Published in final edited form as:
Psychol Sci. 2005 March ; 16(3): 228–235.

Hearing What the Eyes See:

Auditory Encoding of Visual Temporal Sequences

Sharon E. Guttman, **Lee A. Gilroy**, and **Randolph Blake**
Vanderbilt University

Abstract

When the senses deliver conflicting information, vision dominates spatial processing, and audition dominates temporal processing. We asked whether this sensory specialization results in cross-modal encoding of unisensory input into the task-appropriate modality. Specifically, we investigated whether visually portrayed temporal structure receives automatic, obligatory encoding in the auditory domain. In three experiments, observers judged whether the changes in two successive visual sequences followed the same or different rhythms. We assessed temporal representations by measuring the extent to which both task-irrelevant auditory information and task-irrelevant visual information interfered with rhythm discrimination. Incongruent auditory information significantly disrupted task performance, particularly when presented during encoding; by contrast, varying the nature of the rhythm-depicting visual changes had minimal impact on performance. Evidently, the perceptual system automatically and obligatorily abstracts temporal structure from its visual form and represents this structure using an auditory code, resulting in the experience of "hearing visual rhythms."

People's interpretation of the world depends on information delivered through multiple senses. Over the past 40 years, numerous studies have investigated how human perceptual systems coordinate this varied input to create a unified description of reality. Early work suggested that vision predominates in multisensory processing: When visual information conflicts with information from other sensory modalities, vision typically "wins" (e.g., Hay, Pick, & Ikeda, 1965; Howard & Templeton, 1966).

Recent work, however, converges on a more balanced view: Although vision dominates audition for processing spatial information, audition often dominates vision for processing temporal information (e.g., Bertelson & Aschersleben, 1998; Kitagawa & Ichihara, 2002; Kitajima & Yamashita, 1999; Recanzone, 2003; Repp & Penel, 2002; Wada, Kitagawa, & Noguchi, 2003). For example, a repetitive sound (auditory flutter) presented simultaneously with a flickering light causes the rate of perceived visual flicker to shift toward the auditory flutter rate (e.g., Shipley, 1964; Welch, DuttonHurt, & Warren, 1986). Known as *auditory driving*, this tendency for visual flicker to become perceptually synchronized with auditory flutter occurs even though the flutter and flicker rates are easily distinguished when presented in succession (Recanzone, 2003). Auditory dominance of temporal perception also shows up in the most simple of experiences: A single flash and single audible click occurring in close temporal proximity shift perceptually toward temporal coincidence, with misperception of the visual stimulus accounting for the bulk of the shift (Fendrich & Corballis, 2001).

The conditions yielding dominance of vision versus dominance of hearing likely reflect divergent specializations of visual and auditory processing, respectively. According to the

Address correspondence to Sharon E. Guttman, Department of Psychology, Vanderbilt University, 111 21st Ave. South, 301 Wilson Hall, Nashville, TN 37203; e-mail: sharon.guttman@vanderbilt.edu.

modality-appropriateness hypothesis (Welch, 1999; Welch & Warren, 1980), perception gives precedence to the "best" sensory modality for the task at hand: vision for spatial judgments and audition for temporal judgments. Intersensory conflicts are resolved through subjugation of the less reliable sense—as implied by auditory driving—and possibly even through sensory recalibration (Kitagawa & Ichihara, 2002; Recanzone, 1998, 2003).

Largely ignored, however, has been another potential consequence of sensory specialization: cross-modal encoding of unisensory input into the "appropriate" modality. Might people "see" the spatial layout of an auditory array or—as we investigated—"hear" the temporal structure of visual changes? While concentrating on visual sequences consisting of temporally random contrast changes, we noticed a natural tendency to hear the temporal sequencing of these changes as well. Of course, the auditory referencing of visual events is not an entirely new experience; for example, many people engage in subvocalized speech when reading text. However, what we encountered when watching visual sequences had a markedly different flavor: It arose automatically, unintentionally, and without learning or practice. Most notably, we could not ignore the auditory rhythm implied by the visual changes.

In the current article, we present three experiments examining the reality of "hearing visual rhythms." Specifically, these experiments investigated the idea that rhythm (technically termed temporal structure) portrayed solely by visual input receives automatic, obligatory encoding in the auditory domain.

