., 1991), and (c) it requires only a minimum of acoustic information (Remez, Fellowes, & Pisoni, 1998). However, beyond 180 ms, the worse the lip–speech synchronization, the weaker the effect (Munhall et al., 1996; Soto-Faraco & Alsius, 2009).

3. Spatial congruence. In the ventriloquist illusion, the greater the spatial distance between the origin of the sound and visual information, the weaker the effect (Jack & Thurlow, 1973).
4. Affective congruence. de Gelder, Pourtois, and Weiskrantz (2002) studied audiovisual integration in two patients with blindsight by presenting them with two classes of audiovisual stimuli: pairs with naturalistic relationships (an emotional voice accompanying an emotional face) and pairs with semantic but nonnaturalistic relationships (an emotional voice accompanying an emotional scene). Pair type was crossed with two types of congruency (congruent and incongruent), and EEG measures were recorded as patients identified the gender of the voice. Under free-viewing visual conditions, incongruence decreased the amplitude of auditory event-related potentials (ERPs) for both types of pairs. When presenting the visual stimulus to a patient's blind spot, this effect occurred only for the naturalistic pairs. Therefore, whereas both striate and nonstriate circuitry were sensitive to congruency with respect to the emotional tenor of naturalistic face-voice pairings, only striate circuitry was sensitive to this relationship for the nonnaturalistic face-scene pairings.

The observation that many of these four cues were obtained using speech stimuli raises the question of the degree to which the audiovisual identity decision pertains to nonspeech stimuli. To explore this question, Vatakis and Spence (2008) presented video clips of object action and musical notes, some of which were congruent (e.g., the sight of a piano key being struck accompanied by the corresponding sound) and some of which were not (e.g., the sight of the piano accompanied by a guitar pluck). They asked participants to report which modality stream had been presented first. Their results (as well as Radeau & Bertelson's, 1977) implied that the unity assumption does not apply to nonspeech stimuli.

Although consistent with the literature, this conclusion renders the Schutz-Lipscomb illusion even more puzzling; it suggests that this break from optimal integration cannot be explained by invoking the identity decision. However, we believe that the causal relationship between auditory and visual information specifying an impact event serves as a powerful cue for binding, sufficient to trigger the unity decision even for nonspeech stimuli. The idea that causality may play a critical role in changing integration heuristics is consistent with previous work showing that humans are highly attuned to such relationships both within (Michotte, 1963) and across (Fisher, 1962; Guski & Troje, 2003) modalities.

Binding by Causality

The most conspicuous example of cross-modal causality is the perception of a causal link between a visible impact and a percussive sound. It has been investigated by Arrighi, Alais, and Burr (2006), who created drumming videos in which they desynchronized the sound track and the video by different degrees, finding that participants perceived maximal synchrony when the audio lagged slightly behind the video. When they replaced the video of the drummer with dots whose movements matched those of the drummer, the lags required for perceived synchrony were not

meaningfully different. But when the dots moved at a constant speed (rendering them artificial and weakening their causal relationship with the tones), the required lags differed significantly.

Cross-modal causality may also play a role in Sekuler, Sekuler, and Lau's (1997) discovery of the effect of sound on the perception of bouncing: Two circles approach each other, overlap briefly and then continue on their respective paths. In this ambiguous display, the balls could be seen as either "bouncing off" or as "passing through" one another. A tone played at the moment of overlap increased the likelihood that the event was seen as a bounce. Watanabe and Shimojo (2001; see also Shimojo et al., 2001) showed that when the sound coinciding with the moment of overlap was "flanked" (preceded and followed) by identical tones 300 ms before and after, the effect disappeared. However, when the pitch of these flankers differed from the pitch of the critical tone, the effect was restored. This is because auditory grouping occurs before intersensory pairing (Keetels, Stekelenburg, & Vroomen, 2007).

The work of Stekelenburg and Vroomen (2007) supports the notion that cross-modal causality plays a central role in cross-modal integration. Using ERP as their response variable, they compared the cross-modal integration of audiovisual speech to nonspeech pairings such as an audiovisual clap or an audiovisual clink of a spoon against a cup. They found that the suppression and speeding up of the N1 component of the ERP was (a) larger for the impact-percussion pairs than for the audiovisual speech pairs, (b) unaffected by audiovisual incongruity (e.g., pairing the sound of a clap with the sight of a spoon hitting a cup), and (c) not observed for either the audiovisual tearing of paper or sawing of wood. They conjectured that the latter effect is due to the absence of visible anticipatory motion, an idea we revisit in the General Discussion.

All this suggests a hypothesis of *binding by causality*: Percussive sounds have a propensity to bind with the visible movements that could have caused them. The second finding of the study by Stekelenburg and Vroomen (2007)—that the effects observed when using incongruent hand clap-cup strike events were no different from the effects observed when using congruent events—may initially seem to undermine this notion. However, it is important to note that both types of events involve seen impacts producing percussive sounds. Therefore, it is possible that cross-event pairings based on different impact events are not perceived as "incongruent" given that they are in fact qualitatively similar. Therefore, to test the idea of a privileged relationship for causally related sights and sounds, we manipulated the degree to which the seen gesture could have produced the heard sound, first by manipulating timbre (Experiment 1), and second by the temporal alignment of the auditory and visual information (Experiment 2).

This series of four experiments addressed two alternative hypotheses to our notion of binding by causality as an explanation for the illusion reported by Schutz and Lipscomb:

- (a) *The uncertainty hypothesis*: If the participants were more certain about the duration of the impact gestures than the duration of the percussive sounds, then the illusion is in fact consistent with the optimal integration hypothesis. To evaluate this claim, we included an audio-alone condition in each experiment to examine whether, in the absence of visual information, the duration ratings of influenced sounds were more variable than the duration ratings of uninfluenced sounds.

(b) *The response-bias hypothesis*: If the gestures, by being suggestive of greater or lesser duration, affected the ratings but not the perceived durations, we would say that the illusion is due to response bias. As shown by Arieh and Marks (2008), certain patterns of cross-modal influence may in fact be wholly explained by decisional rather than sensory shifts. One way to assess this account is to explore manipulations to the original stimuli by removing the illusion. Therefore, each experiment contains at least one condition pertinent to the response-bias hypothesis, and we address the issue directly in Experiment 4.

Experiment 1: Percussive Versus Sustained Sounds

In this experiment, we asked whether the illusion, originally obtained with marimba tones, occurs with sustained sounds. If binding by causality plays a role in cross-modal integration, a gesture should only influence sounds that it could have produced: Impact gestures should affect the perception of percussive sounds but not of sustained ones.

Method

The experiment took place in a quiet room using an Apple Macintosh G4 computer running the MAX/MSP program, which controlled stimulus presentation. Stimuli were presented on a ViewSonic 19-in. E790B monitor (resolution: 1,280 × 1,024; refresh rate: 85 Hz) and Sennheiser HD 580 Precision headphones. Participants could adjust loudness during the warm-up period.

Stimuli

Each stimulus had an auditory and a visual component:

Auditory component. We used two percussive (piano, marimba) and four sustained (clarinet, French horn, voice, and white noise) sounds. As Figure 1 shows, percussive sounds began with a sharp

attack followed by an exponential decay. We used two versions (*long* and *short*) of each timbre. Their perceived durations roughly approximated those of two of the marimba tones used by Schutz and Lipscomb: E1 (~82 Hz) and D4 (~587 Hz).

Visual component. We derived the visual stimuli for the gestures from Schutz and Lipscomb, depicting a marimbist performing single marimba notes using long or short gestures on the marimba tone E1 (~82 Hz). As Figure 2 shows, the videos included the performer's head and torso in addition to the hand and arm motion involved in the impact. In the audio-alone conditions, the visual component was a blank screen.

Conditions

We presented stimuli in two conditions:

Audiovisual. We crossed the 12 sounds (6 timbres with 2 levels of duration) with the two visible gestures, for a total of 24 audiovisual stimuli.

Audio-alone. The 12 sounds were also presented alone.

Participants and Procedure

We recruited 26 participants from introductory courses in psychology; they participated for course credit. They 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. Stimuli were presented 6 times within six blocks (one presentation per block) under two conditions: (a) as audiovisual stimuli combining the visible gesture and sound, and (b) as audio alone. Block and stimuli order were both randomized.

Participants rated the duration of the sounds by using an unmarked 101-point slider (displayed on-screen after each stimuli), with endpoints *Short* and *Long*. To ensure that they were attending to the visual information, they also rated the degree to which the

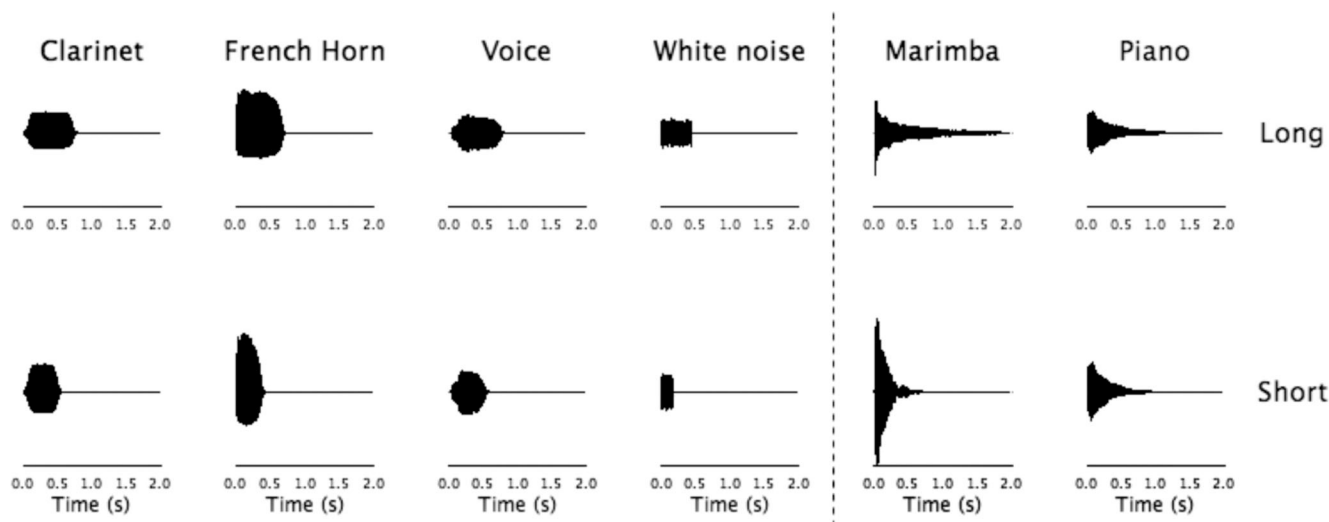


Figure 1. Amplitude envelopes of timbres used in Experiment 1, each in two versions, long (top) and short (bottom). Sounds with a continual driving force (left of the dashed line) were sustained between attack and decay, whereas percussive sounds (right of the dashed line) begin decaying immediately after the attack. Each panel shows a tone's amplitude over time on a common scale, with the y-axis suppressed so as to most clearly display tone onset.



Figure 2. The videos showed the upper body of the marimbist, including full stroke preparation and release. Reprinted from Schutz and Lipscomb (2007).

auditory and visual components of the stimulus agreed by using a second slider with endpoints *Low agreement* and *High agreement*. Previous research has shown that this approach does not interfere with the primary task (Rosenblum & Fowler, 1991; Saldaña & Rosenblum, 1993). Because the purpose of these ratings was only to draw the participants' attention to the visual component of the stimulus, we did not analyze them.

Results

Data Analyses

Our conclusions are based on linear mixed effects models (also known as multilevel analyses or hierarchical linear models) estimated by re-

stricted maximum likelihood, 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; Maxwell & Delaney, 2004, Part IV). For each set of data, we obtained estimates of effects from a *minimal adequate* (or reduced) model, which (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, and (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 standard error). To these we add a 95% confidence interval (CI), as well as a p value for a test of the null hypothesis that the effect in question is zero. 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); by minimizing the role of null hypothesis statistical tests, we implement the recommendations of the American Psychological Association Task Force on Statistical Inference (Wilkinson, 1999).

The Binding by Causality Hypothesis

Figure 3 shows that gesture length significantly affected duration ratings for three of four percussive sounds but not for any of

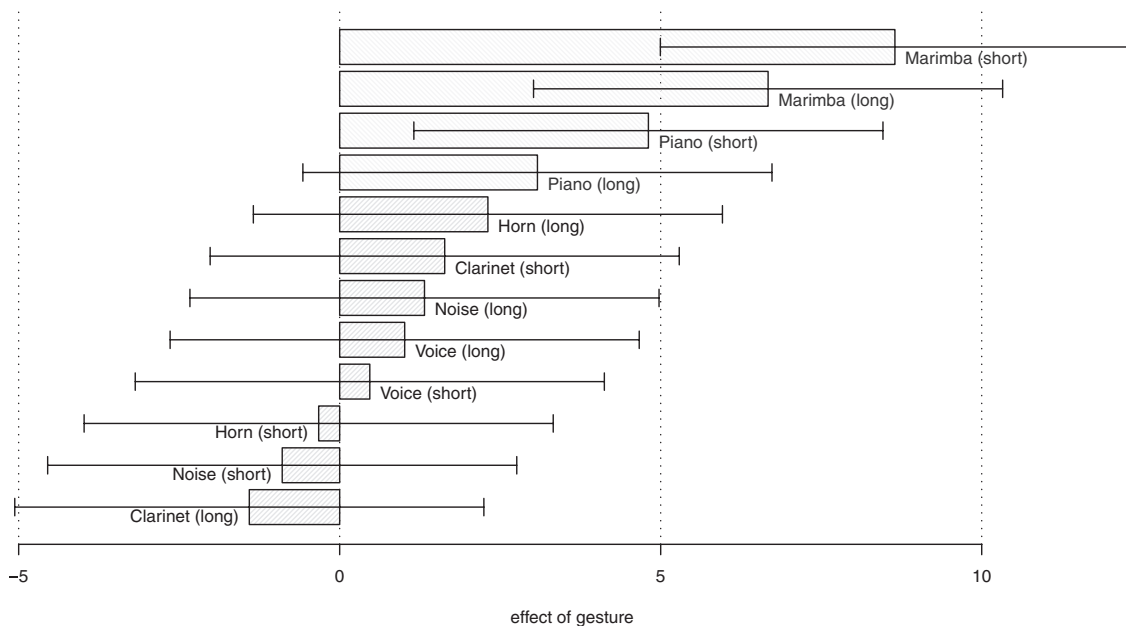


Figure 3. Experiment 1: Degree of visual influence on the 12 sounds. The x -axis represents the effect of the gesture, calculated by subtracting the ratings of a sound when paired with the short gesture from ratings of the same sound when paired with the long gesture. The gestures exerted a strong influence on the marimba sound, a moderate influence on the percussive piano sound, and no influence on the perception of sustained sounds of the clarinet, voice, French horn, and white noise. Error bars represent 95% confidence intervals. (See Figure 6 for a somewhat different analysis.)

the eight sustained sounds (two durations each of the clarinet, voice, French horn, and white noise). Furthermore, the effects on the four percussive sounds were larger than the effects on the sustained sounds. This is in line with binding by causality.

We first assessed the likelihood that the effect on the four percussive sounds would be higher than the effect on the eight other sounds. To do this, we assumed that the 12 effects were normally and identically distributed (i.e., no differential effect on percussive sounds) and determined by simulation (using the extreme-value distribution) that a particular set of four means would be higher than the other eight with a probability of $p = .002$. The average effect of gesture on the marimba sounds was 7.5 (± 3.1) points (95% CI = 1.4, 14.0; $p = .01$), higher than their average effect on the piano sounds. The average effect of gesture on percussive sounds was 6.9 (± 2.5) points higher than their effect on sustained sounds (95% CI = 2.4, 11.0; $p = .004$).

The Uncertainty Hypothesis

Uncertainty about a stimulus feature is measured by the variability of observer responses (Alais & Burr, 2004; Ernst & Banks, 2002). If the uncertainty hypothesis were true, the variability of duration ratings of percussive sounds in the audio-alone condition would predict the audiovisual data. To find out, we analyzed the ratings of the 12 audio-alone stimuli (shown in Figure 4); for each, we measured the variability of audio-alone ratings by taking the square root of standardized residuals—a standard measure for the assessment of heteroscedasticity (see Cleveland, 1993, p. 105). Figure 5 plots the effects of gesture on duration ratings (which are the same as in Figure 4) as a function of the variability of audio-alone ratings. If the uncertainty hypothesis were true, then (a) in the audio-alone condition, percussive sounds would be more

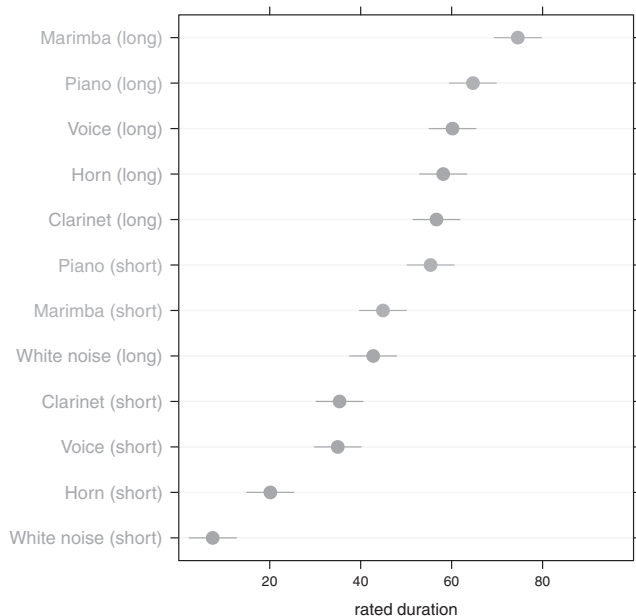


Figure 4. Duration ratings of the 12 sounds in the audio-alone trials of Experiment 1. The rating scale used ranged from 0 (short) to 100 (long). The digits that follow sound names indicate the short (1) or the long (2) version of the sound. Error bars represent 95% confidence intervals.

variable than sustained sounds. This was not the case: Percussive sounds were only 0.02 (± 0.03) points more variable than sustained sounds (95% CI = -0.04, 0.07; $p = .5$). (b) Sustained sounds would cluster in the lower left quadrant of Figure 3, and the percussive sounds would cluster in the upper right quadrant. They do not: Two of the percussive sounds are above the median variability, and two are below. (c) Finally, a regression of the audiovisual data on the variability of duration ratings of percussive sounds in the audio-alone condition gave $R^2 = .09$; $R_{\text{adj}}^2 \approx 0.2$. Another way to reach the same conclusion is the following: If the true value of this effect were zero for these eight sounds, we would expect only half of these to be positive; under this expectation, the probability that five or more are positive (i.e., three or fewer are negative) is $p = \sum_{i=1}^3 \binom{8}{i} 0.5^8 = .36$. In other words, we find no support for the uncertainty hypothesis.

The Response-Bias Hypothesis

If the gestures merely affected the ratings by suggestion, then they would have had the same effect on the sustained sounds as on the percussive sounds. That the effect on sustained sounds is negligible suggests that the difference in the magnitude of the illusion for the two types of sound cannot be due to response bias.

Effect of Perceived Sound Duration on the Illusion

To answer this question, we computed a Kendall rank correlation (τ) between the effect of gesture in the audiovisual condition (see Figure 3) and the corresponding mean ratings of duration in the audio-alone condition (see Figure 4) for all 12 sounds. The wide range of the latter (from 7.5 to 74.5) reassures us that a small value of τ is not due to restriction of range. The correlation was $\tau = .42$, which, for the null hypothesis that the two orders are independent, gives $p = .06$. This is an indication that sound duration may have an effect on the illusion, but the evidence is not strong enough to draw a conclusion.

Effect of Video on the Sensitivity of Duration Ratings

Having established that gesture type affects duration ratings, we wondered about its effect on sensitivity. Therefore, we compared ratings of duration in the audiovisual trials with ratings of duration in the audio-alone trials. These audio-alone ratings ranged widely, as we saw in Figure 4. What then is the functional relation between audiovisual and audio-alone ratings? The results, summarized in Figure 6, show that the slope of the linear function relating the audiovisual to the audio-alone trials is marginally less than 1:0.88 (± 0.08) points (95% CI = 0.73, 1.03; $p \approx .05$). This suggests that the presence of video may reduce the sensitivity of participants to the durations of the sounds.

Discussion

Experiment 1 produced two principal results: (a) The Schutz and Lipscomb illusion does not occur with sustained sounds, only with

² Here we use ≈ 0 to indicate that within the computational precision available, for all practical purposes, the value is not meaningfully different from zero.