Previous research on modality effects in rhythm processing indicates that auditory rhythmic stimuli produce better shortterm memory than comparable visual stimuli (Glenberg, Mann, Altman, Forman, & Procise, 1989), particularly when musically structured (Glenberg & Jona, 1991). These effects could be attributable to the use of modality-specific temporal codes, which are more reliable in audition than in vision, or to the cost of translating visual stimuli into a different format (either auditory or amodal in nature). Collier and Logan (2001) tested these hypotheses explicitly by having participants match two rhythmic sequences either within or across sensory modalities. Performance in all conditions converged at slower presentation rates, suggesting gradual translation to a universal code. Even at the fastest rate, however, performance in the cross-modal conditions—which contained one (presumably well-encoded) auditory sequence—typically did not exceed performance with two visual sequences. From these findings, Collier and Logan concluded that rapidly presented rhythmic sequences become encoded in a modality-specific manner, thus introducing a cost for comparing rhythms across different modalities.

In the experiments we report here, the idea of rapid, obligatory cross-modal encoding is addressed using a paradigm probably more sensitive to underlying representations of temporal structure. In our experiments, observers performed a same/different discrimination task concerning the rhythms of two visual sequences. We assessed temporal representations by measuring the extent to which both task-irrelevant auditory information and task-irrelevant visual information interfered with performance.

In interference paradigms, the extent of task disruption reflects the extent to which the representation of the to-be-encoded information overlaps with the representation of the irrelevant information. If visually presented temporal sequences automatically become represented in an auditory manner, then incongruent auditory information should impair processing of the visual stimuli. By contrast, if comparison of two visual temporal sequences utilizes visual representations, then incongruent auditory signals should have minimal effect; however, manipulations that disrupt the visual similarity of the two sequences—even if along a task-irrelevant dimension—should disrupt processing.

EXPERIMENT 1

This first experiment examined whether task-irrelevant auditory stimulation disrupts encoding of visual temporal sequences. Observers performed a sequence-discrimination task in silence, in the presence of congruent auditory information, or in the presence of incongruent auditory information.

This experiment and the others that follow differ in several ways from previous investigations of modality effects in rhythm processing. First, earlier experiments used predominantly slow, musiclike rhythms (e.g., Collier & Logan, 2001; Glenberg & Jona, 1991; Glenberg et al., 1989). By contrast, we used stimuli with stochastic temporal structure whose successive beats, while creating distinct, perceptible temporal structure, occurred too closely in time for controlled, strategic recoding. Therefore, any auditory-visual interactions observed may be attributed to rapid, automatic processes. Second, visual stimuli in previous experiments consisted of sequences of simple, repetitive light flashes. Our experiments used complex visual stimuli in which *changes* signified the "beats." These stimuli impart rich visual information (e.g., contrast summation over time) in addition to the relevant temporal structure, which may bias performance against the use of auditory codes, if such codes exist. Finally, unlike earlier work, our experiments intermixed different trial types unpredictably, precluding shifts in strategy based on trial type. Thus, representations revealed should be those arising naturally and unintentionally.

Method

Observers—The three authors and 7 naive observers participated in this experiment. All 10 observers had normal or corrected-to-normal vision and normal hearing.

Apparatus—Testing occurred in a darkened, quiet room. The visual stimuli appeared on a MultiSync XE monitor (21 in.; 1024 × 768 pixels; 75 Hz) positioned at eye level, 80 cm from the observers' eyes. The sound stimuli were generated by a pair of Apple Pro speakers, positioned to either side of the monitor.

Stimuli—Figure 1 contains a schematic depiction of the visual and auditory stimuli. The visual stimuli consisted of vertically oriented, even-symmetric Gabor patches (frequency = 0.75 cycles/deg; *SD* = 0.5°; contrast = 80%), with a visible diameter of approximately 2.5°. Over time, the Gabor patches reversed in contrast; the timing of the contrast reversals defined a visual "rhythm."[1]

Each visual rhythm consisted of 17 frames, presented at a frequency of 9.4 Hz (i.e., 106.7 ms/frame). Following the initial frame, contrast reversals occurred on 8 of the 16 remaining frames, resulting in rhythms with 8 visual beats. The distribution of the beats depended on a random point process, with the constraints that no more than 4 visual beats could occur consecutively and no more than 4 consecutive frames could be presented without a visual beat.

On *same* trials, identical visual rhythms appeared within the two sequential presentations (see Fig. 1a). On *different* trials, the two visual rhythms were defined by different random point processes with a correlation of 0 (see Fig. 1b); hence, the timing of any given beat, relative to the start of the sequence in question, bore no relation to the timing of beats in the other sequence. However, both rhythms started and ended with the same Gabor patch, such that only the timing of the contrast reversals differentiated the two sequences.

[1]We use the term "rhythm" here for simplicity; however, we should reiterate that the stimuli did not contain musiclike structure.