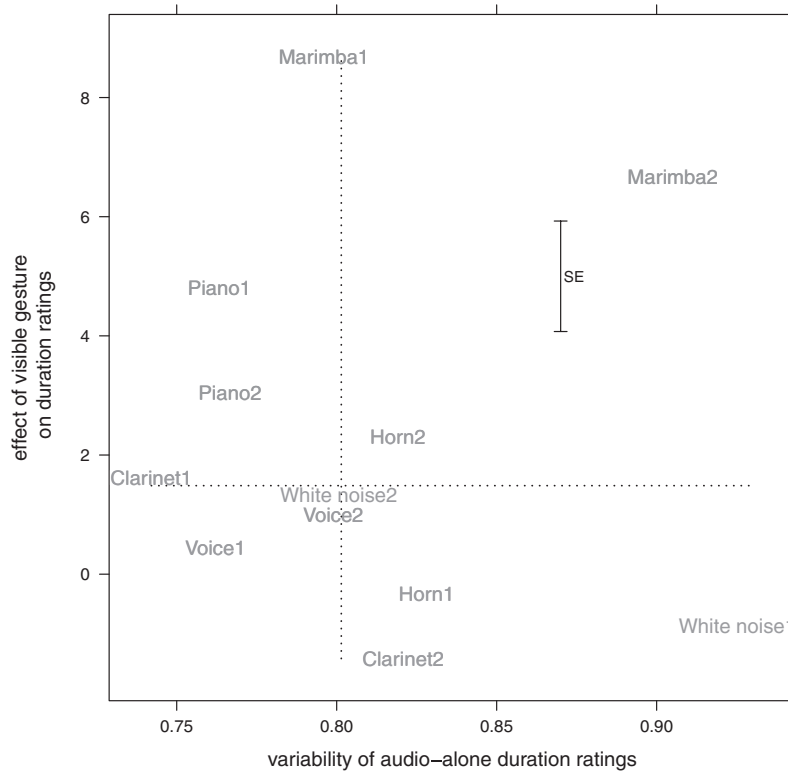


Figure 5. Evidence that uncertainty did not affect the results of Experiment 1. Visual influence on duration ratings of the 12 sounds in the audiovisual condition (y-axis) as a function of the variability of duration ratings when the same sounds were presented in the audio-alone condition (x-axis). The vertical dotted line passes through the median variability, and the horizontal dotted line passes through the median effect. If uncertainty affected the illusion, the data for the percussive sounds would cluster in the upper right quadrant and the data for the sustained sounds would cluster in the lower left. The *SE* bar is the average of the standard errors used to determine the confidence intervals in Figure 3.

percussive ones. (b) The illusion is more powerful when the type of event portrayed by the visible impact more closely matches the timbre of the accompanying sound. In other words, gestures depicting the striking of a solid object more strongly influence percussive sounds produced by striking a solid object (marimba bar) than percussive sounds produced by striking a taut string (piano).

In light of these results, we can update the hypothesis of binding by causality. The refined hypothesis stipulates that when a visible event and a sound occur in temporal proximity, $C(\text{visible event} \rightarrow \text{sound})$, and that such impressions can be ordered. If this is the case, we can safely assume that $C(\text{visible impact} \rightarrow \text{marimba}) > C(\text{visible impact} \rightarrow \text{piano}) > C(\text{visible impact} \rightarrow \text{sustained sound}) \approx 0$. If the magnitude of the illusion depends on the impression of causality, we would expect the results we obtained: the largest illusion with marimba sounds, a weaker illusion with piano sounds, and no illusion with sustained sounds.

We ruled out the uncertainty hypothesis by showing that the variability of ratings for percussive and sustained sounds did not differ in the audio-alone condition, and did not predict the effects we obtained in the audiovisual condition. Likewise, we ruled out the response-bias hypothesis by showing that the effect of gesture

on sustained sounds is negligible. Finally, we noted that the slope of the linear function relating audiovisual and audio-alone trials may be less than 1.0, which—if confirmed in other experiments—would be evidence that dividing the participants' attention between the auditory and visual sources degrades their ability to discriminate among sounds.

Experiment 2: The Effect of Asynchrony

If the hypothesis of binding by causality is correct, then disruptions of the temporal ordering weakening the causal relationship between modalities should weaken the strength of the Schutz–Lipscomb illusion. Accordingly, in this experiment, we manipulated the temporal relationship between the auditory and visual information to examine the relative size of the illusion when the visible impact (a) was synchronous with the sound, (b) preceded the sound, and (c) followed the sound.

Method

The experiment was identical to Experiment 1, except in the ways we describe next.

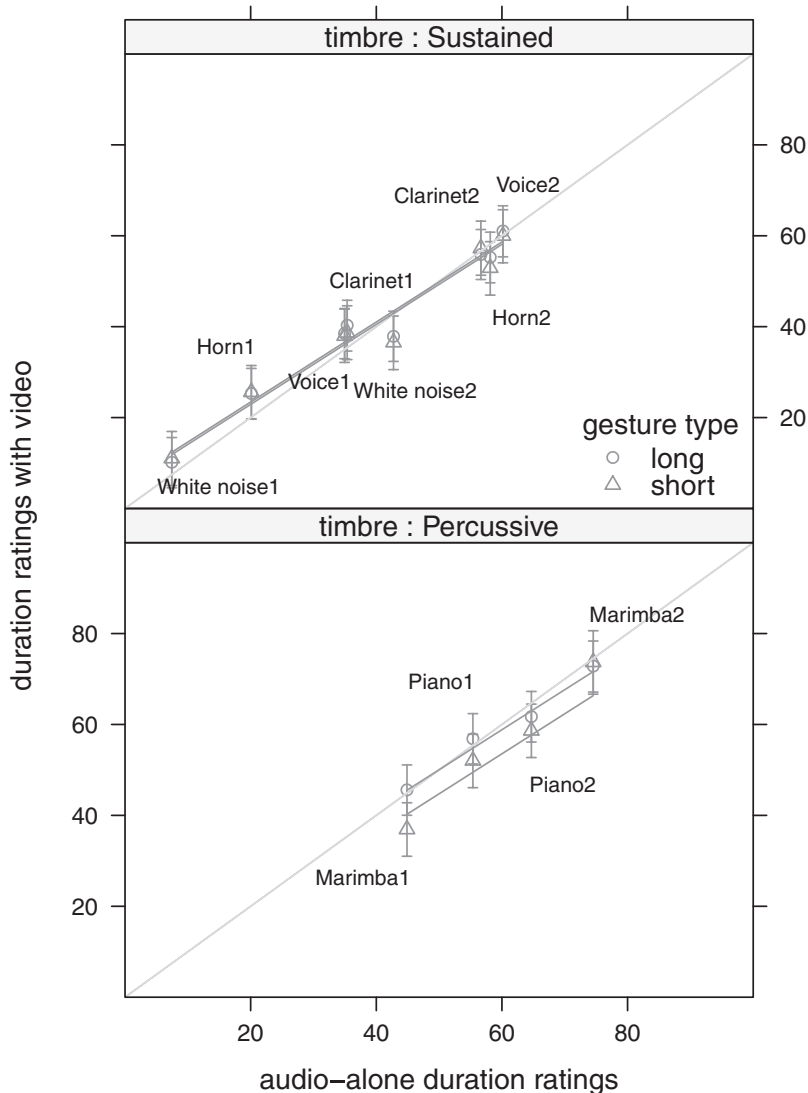


Figure 6. Duration ratings for the 12 sounds for the audiovisual trials in Experiment 1 as a function of ratings of identical sounds in the audio-alone condition, by timbre class (sustained vs. percussive) and visible gesture (long vs. short). Error bars represent $\pm 1 SE$. The lines represent the best-fitting linear regression functions.

Stimuli

The stimuli were derived from the marimba videos and the marimba sound tracks used in Experiment 1.

Auditory component. We used marimba tones of different durations (shown in Figure 7), which we controlled by manipulating (a) *sound termination* by using either tones that decayed naturally, or tones that were manually damped soon after the bar was struck, and (b) *natural decay time*. Because bars tuned to lower frequencies ring longer (Bork, 1995), we varied the frequencies of the marimba sounds. We crossed the two types of sound termination (damped, natural) with three musical pitches: E1 (~82 Hz), D4 (~587 Hz), and G5 (~1568 Hz).

Visual component. The original long and short gestures from the previous experiment served as the visual stimuli (see Figure 2).

The onset of the sound (a) was synchronous with the visible impact, (b) preceded it by 400 or 700 ms, or (c) followed it by 400

or 700 ms. Thus, the levels of offset were: -700, -400, 0, 400, 700 ms (negative values denote the audio-first conditions).³ Synchrony was manipulated by altering the frame in which the tone began sounding (the videos used a recording rate of 30 frames per second). For example, in the 400-ms offset condition, tone onset occurred 12 frames later than in the original (synchrony) condition.

Conditions

We presented the sounds alone or with the video.

Participants and Procedure

We recruited 10 new participants, and paid them \$10 for their participation. They went through 264 trials, organized in five

³ These timings are accurate within a 33-ms window.

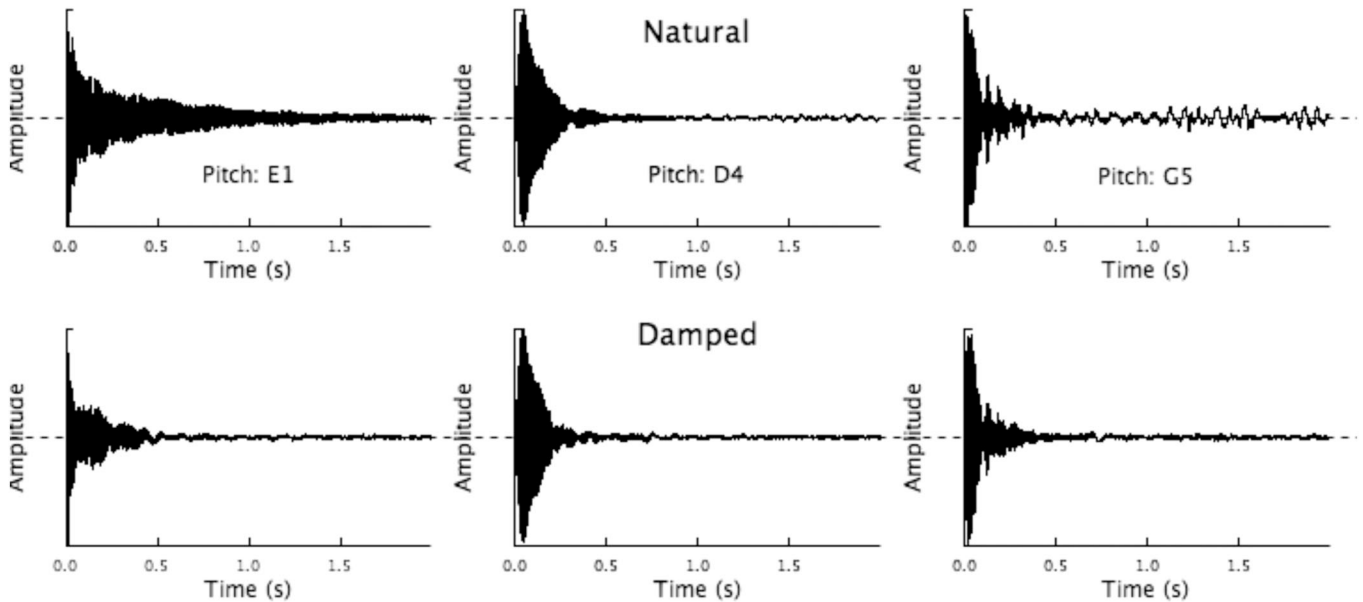


Figure 7. Natural (top row) and damped (bottom row) marimba tones used as auditory stimuli for Experiment 2.

blocks (four audiovisual blocks of 60 trials and one audio-alone block of 24 trials). Each block contained one exemplar of each stimuli (presented in a random order), with the order of blocks within the experiment randomized as well.

Results

The Binding by Causality Hypothesis

As Figure 8 shows, the illusion was absent in the audio-first conditions: The average effect in the audio-first conditions was 7.2 (± 4.4) points (95% CI = -1.1, 16.3; $p = .84$). It was present in the synchrony and the video-first conditions. The effect in the synchrony condition was larger than in the video-first conditions: The difference between the effect at synchrony and the average effect in the video-first conditions was 10.8 (± 2.1) points (95% CI = 6.8, 14.9; $p \approx 0$).

The Uncertainty Hypothesis

As in Experiment 1, the sounds varied widely in perceived duration in the audio-alone condition (see Figure 9). To determine whether the variability of these duration ratings predicted the degree to which they were visually influenced, we plotted the effect of the gesture as a function rating variability (see Figure 10). As is evident from the trend line (labeled “mean”), the magnitude of the illusion in the audiovisual condition does not increase substantially as a function of the variability of audio-alone ratings. (We note that the range of the measures of variability was similar to the range for Experiment 1: 0.74 to 0.95.)

The Response-Bias Hypothesis

Our analysis here is parallel to the one for Experiment 1, in which we found no evidence for response bias. Here, however, even though we found no significant effect in the audio-first conditions (-400 ms, -700 ms in Figure 8), we do have some evidence of response bias. If

the true value of this effect were zero for these 12 sounds, we would expect only half of them to be positive in our data; yet, the effect of gesture was positive for the 12 audio-first conditions, an extremely unlikely event: $p = .5^{12} = .00024$.

Effect of Perceived Sound Duration on the Illusion

As shown in Figure 9, the mean perceived duration of these sounds varied widely: from 25 to 87 (a range of 62, compared with a range of 67 in Experiment 1). This variability allowed us to explore whether tone duration had any effect on the strength of the illusion. In other words, was there consistency among effect sizes (Falissard, 2008) for the different tones across each of the offsets? To this end, we used the intraclass correlation (ICC; Shrout & Fleiss, 1979)—a measure normally used to examine interrater reliability—to examine the reliability of the degree of visual influence on each tone. We obtained a mild degree of disagreement—ICC = -0.12 (95% CI = -0.21, -0.0001; $p \approx .05^4$; computed by bootstrap simulation), indicating the illusion was not strongly affected by tone duration (only a strong positive rating would have suggested otherwise).

Effect of Video on Duration Ratings

Having established that audiovisual asynchrony affects the illusion, we wondered—as we did in Experiment 1—what effect the presence of video might have had on duration ratings. To answer this question, we compared ratings of duration in the audiovisual trials with ratings of duration in the audio-alone trials. These audio-alone ratings ranged widely, as we saw in Figure 9. What then is the functional relation between audiovisual and audio-alone

⁴ Here, we use $\approx .05$ to indicate that within the computational precision available; the effect is, for all practical purposes, marginally significant.

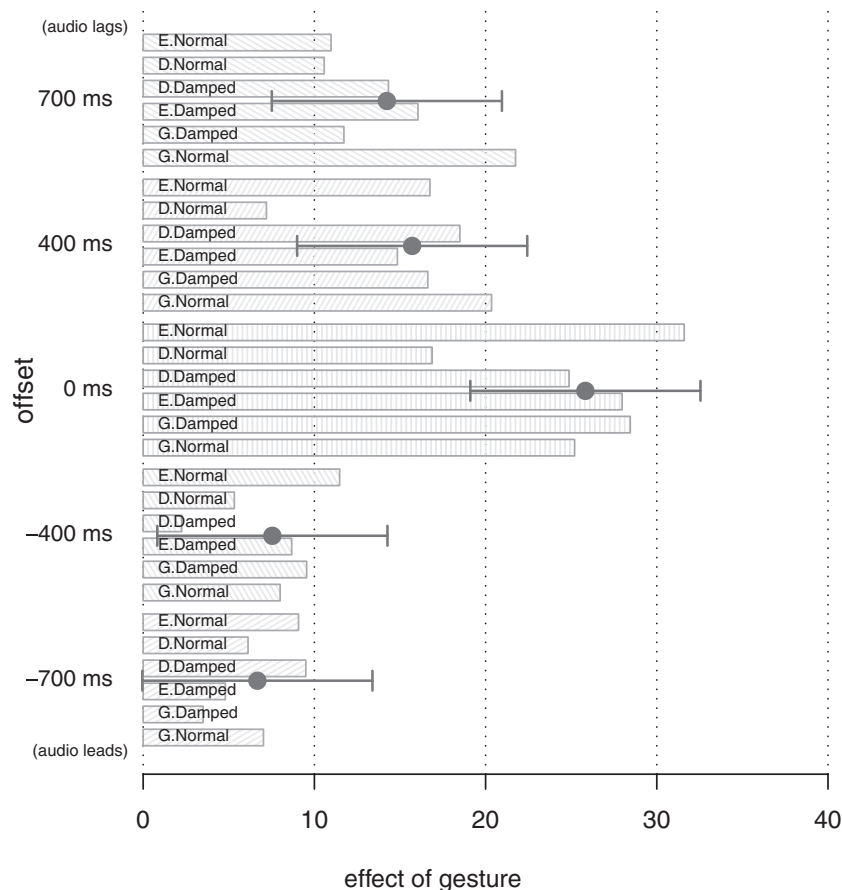


Figure 8. Effect of gesture as a function of auditory–visual asynchrony in Experiment 2. The six bars in each asynchrony group represent the marimba sounds of different durations, in ascending order of perceived duration. The illusion was greatest when the visible impact and the onset of the marimba sound were in sync (offset = 0 ms), smaller when the visible impact preceded the sound (offset \geq 0 ms), and not significant when the sound came first (offset \leq 0 ms). The perceived duration of the sounds had no effect on the magnitude of the illusion (i.e., the variation in the magnitudes of the six bars is not consistent between levels of offset). Error bars represent 95% confidence intervals.

ratings? The results, summarized in Figure 11, show that the slopes of the linear functions relating the audiovisual to the audio-alone trials are less than 1 (although not necessarily significantly so): 0.92 (95% CI = 0.55, 1.3), 0.95 (95% CI = 0.58, 1.3), 0.72 (95% CI = 0.46, 0.99), 0.81 (95% CI = 0.46, 1.2), and 0.82 (95% CI = 0.43, 1.2). It is noteworthy that the offset for which the effect of gesture is largest (offset = 0), the slope is significantly lower than 1 (i.e., its confidence interval does not contain it), albeit marginally so. As in Experiment 1, this suggests that participants were less sensitive to differences in sound duration in the presence of visual information. In addition, it hints at the intriguing possibility that this decrease in sensitivity is modulated by the quality of the link between the visual and the auditory information.

Discussion

Experiment 2 produced three principal results: (a) The Schutz and Lipscomb illusion did not occur when the percussive sound preceded the visible impact. (b) The illusion was weakened but

present when the sound followed the visual impact. (c) The illusion was still present when the sound came 400 ms or 700 ms after the visible impact.

The literature would lead us to expect the first two results: Auditory–visual cross-modal effects are generally weaker when the sound precedes the visual information than when it is simultaneous with the visual information or follows it (in the McGurk effect by Munhall et al., 1996; in the audiovisual perception of drumming by Arrighi et al., 2006; in temporal ventriloquism by Stekelenburg & Vroomen, 2007). The third result is unexpected because the cross-modal effects just cited do not persist beyond a delay of 200 ms. This suggests that a stronger form of cross-modal binding may be occurring in this situation.

It is worth noting that that a delay of 700 ms is not ecologically impossible. At 340 m/s, sound can travel 238 m in 700 ms, 2.2 or 2.3 times a football or a soccer field. Under free-field conditions, a loud sound would be audible at that distance (say, the loudness of a pneumatic hammer 1 m away, roughly 106 dB-SPL, will have

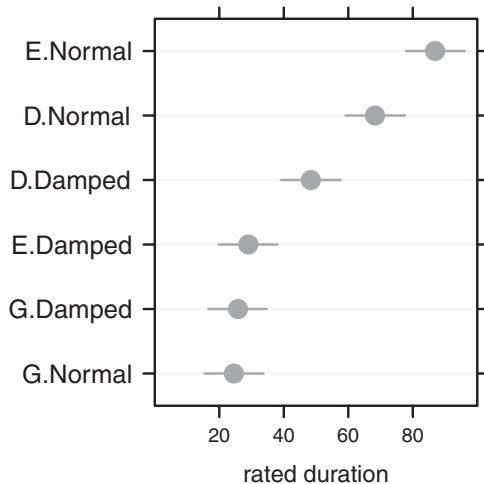


Figure 9. Duration ratings of the six sounds for the audio-alone trials of Experiment 2. Error bars represent 95% confidence intervals.

dropped to the loudness of a quiet restaurant, roughly 48 dB-SPL at 256 m). This observation may be relevant to questions related to the perceived simultaneity of auditory and visual information traveling over distances large enough to induce temporal discrepancies (King, 2005). Results are mixed, with evidence both supporting (Alais & Carlile, 2005; Kopinska & Harris, 2004; Sugita & Suzuki, 2003) and refuting (Lewald & Guski, 2004; Stone et al., 2001) the notion that the perceptual system compensates for differences between the speed of light and sound when gauging onset simultaneity.

As in Experiment 1, these results offer no support for the uncertainty hypothesis's prediction that visual influence is a function of auditory ambiguity. Although the variability of duration ratings differed considerably between the tones, the magnitude of vision's influence was independent of these differences (see Figure 10). Consequently, the uncertainty hypothesis and, therefore, the notion of optimal integration cannot account for these data.

Finally, let us look at our evidence regarding the response-bias hypothesis. We noted that the effects obtained for the negative offsets were positive for all 12 sound–offset combinations, and that the probability of such an event is negligible. Two accounts of this finding can be proposed: (a) There is a small visual effect on auditory judgments even when the sound precedes its visible cause, or (b) the video induced a bias response when rating the duration of sounds preceding it. Our data do not offer a resolution.

Whichever of these accounts may be the case, the main findings of this experiment support the binding by causality hypothesis: The illusion is contingent upon causality. The illusion occurred only when there was a plausible visual–auditory causal link. Furthermore, as with Experiment 1, the effect was graded: It was strongest when the causal link was most plausible (synchrony), weaker when the causal link was less likely (audio-lag), and very weak when the causal link was impossible (audio-lead).