Auditory stimuli consisted of a sequence of eight clicks (2 ms, 800 Hz, 60 dB) following random point processes like those defining the visual sequences. On *congruent* trials, the auditory clicks precisely matched the timing of the visual changes in each of the two sequences (see Fig. 1a). On *incongruent* trials, the auditory clicks followed point processes that were uncorrelated with the visual changes (see Fig. 1b); the auditory sequences could be the same or different over the two intervals. On no-sound trials, the visual sequences appeared without any auditory accompaniment.

Design—The nature of the auditory sequence (congruent with visual sequence, incongruent with visual sequence, or no sound) was the key independent variable. All observers participated in four experimental sessions, each containing 10 *same* and 10 *different* trials of each type.

Procedure—Each trial consisted of the sequential presentation of two visual or auditory-visual sequences, separated by a 1,600-ms interstimulus interval (ISI). The various trial types appeared in random order, and no cue revealed which type to expect.

Observers were instructed to ignore the auditory information and to indicate, by pressing one of two keys, whether the two visual sequences had the same or different temporal structure. An auditory “ping” provided feedback for incorrect responses.

The screen was blank during the ISI and after the second sequence, until response. A fixation cross appeared 2 s after response, at which time observers could initiate the next trial by pressing the space bar.

Results and Discussion

Figure 2 depicts observers' ability to match the visual rhythms in the various auditory conditions, collapsed over *same* and *different* trials. Clearly, the presence and nature of the auditory sequence significantly affected task performance, $F(2, 18) = 90.4$, $p < .001$, $\eta^2 = .91$. Planned comparisons confirmed that the nosound condition differed from both the congruent sound and the incongruent sound conditions, $t(9) = 7.8$, $p < .001$, $\eta^2 = .87$, and $t(9) = 7.0$, $p < .001$, $\eta^2 = .84$, respectively. Incongruent auditory stimulation significantly interfered with observers' ability to track a visual rhythm. This cross-modal interference cannot be attributed to the presence of sound per se: When congruent with the visual rhythm, auditory stimulation enhanced task performance. Furthermore, we doubt that these results are attributable to criterion effects (Marks, Ben-Artzi, & Lakatos, 2003): A d′ analysis—which revealed observers' ability to discriminate *same* trials from *different* trials irrespective of response bias—yielded qualitatively similar results.[2]

Might the effects simply reflect auditory “capture” of the visual rhythms—alteration of the visual input to coincide with the auditory input? On incongruent sound trials, whether the two auditory sequences were the same as or different from one another significantly influenced observers' responses: As would be expected on the basis of auditory capture, more responses matched the similarity of the auditory sequences than would be expected by chance (62.1% vs. 50.0%), $t(9) = 6.0$, $p < .001$, $\eta^2 = .80$. However, results from ancillary experiments do not support auditory capture as the basis of the cross-modal interference effect. In one experiment, observers judged whether or not the visual changes within given sequences occurred in synchrony with the superimposed auditory clicks. Observers experienced no confusion between the visual and auditory beats (mean performance = 98.4% correct). This finding may be contrasted with experiments on auditory driving (Recanzone, 2003), in which

[2]Similar d′ analyses rule out decisional biases as the explanation for the results of Experiments 2 and 3.

simultaneously presented visual flicker and auditory flutter appear to occur at the same rate, despite being distinguishable if presented sequentially. In another follow-up study, analogous to Experiment 1, we had observers track auditory rhythms presented either in isolation, with congruent visual sequences, or with incongruent visual sequences. Visual input had virtually no impact on performance, $F(2, 18) = 1.2$, $p = .33$, $\eta^2 = .11$, with observers averaging 94.6% correct performance across all conditions. The high fidelity of auditory rhythm tracking suggests that visual rhythm tracking during concurrent auditory stimulation—if dependent on auditory capture—should have reflected the similarity of the two auditory sequences more closely than observed (94.6% as opposed to 62.1%). We conclude, therefore, that the strong cross-modal interference found in Experiment 1 does not arise from auditory distortion of the visual sequences.

EXPERIMENT 2

The robust cross-modal interference seen in Experiment 1, as well as the performance enhancement effected by congruent auditory information, reinforced our intuition that visual temporal structure is automatically (though perhaps imperfectly) encoded using an auditory representation. By this hypothesis, the nontemporal properties of the visual stimulus should affect rhythm processing only to the extent that they facilitate or disrupt the extraction of distinct changes for auditory encoding.