Experiment 3: Which Part of the Gesture Is Responsible for the Illusion?

In Experiment 3, we determine which portion of the gesture (pre- or post-impact) is responsible for the illusion.

Method

The experiment was identical to Experiments 1 and 2, except in the ways we describe next.

Stimuli

We used the auditory and visual stimuli from the 0-ms offset of Experiment 2 to create a segment (pre-impact) video that shows the gesture prior to the impact and freezes when the sound begins and a segment (post-impact) video that starts frozen on the moment of impact, then displaying the post-impact gesture beginning at the onset of the marimba tone. The segment (both) stimulus consisted of the original videos with the complete gesture.

Auditory component. Sounds consisted of long, short, and damped marimba sounds performed at two pitch levels: pitch:E1 (~82 Hz) and pitch:D4 (~587 Hz), which were used in Experiment 2.

Visual component. The long and short gestures used in Experiment 2 served as the visual stimuli.

Participants and Procedure

Twenty-nine new participants from introductory courses in psychology received course credit for their participation. Each went through 324 trials in 18 blocks: 6 audiovisual blocks (36 trials each), 6 audio-alone blocks (6 trials each), and 6 video-alone blocks (12 trials each). Both the order of blocks and the order of trials within each block were randomized. During the video-alone condition (used only in this experiment), participants were asked to rate the relative duration of the gesture on the same scale used for the audiovisual and audio-alone presentations.

Results

The left panel of Figure 12 summarizes the results of the main part of the experiment, the audiovisual condition. It shows that the visual influence is due to the post-impact segment: Gestures affected ratings only when presented concurrently with the sound (e.g., the both and the post-impact conditions). The right panel summarizes the results of the video-alone condition. It shows that the participants could tell which gesture was long and which was short from the information in both segments, although the difference was larger when the post-impact segment was visible.

Importance of the Post-Impact Portion of the Gesture

In the post-impact segment in the audiovisual condition, the gesture exerted a 12-point (± 2.1) influence on duration ratings (95% CI = 8.5, 16.5; $p \approx 0$). This is only slightly less than its influence when both segments (the whole gesture) were visible: a 14-point (± 2) effect. In contrast, the pre-impact segment showed a negligible 3.1-point (± 2.1) effect (95% CI = ~ 1.0 , 7.1; $p = .13$). This lack of influence does not reflect visual ambiguity, as illustrated by the 21.9-point (± 5.5) difference between ratings of the pre-impact gestures when presented as video-alone (95% CI = 11, 33; $p \approx 0$).

Difference Between Segments

The effect of the post-impact portion of the gesture in the audiovisual condition was 9.4 (± 1.6) points larger (95% CI = 6.3,

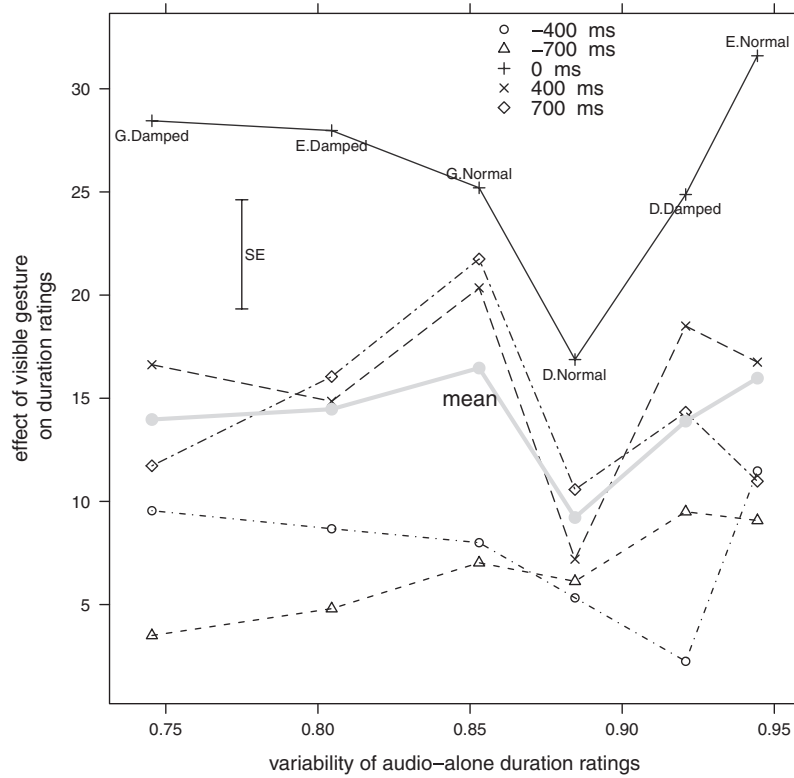


Figure 10. Evidence that uncertainty was unrelated to degree of influence in Experiment 2. Effect of gesture on the duration ratings of the six sounds in the audiovisual trials as a function of the variability of duration ratings in the audio-alone trials, plotted separately based on degree of offset. The gray line labeled “mean” represents the mean effect for each sound, collapsed across the five levels of offset. It does not show the upward trend one would expect if the uncertainty hypothesis were true. The ordering of the variability of the ratings of sounds differs from the ordering of their durations (see Figure 9). The SE bar is the average of the standard errors used to determine the confidence intervals in Figure 8.

12.5; $p \approx 0$) than the effect of the pre-impact portion of the gesture. In contrast, in the both condition, the gesture had a negligibly larger effect than in the post-impact condition: a 1.7-point (± 1.6) difference (95% CI = $\sim 1.5, 4.9$; $p = .3$), again reflecting that the post-impact segment of the gesture was responsible for the visual influence.

The gestures were discriminable in the video-alone condition, as shown in Figure 12. There was a 49-point (± 5.5) effect of gesture length (95% CI = 38, 59; $p \approx 0$) in the post-impact segment and a slightly larger 53-point (± 5.5) difference (95% CI = 42, 64; $p \approx 0$) when both segments were visible. The effect for the pre-impact segment was smaller, 21.9 (± 5.5) points (95% CI = 11, 33; $p \approx$

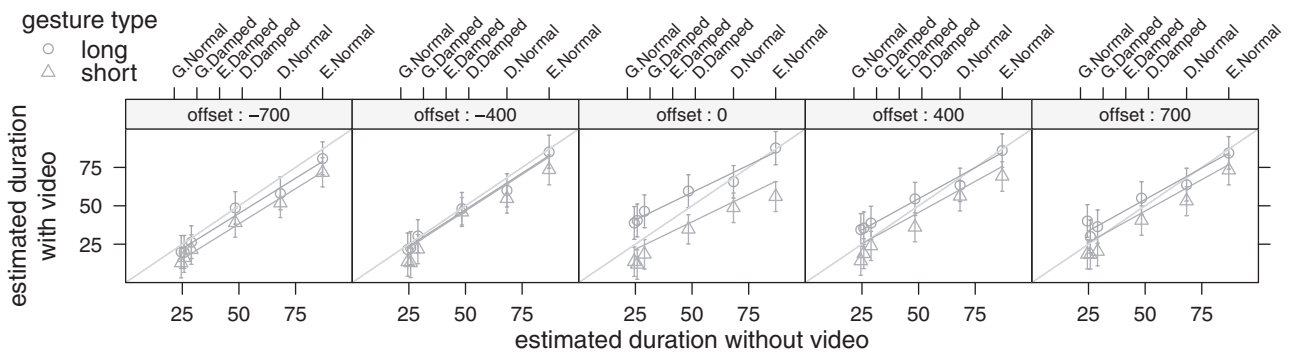


Figure 11. Duration ratings for the six sounds for the audiovisual trials in Experiment 2 as a function of ratings of identical sounds in the audio-alone condition, by offset and visible gesture (long vs. short). Error bars represent ± 1 SE. The lines represent the best-fitting linear regression functions.

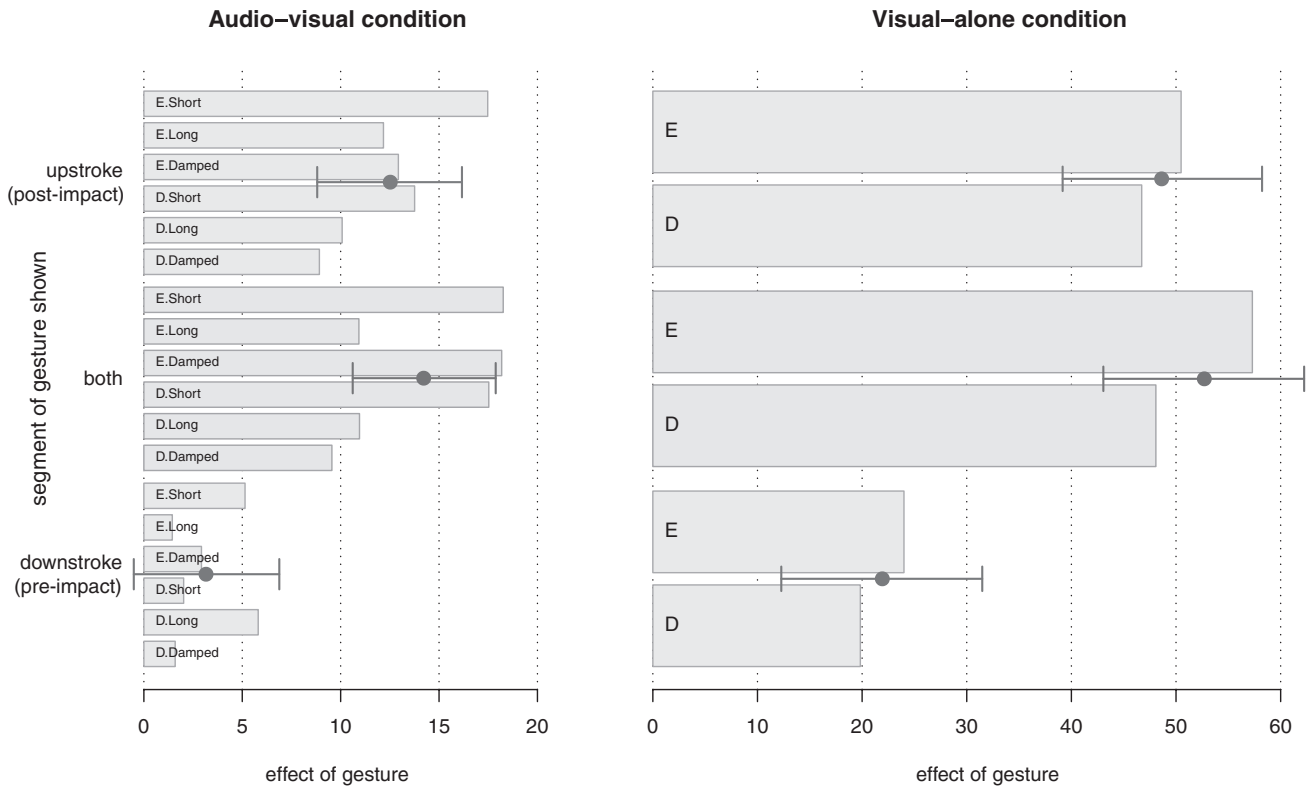


Figure 12. Experiment 3. *Left—the audiovisual condition:* Effect of gesture as a function of segment (pre-impact, post-impact, both). The six bars in each group represent the ratings of each marimba sound, in ascending order of perceived duration. The illusion was large when the post-impact portion of the gesture was visible, and much smaller when it was not (e.g., the post-impact condition). The perceived duration of the sounds had no effect on the magnitude of the illusion (i.e., the variation in the magnitudes of the six bars is not consistent between levels of segment). *Right—the video-alone condition:* Ratings of gesture length as a function of gesture type. The two bars in each group show ratings for the gestures used on two different pitch levels. The stroke length was discriminable for both segments, but the information in the post-impact segment was more diagnostic. Note the difference in scale between the left and right panels. Error bars represent 95% confidence intervals.