Experiment 2 tested this idea by pitting task-irrelevant auditory information against task-irrelevant visual information in their capacity to interfere with rhythm discrimination. To induce visual interference, we varied across sequences the nature of the visual changes giving rise to the temporal structure; this manipulation differentiated the sequences' visual representations without affecting any corresponding auditory codes. If visual representations underlie the processing of visual temporal structure, then stimulus variations along the task-irrelevant visual dimension would be expected to impair rhythm discrimination performance. However, if people automatically abstract an auditory code from visual changes, regardless of the nature of those changes, then task-irrelevant auditory information would be expected to interfere with rhythm discrimination to a greater extent than task-irrelevant visual information.

Method

Once again, observers made same/different judgments concerning the rhythm of two sequences that were depicted by a series of visual changes. However, the rhythms appeared with or without *task-irrelevant auditory information* and with or without *task-irrelevant visual information*.

In this experiment, task-irrelevant auditory information consisted of sequences of eight clicks that were always incongruent with the timing of the visual changes. When auditory rhythms were present, observers always heard different sequences during the two intervals; when auditory rhythms were absent, no sound accompanied the visual stimuli.

Task-irrelevant visual information arose through variations in the nature of the visual changes giving rise to the two successive visual rhythms. Rhythms could be portrayed by the timing of contrast reversals or by the timing of 90° orientation changes. On trials without task-irrelevant visual information, the same type of visual change depicted both visual rhythms: Trials consisted of two sequences of contrast reversals or two sequences of orientation reversals. On trials with task-irrelevant information, different types of visual change were used in the two sequences: That is, contrast reversals depicted the first rhythm and orientation reversals the second rhythm, or vice versa. As before, *same* trials contained two sequences of identically timed changes, whereas *different* trials contained two sequences of differently timed changes.

We instructed observers to ignore the nature of the visual changes, as well as auditory input, in making their same/different judgments about the visually defined rhythms.

The same 10 observers who participated in Experiment 1 completed four sessions of 80 randomly ordered trials (10 *same* and 10 *different* trials of each type). Experiment 2 matched Experiment 1 in all other aspects of the methodology.

Results and Discussion

Figure 3 summarizes observers' ability to discriminate two successive visual rhythms under the various auditory and visual conditions; as before, the data are collapsed over *same* and *different* trials. Task-irrelevant information in both the auditory and visual domains significantly reduced task performance, $F(1, 9) = 48.2$, $p < .001$, $\eta^2 = .72$, and $F(1, 9) = 19.2$, $p < .01$, $\eta^2 = .06$, respectively; the interaction between these factors did not approach significance, $F < 1$. However, incongruent auditory information on its own produced much greater interference with rhythm discrimination than did varying the visual stimuli along a task-irrelevant dimension, $t(9) = 4.4$, $p < .01$, $\eta^2 = .68$, even though the task is inherently and solely visual in nature.

These results imply that the human perceptual system may indeed encode visual rhythm sequences in an essentially auditory manner: Incongruent auditory information substantially impeded rhythm memory, even though this information was irrelevant to the visual task. Moreover, this auditory encoding of visual temporal structure appears to be obligatory. Despite realizing that auditory inputs could be confusing, observers were unable to ignore the sounds and rely exclusively on the visual sequences. When the same type of change portrayed both rhythms, these visual sequences contained potential information that could be divorced from the temporal structure per se (e.g., contrast summation over time). Nonetheless, adding task-irrelevant visual information impaired rhythm discrimination only slightly, suggesting primary reliance on temporal structure encoded in an auditory format.

EXPERIMENT 3

The discrimination of visual rhythms involves several component processes, including perceptual encoding of the first sequence's temporal structure, retention of this information in working memory, and comparison of this information with the second sequence's temporal structure (either during its presentation or following perceptual encoding). Though we assumed that cross-modal interference disrupts the encoding of a durable representation, this assumption had to be tested. To this end, we performed Experiment 3, in which task-irrelevant auditory information accompanied only the first visual sequence (i.e., encoding), only the second visual sequence (i.e., retrieval and comparison), both sequences, or neither sequence.

Method

Experiment 3 employed the same observers, stimuli, and procedure as Experiment 1, with the exception that the design involved the factorial combinations of two variables: *sound at encoding* (present or absent) and *sound at retrieval* (present or absent). All auditory sequences were incongruent with the concurrent visual sequences; when present during both intervals, the two auditory sequences differed from one another.

Results and Discussion

Figure 4 summarizes performance in the various conditions. When an incongruent auditory sequence accompanied encoding of the first visual rhythm, observers showed a large deficit in performance, $F(1, 9) = 16.6$, $p < .01$, $\eta^2 = .48$. Additionally, incongruent auditory information

during retrieval (i.e., accompanying the second visual rhythm) diminished task performance slightly, though not significantly, $F(1, 9) = 4.8$, $p = .06$, $\eta^2 = .06$. A direct pair-wise comparison indicated that sound at encoding affected rhythm matching significantly more than sound at retrieval, $t(9) = 2.5$, $p < .05$, $\eta^2 = .41$. The interaction between the two factors did not reach significance, $F(1, 9) = 1.8$, $p = .21$, $\eta^2 = .01$.

In sum, task-irrelevant auditory information primarily affects the perceptual system's ability to encode visual temporal structure. This finding supports the notion that temporal rhythm information—even when presented visually—is automatically and involuntarily registered and remembered using an auditory code. Additionally, auditory stimulation may interfere with the retrieval of previously stored visual rhythms or the comparison of two temporal structures; however, as the effect of auditory stimulation during the second sequence may simply reflect encoding of the second visual rhythm, this latter conjecture remains to be tested.

GENERAL DISCUSSION

The perception of a unitary environment depends critically on interactions among the senses. The experiments presented here suggest that obligatory cross-modal encoding may be one type of sensory interaction that, though often overlooked, plays a role in shaping people's perceptual reality.

Experiment 1 indicated that rhythmic auditory sequences disrupt processing of visual temporal structure. Experiment 2 further demonstrated that this auditory interference far outweighs the impact of varying the nature of the stimulus changes giving rise to visual temporal structure. Experiment 3 confirmed that cross-modal interference impairs encoding of the temporal structure, rather than (or in addition to) its retrieval. Together, these findings suggest that the human perceptual system abstracts temporal structure from the nature of its visual "messenger," automatically representing this structure using an essentially auditory code.

In some respects, the idea of visual-to-auditory cross-encoding seems familiar. Many people experience subvocal speech when reading, and accomplished musicians report hearing music when viewing a musical score (Brodsky, Henik, Rubinstein, & Zorman, 2003). However, these experiences, which may better be termed cross-modal *recoding*, differ significantly from the phenomenon reported here. Both subvocal reading and auditory imagery from musical notation develop only after considerable practice. By contrast, the cross-modal encoding of visual temporal structure demonstrated in the present experiments arose without explicit learning or practice. This makes sense because, unlike the auditory experiences that accompany the viewing of written words or a musical score, the auditory representation of temporal structure bears a natural, nonarbitrary relationship to the inducing visual stimulus.

Perhaps most important, whereas the auditory recoding of text or music likely reflects an effortful processing strategy (at least prior to extensive practice), the cross-modal encoding of visual temporal structure appears to be automatic and obligatory. The visual rhythms we used were presented too rapidly to realistically allow an effortful recoding strategy. Furthermore, such a strategy would prove suboptimal if observers could instead take advantage of visual representations, which would be unaffected by auditory stimulation. In a study supporting this strategic-versus-obligatory distinction, Brodsky et al. (2003) found that short-term memory for musical notation was disrupted by effortful coarticulation—indicating that auditory imagery plays a role in this task—but not by passive auditory input. As we have discussed, passive auditory information in our task had a profound impact on the encoding of visual temporal structure.

Obligatory cross-modal encoding also differs from other types of auditory-visual interactions discussed in the literature. Previously demonstrated effects include the disambiguation of visual motion displays with ecologically valid sound stimuli (Sekuler, Sekuler, & Lau, 1997) and the auditory induction of illusory flashing in an unambiguous disk of light (Shams, Kamitani, & Shimojo, 2000, 2002). The ventriloquism effect (e.g., Howard & Templeton, 1966) and auditory driving (e.g., Recanzone, 2003; Shipley, 1964) both reflect capture of one sensory modality by conflicting information in another, resulting in a unitary perceptual experience. In all of these cases, information in one sensory modality dramatically affects perception in another sensory modality. In the current study, auditory information did not alter perception of the relevant visual stimuli; the auditory and visual rhythms maintained their perceptual distinctiveness (see Experiment 1). Thus, the auditory-visual interactions did not result in the perceptual experience of unity.[3] Rather, the auditory information interfered with the encoding of a durable representation of visual temporal structure, suggesting that this nominally visual representation was essentially auditory in nature.

Nonetheless, cross-modal encoding may have important implications for the processing of a unitary reality. Previous research on spatial processing suggests the existence of a unified neural representation of visual and auditory space (Knudsen & Brainard, 1995). This finding, as well as the phenomenon reported here, suggests that the critical modules of perceptual processing involve dimensions of the environment such as space and time, rather than strict sensory segregation (see also Shimojo & Shams, 2001). A perceptual module for processing time, for example, would facilitate the construction of a unitary temporal framework from multiple sensory modalities. Dominance effects may arise from differential weightings of inputs to the module, as suggested by the modality-appropriateness hypothesis (Welch, 1999; Welch & Warren, 1980). Furthermore, given the effectiveness with which auditory information reflects time, the representations arising from such a module could well engender an auditory (rather than amodal) character, resulting in the experience of “hearing” visual temporal structure.