0), but still larger than the largest audiovisual influences in this experiment.

The Uncertainty Hypothesis

As in the previous experiments, the ratings of the durations of the sounds in the audio-alone condition varied widely (see Figure 13). Figure 14 shows the effect of gesture as a function of the variability of these ratings. It is evident from the trend line (labeled “mean”) that the magnitude of the illusion in the audiovisual condition does not increase as a function of the variability of audio-alone ratings. (We note that the range of the measures of variability was similar to the range for the previous experiments: 0.73 to 0.94).

The Response-Bias Hypothesis

As in Experiment 2, although we found no significant effect in the pre-impact condition (see Figure 12), we do have some evidence of response bias. If the true value of this effect were zero for these six sounds, we would expect only half of them to be positive in our data; that the effect of gesture was positive for these six conditions is unlikely: $p = .5^6 = .016$.

Effect of Perceived Sound Duration on the Illusion

As in Experiment 2, we tested whether the order of effects for the six sounds (see Figure 12) was consistent across video segments. (As Figure 13 shows, the mean perceived durations of these sounds varied widely: from 23 to 78, a range of 55, compared with 67 and 62 in the first two experiments.) We computed the ICC and obtained an ICC = .6 (95% CI = -.2, .8; $p \approx .05$; computed by bootstrap simulation); there is no evidence of agreement, and we conclude that perceived sound duration did not have an effect on the illusion.

Effect of Video on Duration Ratings

In this experiment too, audio-alone ratings ranged widely, as we saw in Figure 13. The functional relation between audiovisual and audio-alone is summarized in Figure 15. It shows that the slope of the linear function relating the audiovisual to the audio-alone trials is definitely less than 1: 0.78 (± 0.02) points (95% CI = 0.76, 0.81; $p > .05$). This suggests that the presence of video reduced the sensitivity of participants to the durations of the sounds.

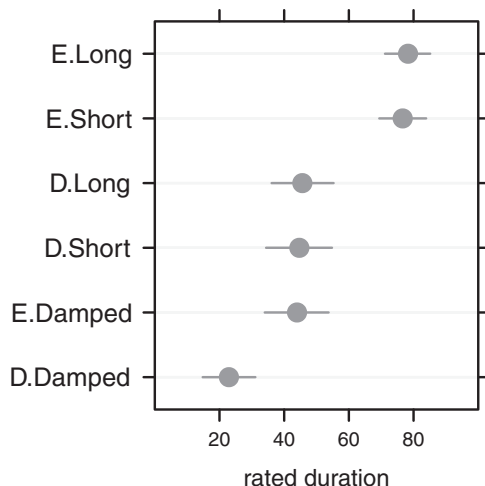


Figure 13. Duration ratings of the six sounds for the audio-alone trials of Experiment 3. Error bars represent 95% confidence intervals.

Discussion

These results suggest that the illusion is driven by the post-impact portion of the gesture: the visual information presented concurrently with the sound. Two observations support this claim: (a) The effect of the pre-impact gesture is much smaller than the effect of the post-impact gesture. (b) The effect of gesture is just about the same when the entire gesture is visible as when only the post-impact segment is visible. The slight visual influence in the pre-impact condition is analogous to the effects observed in the audio-first conditions of Experiment 2.

An alternative interpretation to our claim that the post-impact portion of the gesture governs the illusion is that participants were actually imagining the pre-impact gesture when seeing only the post-impact portion. However, it seems unlikely that participants were imagining the pre-impact gesture when seeing the post-impact portion without also doing the reverse (imagining the post-impact gesture when seeing only the pre-impact portion).

It would be interesting to compare these results with those of participants who had never seen the pre-impact portion of the stroke before evaluating the post-impact segments. Nevertheless, as long as the illusion remained substantially larger in the post-impact condition, our conclusions regarding the importance of this segment would remain unchanged. In fact, subsequent research using a single moving dot mimicking the striking gestures used here also found post-impact motion to be more important than pre-impact motion, even when using full gestures (Armontrout, Schutz, & Kubovy, 2009).

This experiment provides further evidence against the uncertainty hypothesis: The uncertainty in the rating of the durations of the sounds in the audio-alone condition was unrelated to the magnitude of the illusion in the audiovisual condition (see Figure 14). This further supports the hypothesis of binding by causality. When an event yields both visible and audible effects, the two are perceptually bound as a result of their causal relationship, and can influence each other even though the quality of the information in each modality would suggest otherwise.

Finally, we revisit the response-bias hypothesis. We noted that the six effects obtained for the pre-impact condition were positive, and that the probability of such an event is small. Just as in Experiment 2, we can propose two accounts of this finding: (a) The visible difference between long and short gestures had an effect on the duration ratings, or (b) this information produces a 3-point response bias on the rating of the duration of the sound that preceded it. Our data do not allow us to adjudicate between the two accounts.

Experiment 4: A Test of the Response-Bias Hypothesis

Even though the three experiments described so far offer compelling evidence against the response-bias hypothesis, we designed Experiment 4 as a final test. We compared the effect of the videotaped gesture to the effect of suggestive text—the written words *long* and *short*. Assuming that text cannot alter the perception of concurrent auditory information, if we found an influence of text similar to that of the gesture, then a top-down influence could adequately explain the illusion, eliminating any need for a hypothesis of binding by causality.

Method

This experiment followed the methodology used in the first three experiments.

Stimuli

We used the same stimuli as in the synchrony condition of Experiment 2. We created the text condition by replacing the long and short gestures with the words *long* and *short*.

Auditory component. The sounds were the same as in Experiment 2: two types of marimba tones (natural, damped) performed at three pitches: E1 (~82 Hz), D4 (~587 Hz), and G5 (~1568 Hz).

Visual component. There were two visual conditions: (a) display (video), the long and short gestures used in Experiment 2; and (b) display (text), in which we replaced the videos of gestures with the text *long* and *short* written in black on white. The text was visible for the duration of the original videos. This maximized the chance of a visual influence: The marimba tones began approximately 1 to 1.5 s after the text, giving participants the time to read it before hearing the sound.

Participants and Procedure

Twenty-four new participants from introductory courses in psychology received course credit for their participation. We discarded the data of 2 participants who gave the same response on all trials. Stimuli were presented six times each, in blocks organized into two conditions: (a) audiovisual and (b) audio alone. The seven blocks included three blocks of display:gesture (72 trials), three blocks of display:text (72 trials), and one block of audio-alone (36 trials), for a total of 180 trials per participant.

Results and Discussion

For the sake of brevity, we focus our analysis mostly on assessing whether the text condition affected duration ratings. Figure 16

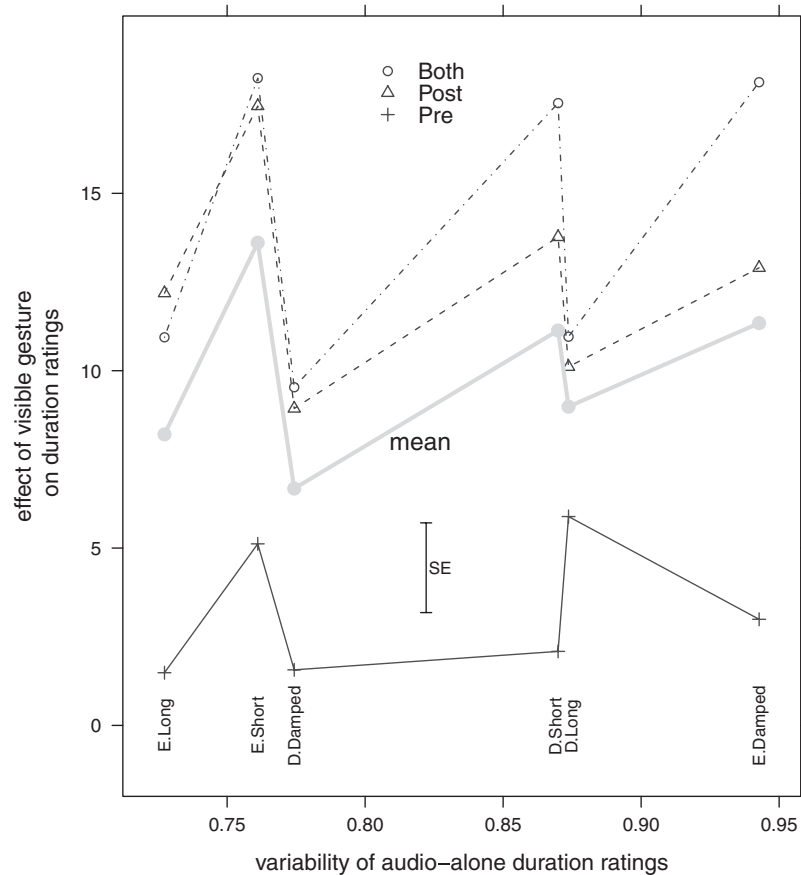


Figure 14. Evidence that uncertainty did not affect the results of Experiment 3. Effect of gesture on the duration ratings of the six sounds in the audiovisual trials as a function of the variability of duration ratings in the audio-alone trials and the segment(s) shown. The gray line labeled “mean” represents the mean effect for each of the six sounds. It does not show the upward trend one would expect if the uncertainty hypothesis were true. The ordering of the variability of the ratings of sounds differs from the ordering of their magnitudes (see Figure 13). The *SE* bar is the average of the standard errors used to determine the confidence intervals in Figure 12.

summarizes the results. It shows that the visual influence was larger with the videos than the text.