Does there exist physiological evidence for a multimodal temporal processing module? In a study of auditory and visual rhythm reproduction, Penhune, Zatorre, and Evans (1998) found that visual rhythms produced activation in regions of multimodal cortex (superior temporal sulcus and insula); interestingly, however, auditory rhythms did not give rise to similar activation. Nonetheless, multimodal cortex may contain neural machinery supporting generalized temporal processing. Alternatively, auditory cortex may itself contain such machinery. Recent studies demonstrate that primary auditory cortex can be activated by visual stimuli that merely imply sound (e.g., Calvert et al., 1997). Might visual rhythms also activate auditory cortex? Clearly, further research is needed to unravel the neural underpinnings of cross-modal temporal processing.

In conclusion, the present results imply that visual temporal structure is automatically and effortlessly transformed from its inherently visual form into an accurate auditory representation. This process of transformation could be construed as a form of synesthesia, wherein stimulation of one modality evokes sensory experiences in another (Robertson, 2003). Unlike in most forms of synesthesia, however, the relation between the auditory representation and visual temporal structure is not arbitrary but is, instead, isomorphic. People tend to hear rhythms in the mind's ear that are synchronized with rapidly occurring visual changes. Think about this tendency the next time you watch a conductor's arm movements coordinating a musical passage or see a naval ship flashing Morse code signals.

[3]Interestingly, several observers reported experiencing a complex rhythmic gestalt that combined the auditory and visual inputs. However, information from the two senses remained clearly distinguishable.

Acknowledgments

The authors thank the observers who participated in this study and Larry Marks, James Cutting, and an anonymous reviewer for helpful comments on an earlier draft of this article. This work was supported by National Eye Institute Grants EY07760 to R.B. and EY07135 to the Vanderbilt Vision Research Center.

REFERENCES

Bertelson P, Aschersleben G. Automatic visual bias of perceived auditory location. Psychonomic Bulletin & Review 1998;5:482–489.

Brodsky W, Henik A, Rubinstein B-S, Zorman M. Auditory imagery from musical notation in expert musicians. Perception & Psychophysics 2003;65:602–612. [PubMed: 12812282]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff OW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science 1997;276:593–596. [PubMed: 9110978]

Collier GL, Logan G. Modality differences in short-term memory for rhythms. Memory & Cognition 2001;28:529–538.

Fendrich R, Corballis PM. The temporal cross-capture of audition and vision. Perception & Psychophysics 2001;63:719–725. [PubMed: 11436740]

Glenberg AM, Jona M. Temporal coding in rhythm tasks revealed by modality effects. Memory & Cognition 1991;19:514–522.

Glenberg AM, Mann S, Altman L, Forman T, Procise S. Modality effects in the coding and reproduction of rhythms. Memory & Cognition 1989;17:373–383.

Hay JC, Pick HL, Ikeda K. Visual capture produced by prism spectacles. Psychonomic Science 1965;2:215–216.

Howard, IP.; Templeton, WB. Human spatial orientation. Wiley; New York: 1966.

Kitagawa N, Ichihara S. Hearing visual motion in depth. Nature 2002;416:172–174. [PubMed: 11894093]

Kitajima N, Yamashita Y. Dynamic capture of sound motion by light stimuli moving in three-dimensional space. Perceptual and Motor Skills 1999;89:1139–1158. [PubMed: 10710763]

Knudsen EI, Brainard MS. Creating a unified representation of visual and auditory space in the brain. Annual Reviews of Neuroscience 1995;18:19–43.

Marks LE, Ben-Artzi E, Lakatos S. Cross-modal interactions in auditory and visual discriminations. International Journal of Psychophysiology 2003;50:125–145. [PubMed: 14511841]

Penhune VB, Zatorre RJ, Evans AC. Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. Journal of Cognitive Neuroscience 1998;10:752–765.

Recanzone GH. Rapidly induced auditory plasticity: The ventriloquism aftereffect. Proceedings of the National Academy of Sciences, USA 1998;95:869–875.

Recanzone GH. Auditory influences on visual temporal rate perception. Journal of Neurophysiology 2003;89:1078–1093. [PubMed: 12574482]

Repp BH, Penel A. Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. Journal of Experimental Psychology: Human Perception and Performance 2002;28:1085–1099. [PubMed: 12421057]

Robertson LC. Binding, spatial attention and perceptual awareness. Nature Reviews Neuroscience 2003;4:93–102.

Sekuler R, Sekuler AB, Lau R. Sound alters visual motion perception. Nature 1997;385:308. [PubMed: 9002513]

Shams L, Kamitani Y, Shimojo S. What you see is what you hear. Nature 2000;408:788. [PubMed: 11130706]

Shams L, Kamitani Y, Shimojo S. Visual illusion induced by sound. Cognitive Brain Research 2002;14:147–152.

Shimojo S, Shams L. Sensory modalities are not separate modalities: Plasticity and interactions. Current Opinion in Neurobiology 2001;11:505–509. [PubMed: 11502399]

Shipley T. Auditory flutter-driving of visual flicker. Science 1964;145:1328–1330. [PubMed: 14173429]

Wada Y, Kitagawa N, Noguchi K. Audio–visual integration in temporal perception. International Journal of Psychophysiology 2003;50:117–124. [PubMed: 14511840]

Welch, RB. Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In: Aschersleben, G.; Bachmann, T.; Musseler, J., editors. Cognitive contributions to the perception of spatial and temporal events. Elsevier; Amsterdam: 1999. p. 371.-387.

Welch RB, DuttonHurt LD, Warren DH. Contributions of audition and vision to temporal rate perception. Perception & Psychophysics 1986;39:294–300. [PubMed: 3737359]

Welch RB, Warren DH. Immediate perceptual response to intersensory discrepancy. Psychological Bulletin 1980;68:638–667.

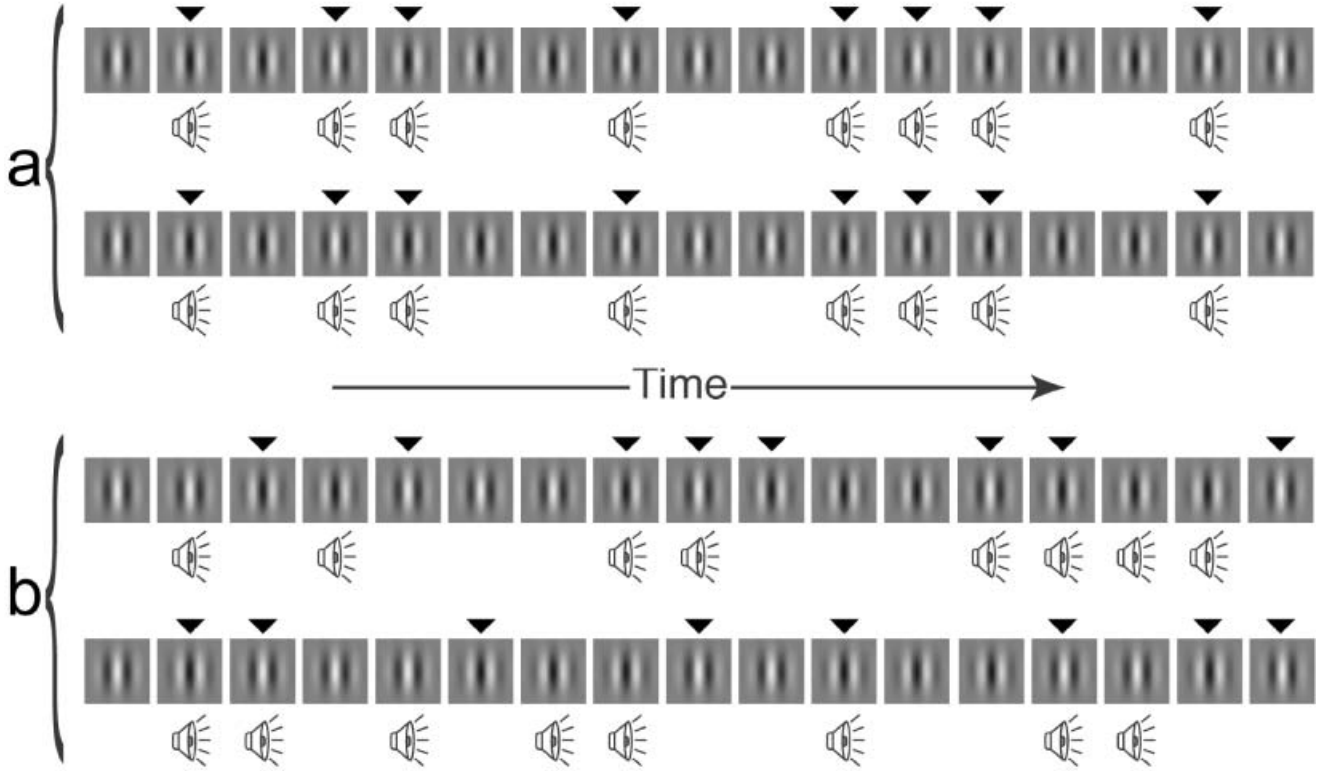

Fig 1.