The 4.7-point (± 2.9) effect of the text was negligible (95% CI = $-0.9, 10.5$; $p = .1$), whereas the 10.5-point (± 2.9) effect of the video (95% CI = $4.7, 16.2$; $p = .002$) tells us that it affected the ratings. The latter was 5.7 (± 2.5) points larger (95% CI = $0.8, 10.5$; $p = .02$) than the effect of the text. (There is only weak evidence that the ratings in the text condition were greater than zero: $p = 5 \times 0.56 = .08$.) Therefore, we conclude that it is improbable that the illusion is accounted for by bias.

The visual influence from the gesture was smaller than that observed in the previous experiments. This may be due to that fact that the written text explicitly called attention to our visual manipulation more directly than the visible gestures, giving participants greater conscious awareness of the role of the visible gestures in their judgments of tone duration. If so, these results would parallel those of Schwarz and Clore (1983), who found that although one’s current affective state influences evaluations of life satisfaction, this influence is eliminated when the source of this affective information is brought to conscious awareness.

General Discussion

We summarize our studies in Table 1 and Figure 17. Before assessing three possible accounts of the Schutz–Lipscomb illusion (binding by causality, uncertainty/optimal integration, response bias), we articulate the three main conclusions that can be drawn from this series of studies:

1. The illusion is conditioned on causality. The visible gesture affects only the perceived duration of sounds it could have caused. The illusion replicated with percussive sounds but not sustained sounds in Experiment 1, and under the synchrony but not the (ecologically impossible) audio-lead condition in Experiment 2.
2. The strength of the illusion is proportional to the degree of audiovisual plausibility. The degree of influence reflects the relative probability that the visible gesture could have caused the sound. The largest influence occurs when it is most likely (marimba timbre in Experiment 1; synchrony condition in Experiment 2), a smaller

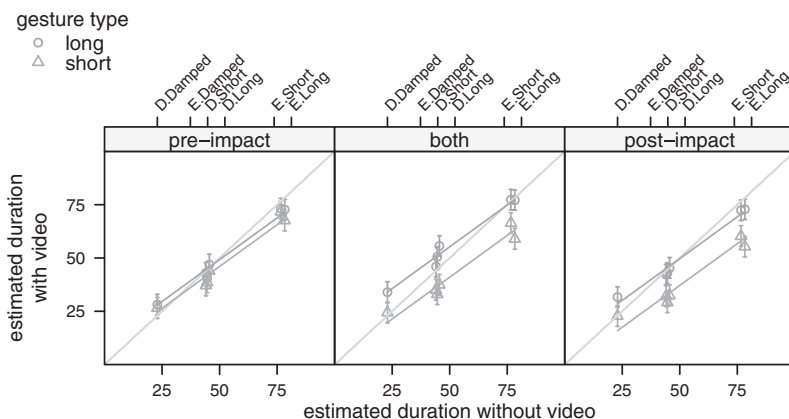


Figure 15. Duration ratings for the six sounds for the audiovisual trials in Experiment 3 as a function of ratings of identical sounds in the audio-alone condition, by segment (downstroke, both, and upstroke) and visible gesture (long vs. short). Error bars represent $\pm 1 SE$. The lines represent the best-fitting linear regression functions.

but notable influence occurs when it is *possible* (piano timbre in Experiment 1; audio-lag condition in Experiment 2), and no meaningful influence occurs when the gesture could *not* have caused the sound (non-percussive timbres in Experiment 1; audio-lead condition in Experiment 2).

3. The post-impact gesture is paramount. By and large, the component of the gesture affecting perceived duration is the one concurrent with the sound (Experiment 3).

The Uncertainty Hypothesis

The uncertainty hypothesis (e.g., the traditional notion of optimal integration) argues that the Schutz–Lipscomb illusion stems from uncertainty regarding the duration of percussive sounds. Because they decay gradually without a clear offset, this seems a plausible account, parsimoniously reconciling the illusion with the vast literature on optimal integration. Experiments 1–3 explored this idea by comparing the variability of ratings (the standard index of perceptual uncertainty) for sounds presented in the audio-alone condition against the degree to which these same sounds were visually influenced in the audiovisual condition. As uncertainty was unrelated to influence in each case, the uncertainty hypothesis, and by extension the notion of optimal integration, cannot account for these data.

The Response-Bias Hypothesis

The response-bias hypothesis states that the basis of the Schutz–Lipscomb illusion is decisional, rather than sensory. Given the inherent difficulty of distinguishing between the two (Arieh & Marks, 2008), such an account is not unreasonable. Like uncertainty, a response-bias account would parsimoniously reconcile the Schutz–Lipscomb illusion with the literature on optimal integration, removing any need to resort to the notion of binding by causality.

Experiment 4 ruled out the most blatant form of such bias, the direct effect of words on duration ratings. Nevertheless, it does

not address a more subtle form of the hypothesis: that the effect requires relevant gestural input, but that it is not mediated by perception. Indeed, in two experiments, we found unmistakable evidence of a weak effect of gesture in conditions with a weak visual–auditory causal link: in the -700 -ms and the -400 -ms conditions of Experiment 2 and in the pre-impact condition of Experiment 3. However, this in itself is not sufficient to support the response-bias hypothesis given that the magnitude of the illusion in these conditions was significantly less than the magnitude of the illusions’ replication (included in each experiment). Therefore, although it is possible that there may be an *element* of bias in the original illusion, a bias-alone explanation cannot account for its size nor its sensitivity to manipulations breaking the causal connection between modalities.

The Binding by Causality Hypothesis

As neither the uncertainty nor the response-bias hypotheses can fully explain our results, we believe that binding by causality represents the only coherent explanation of these data and, by extension, the Schutz–Lipscomb illusion. (Note: The idea that causality is an important principle governing integration is not without precedent—see our previous discussions of the “identity decision” and section on binding by causality.) From these experiments, it is clear that the detection of a causal relationship changes the process of integration in a manner incompatible with a traditional understanding of optimal integration.

However, that is not to say that the idea of optimal integration is utterly wrong—merely that the definition of *optimal* may depend on the nature of the event in question in addition to the quality of information available. Under those circumstances in which visual information is clearly relevant, it is not unreasonable to weight its input more strongly, irrespective of its relative quality. We consider the extent to which such special cases provide insight into the design, function, and day-to-day operations of cross-modal systems to be fruitful questions for future research.

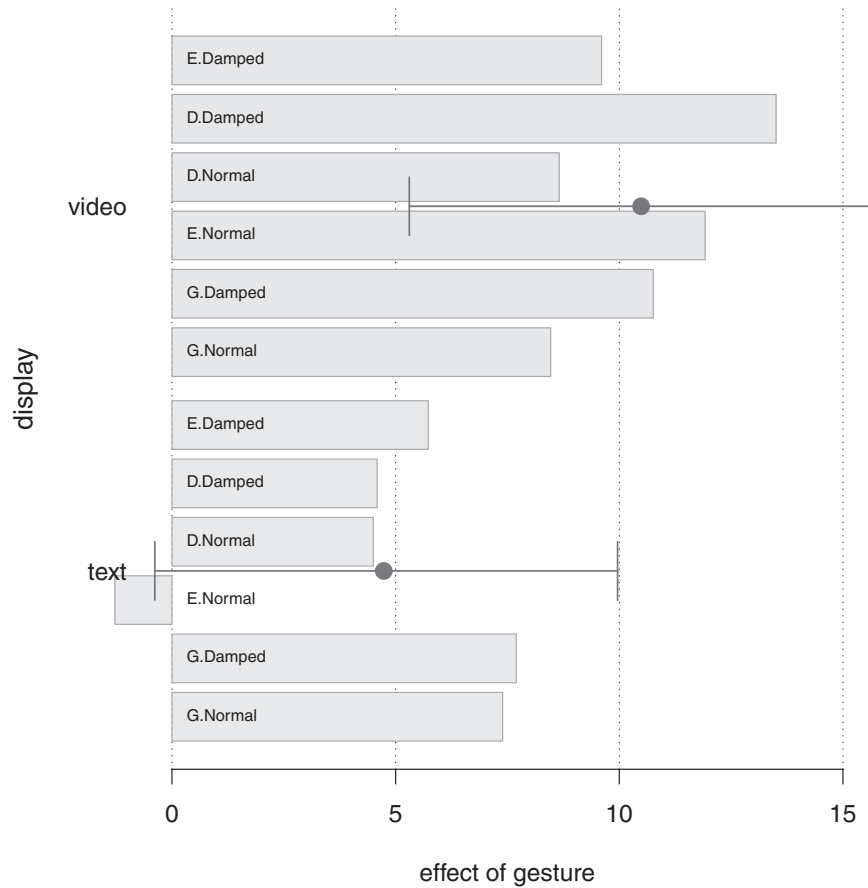


Figure 16. Experiment 4. Effect of gesture as a function of display type (video or text). The six bars in each group represent the marimba sounds of different durations, in ascending order of perceived duration. The illusion was large when the video was shown, and much smaller when text was shown. The perceived duration of the sounds had no effect on the magnitude of the illusion. Error bars represent 95% confidence intervals.

Additional Observations

We have evidence that one cannot counteract the effects of binding by causality, as our experiments suggest that one does not have the option to voluntarily undo the cross-modal influences that follow. We gave our participants instructions to ignore gestures when rating tone duration, but they were able to do so only when it was unlikely that the gesture caused the tone.

In addition, we found that participants were less sensitive to differences in auditory duration in the presence of visual information. In each experiment, the slopes of the linear relation between ratings obtained in the audiovisual condition and the audio-alone condition were less than 1. This reduction in sensitivity may reflect competition for attentional resources when auditory and visual information must be independently evaluated (Duncan, 1980) or when switching attention between modalities (Spence, Nicholls, & Driver, 2001).

The small visual influences observed in the absence of a fully specified causal link (e.g., the sustained sounds in Experiment 1 or the audio-lead condition in Experiment 2) may in fact reflect the presence of a weaker form of binding requiring only partial cues (e.g., either temporal proximity or agreement regarding event type). Such an explanation could bridge the gap between our data and the results of previous experiments exploring cross-modal influences using tone beeps and light flashes that lack any relationship beyond their temporal proximity (Sekuler et al., 1997; Shams et al., 2002; J. T. Walker & Scott, 1981; Welch & Warren, 1980), and would be consistent with the previously mentioned literature on the “identity decision” (Bedford, 2001a, 2001b; Spence, 2007; Vatakis & Spence, 2008; Vroomen, 1999;

Table 1
Summary of Design for Experiments 1–4

Experiment	Strength of the causal relationship		
	Strong	Weak	None
1	Marimba (original)	Piano	Sustained (french horn, clarinet, voice, white noise)
2	Synchrony (original)	Audio lags	Audio leads
3	Both	Post-impact	Pre-impact
4	Video		Text

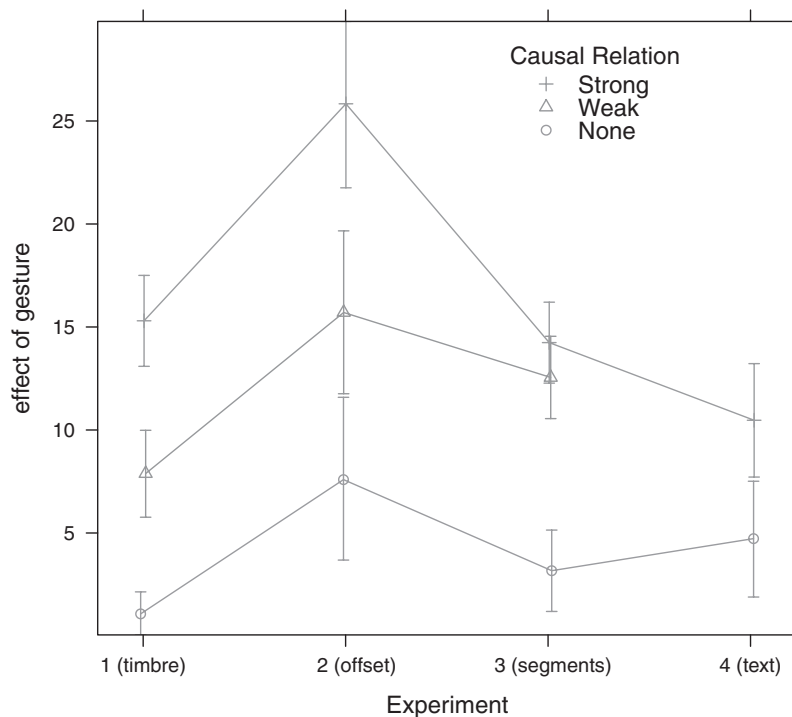


Figure 17. Summary of the experiments by condition. Conditions thought to represent a response bias are represented with circles, conditions replicating the original experiment with crosses, and “other” conditions (e.g., piano timbre in Experiment 1 or audio-lag condition in Experiment 2) with triangles. Error bars represent ± 1 SE.

Welch & Warren, 1980). Such an account would also explain why vision does not affect auditory judgments of tone duration when information from each modality shares only the partial cue of temporal proximity (J. T. Walker & Scott, 1981), whereas it does when sharing both the cues of temporal proximity and event-type agreement. At this point, such an explanation is obviously speculative, but it does suggest possibilities for future work to better explore the role of causality in sensory integration.

Finally, we must point out two limitations of this work, also suggestive of future research: (a) we have not shown that other types of causal links can trigger cross-modal binding and hence cross-modal influences; and (b) we do not know whether visual events that could *not* have produced percussive sounds would fail to integrate. Our hypothesis could be extended into a theory by finding a range of gestures that are integrated only with those sounds that they could have caused. For example, it may be possible to observe binding of sustained tones (e.g., those of a clarinet or French horn) and the motions perceived to have caused them. Finding that these same motions fail to bind with percussive sounds (which they could not have caused) would broaden the work we have presented here, which is based solely on impact gestures and largely on percussive sounds. Because visual information is known to affect ratings of sung intervals (Thompson, Russo, & Quinto, 2008) and judgments of tension in clarinet performances (Vines, Krumhansl, Wanderley, & Levitin, 2006), it is not unlikely that such audiovisual pairs could be discovered.

Final Remarks

Although the origin of the privileged relationship between impact gestures and percussive sounds is not yet clear, we believe that it is likely a perceptual adaptation rather than a learned association. Percussive sounds produced by impacts are common in our environment (branches breaking, rocks falling, footsteps, etc.), and were likely just as important to our evolutionary ancestors as they are to us today. Our results may also reflect cross-modal Gestalt principles (Spence, Sanabria, & Soto-Faraco, 2007) integrating auditory and visual information into a single “impact event.” It is also possible that this relationship is learned (or strengthened) from repeated exposure to impact events common in our environment, such as the slamming of car doors, dropping of objects, and sound of feet stepping (Saygin, Driver, & de Sa, 2008).

Regardless of its origin, it makes sense for the perceptual system to treat artificial, ecologically unrelated information such as tone and light pairs by defaulting to a “best information wins” heuristic, while reserving a privileged override for multimodal information clearly specifying a common cause. This approach captures the best of both worlds, deferring to the stronger modality except when there is a compelling reason to do otherwise. Such a design favors information utility over information quality, which is not uncommon in the design of our perceptual system—for example, our greater sensitivity to high frequencies allows for better detection of the most important frequency information (Fletcher & Munson, 1933).

In conclusion, the discovery of the illusion by Schutz and Lipscomb (2007) was prompted by a debate among two schools of thought about marimba performance. Some thought that by modulating one's gesture, one could lengthen the duration of marimba notes; others claimed it could not as this is physically impossible. As is so often the case, those who were wrong in theory were right in practice. And ultimately, this failed attempt at the physically impossible was not only aesthetically powerful, but psychologically informative.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.
- Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences, USA, 102*, 2244–2247.
- Arieh, Y., & Marks, L. E. (2008). Cross-modal interaction between vision and hearing: A speed–accuracy analysis. *Perception and Psychophysics, 70*, 412–421.
- Armontrout, J., Schutz, M., & Kubovy, M. (2009). Visual determinants of a cross-modal illusion. *Attention, Perception, and Psychophysics, 71*, 1618–1627.
- 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, England: 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*. Retrieved June 5, 2008, from <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. *Journal of the 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 and 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 and Psychophysics, 62*, 321–332.
- Bork, I. (1995). Practical tuning of xylophone bars and resonators. *Applied Acoustics, 46*, 103–127.
- Cleveland, W. S. (1993). *Visualizing data*. Murray Hill, NJ: AT&T.
- Congedo, M., Lécuyer, A., & Gentaz, E. (2003). The influence of spatial delocation on perceptual integration of vision and touch. *Presence: Teleoperators and Virtual Environments, 15*, 353–357.
- Crawley, M. J. (2007). *The R book*. New York: Wiley.
- de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of the Sciences, USA, 99*, 4121–4126.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*, 272–300.
- Ernst, M. O., & Banks, M. S. (2002, January 24). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.
- Falissard, B. (2008, February 16). *The psy package* (Version 0.7). Retrieved March 23, 2008, from <http://cran.r-project.org/web/packages/psy/psy.pdf>
- Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception and Psychophysics, 63*, 719–725.
- Fisher, G. H. (1962). Phenomenal causality in conditions of intrasensory and intersensory stimulation. *American Journal of Psychology, 75*, 321–323.
- Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America, 5*, 82–108.
- Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision, 5*, 1013–1023.
- Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics, 50*, 524–536.
- Guest, S., & Spence, C. (2003). Tactile dominance in speeded discrimination of textures. *Experimental Brain Research, 150*, 207–207.
- Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. *Perception and Psychophysics, 65*, 789–800.
- Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual–haptic integration. *Perception, 36*, 1523–1533.
- 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.
- 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.
- Keetels, M., Stekelenburg, J., & Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. *Experimental Brain Research, 180*, 449–456.
- King, A. J. (2005). Multisensory integration: Strategies for synchronization. *Current Biology, 15*, R339–R341.
- Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception, 33*, 1049–1060.
- Kreft, I., & Leeuw, J. de. (1998). *Introducing multilevel modeling*. London: Sage.
- Lewald, J., & Guski, R. (2004). Auditory–visual temporal integration as a function of distance: No compensation for sound-transmission time in human perception. *Neuroscience Letters, 357*, 119–122.
- Loftus, G. R. (2002). Analysis, interpretation, and visual presentation of experimental data. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology: Vol. 4. Methodology in experimental psychology* (3rd ed., 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.
- McGurk, H., & MacDonald, J. (1976, December 23). Hearing lips and seeing voices. *Nature, 264*, 746–748.
- Michotte, A. (1963). *The perception of causality* (T. R. Miles & E. Miles, Trans.). New York: Basic Books.
- Miller, E. A. (1972). Interaction of vision and touch in conflict and

- nonconflict form perception tasks. *Journal of Experimental Psychology*, 96, 114–123.
- Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception and Psychophysics*, 58, 351–362.
- Newport, R., Rabb, B., & Jackson, S. R. (2002). Noninformative vision improves haptic spatial perception. *Current Biology*, 12, 1661–1664.
- Radeau, M., & Bertelson, P. (1977). Adaptation to auditory–visual discordance and ventriloquism in semirealistic situations. *Perception and Psychophysics*, 22, 137–146.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Linear mode: Applications and data analysis methods*. London: Sage.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89, 1078–1093.
- Remez, R., Fellowes, J., & Pisoni, D. (1998). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication*, 26, 65–73.
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audiovisual integration. *Proceedings of the Royal Society B: Biological Sciences*, 273, 2159–2168.
- 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.
- Sams, M., Manninen, P., & Surakka, V. (1998). McGurk effect in Finnish syllables, isolated words and words in sentences: Effects of word meaning and sentence context. *Speech Communication*, 26, 75–87.
- 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., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36, 888–897.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513–523.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997, January 23). Sound alters visual motion perception. *Nature*, 385, 308.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147–152.
- Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., & Watanabe, K. (2001). Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science and Technology*, 22, 61–67.
- Shipley, T. (1964, September 18). Auditory flutter-driving of visual flicker. *Science*, 145, 1328–1330.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk–MacDonald illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 580–587.
- Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28, 61–71.
- Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception and Psychophysics*, 63, 330–336.
- Spence, C., Sanabria, D., & Soto-Faraco, S. (2007). Intersensory Gestalten and crossmodal scene perception. In K. Noguchi (Ed.), *Psychology of beauty and Kansei: New horizons of Gestalt perception* (pp. 519–579). Tokyo: Nihon University College of Humanities and Sciences.
- Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, 19, 1964–1973.
- Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B: Biological Sciences*, 268, 31–38.
- Sugita, Y., & Suzuki, Y. (2003, June 26). Implicit estimation of sound-arrival time. *Nature*, 423, 911.
- Thomas, G. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, 28, 163–175.
- Thompson, W. F., Russo, F., & Quinto, L. (2008). Audiovisual integration of emotional cues in song. *Cognition and Emotion*, 22, 1457–1470.
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception and 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. Müseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 389–393). Amsterdam: Elsevier.
- Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audiovisual 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.
- Walker, S., Bruce, V., & O’Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception and Psychophysics*, 57, 1124–1133.
- Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, 12, 109–116.
- Welch, R. B. (1972). The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception and 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. Müseler (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 and Psychophysics*, 39, 294–300.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638–667.
- Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, chapter 25). New York: Wiley-Interscience.
- 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.
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audiovisual simultaneity judgments. *Perception and Psychophysics*, 67, 531–544.

Received November 10, 2007

Revision received October 21, 2008

Accepted November 11, 2008 ■