Schematic diagram of the visual and auditory stimuli used in Experiment 1: (a) two visual sequences constituting a *same* trial accompanied by congruent auditory sequences and (b) two visual sequences constituting a *different* trial accompanied by two different incongruent auditory sequences. The arrows indicate the timing of the visual beats (points in time when the Gabor patches underwent contrast reversals), and the speakers indicate the timing of the auditory clicks. Note that on both *same* and *different* trials, the auditory signals could be either congruent with the visual beats, incongruent with the visual beats, or absent altogether; the two factors were orthogonal, meaning that auditory sequences alone provided no cue for reliable discrimination of same versus different visual sequences.

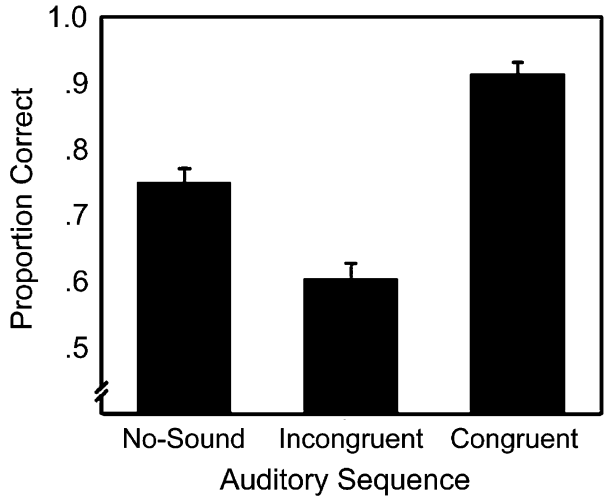

Fig 2.
Results of Experiment 1: proportion of correct responses as a function of the nature of the auditory sequences. Error bars depict standard errors across observers.

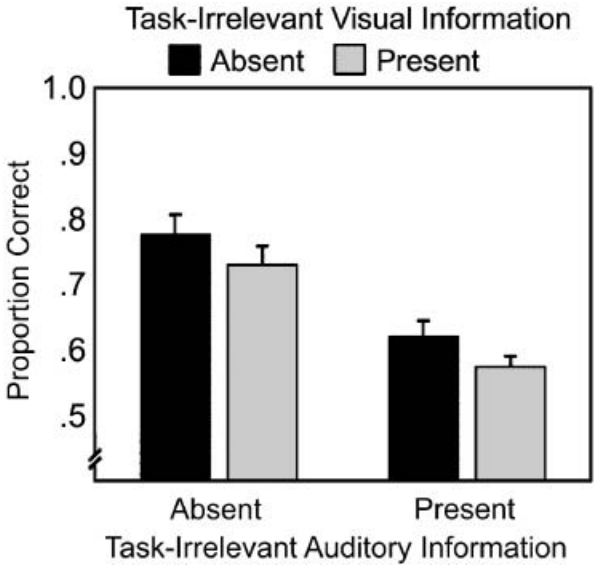

Fig 3.

Results of Experiment 2: proportion of correct responses as a function of the presence or absence of task-irrelevant information in the auditory and visual domains. Error bars depict standard errors across observers.

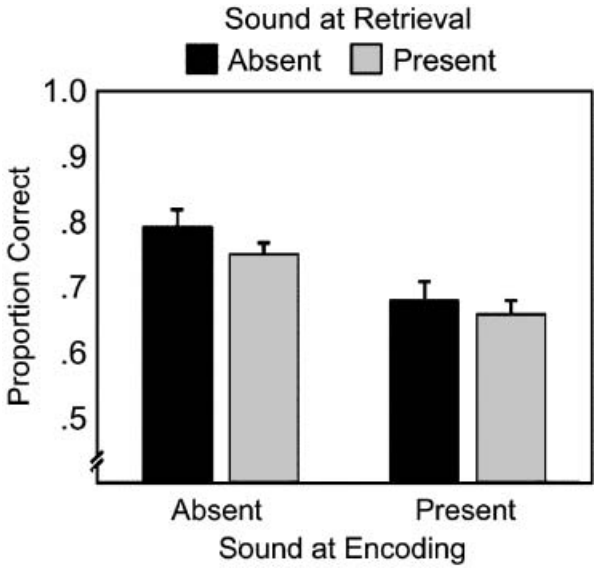

Fig 4.

Results of Experiment 3: proportion of correct responses as a function of the presence or absence of auditory stimulation during the first (encoding) and second (retrieval) visual sequences. Error bars depict standard errors across observers.