

# Auditory Dominance in Temporal Processing: New Evidence From Synchronization With Simultaneous Visual and Auditory Sequences

Bruno H. Repp and Amandine Penel  
Haskins Laboratories

Evidence that audition dominates vision in temporal processing has come from perceptual judgment tasks. This study shows that this auditory dominance extends to the largely subconscious processes involved in sensorimotor coordination. Participants tapped their finger in synchrony with auditory and visual sequences containing an event onset shift (EOS), expected to elicit an involuntary phase correction response (PCR), and also tried to detect the EOS. Sequences were presented in unimodal and bimodal conditions, including one in which auditory and visual EOSs of opposite sign coincided. Unimodal results showed greater variability of taps, smaller PCRs, and poorer EOS detection in vision than in audition. In bimodal conditions, variability of taps was similar to that for unimodal auditory sequences, and PCRs depended more on auditory than on visual information, even though attention was always focused on the visual sequences.

Time differs from most other dimensions of the environment in that people do not have a specific sensory organ for its perception. Some theorists go as far as claiming that events are perceivable but time is not (Gibson, 1975). Yet all events, regardless of their sensory modality, contain temporal information that is registered by the brain: Duration and temporal structure can be appreciated in music, in a silent movie, and in a body massage. Time shares this supramodal nature with space: The size, location, and distance of an object, too, can be heard, seen, or felt.

Despite the supramodality of space and time, it has been claimed that vision and audition are relatively specialized for spatial and temporal processing, respectively (Freides, 1974; Geldard, 1970; Kubovy, 1988; Näätänen & Winkler, 1999; O'Connor & Hermetin, 1972). Empirical evidence supporting this claim comes mainly from two sources: comparisons of the relative sensitivity of each modality to spatial and temporal information, and studies showing dominance of one modality over the other when conflicting spatial or temporal information is presented, as reviewed below. Because our study is concerned with temporal processing, we do not review the evidence for better spatial discrimination in the visual than in the auditory modality. However, we consider briefly the research on visual dominance in the perception of conflicting spatial information, because it illustrates paradigms and hypotheses relevant to all kinds of cross-modal interaction.

## Visual Dominance for Spatial Information

The best known example of an audiovisual interaction showing visual dominance (in most situations) is the ventriloquist effect, first reported by Stratton (1897). Discrepant visual and auditory information about the location of an object can be created by either displacing the visual image with prisms or manipulating binaural cues for auditory localization. Numerous studies have shown that a sound associated with an object appears to come from a direction that is much closer to the object's seen location than to the actual location of the sound source (Bermant & Welch, 1976; Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981; Bertelson, Vroomen, de Gelder, & Driver, 2000; Jack & Thurlow, 1973; Jackson, 1953; Pick, Warren, & Hay, 1969; Radeau & Bertelson, 1976, 1977, 1987; Thomas, 1941; Vroomen, Bertelson, & de Gelder, 2001; Warren, 1979, 1980; Warren, Welch, & McCarthy, 1981; Welch, 1999; Witkin, Wapner, & Leventhal, 1952). Thus, auditory source localization is strongly biased by pertinent visual information, whereas pertinent auditory information seems to exert little or no influence on visual object localization (but see Bertelson & Radeau, 1981; Radeau & Bertelson, 1987; Warren et al., 1981).

The magnitude of the visual bias on auditory judgment depends on both perceptual (structural, noncognitive) and cognitive factors (Radeau & Bertelson, 1977; Welch, 1999; Welch & Warren, 1980). As to the former, visual dominance has been found to decrease when the spatial distance between the visual object and the sound source is increased (Bermant & Welch, 1976; Jack & Thurlow, 1973; Jackson, 1953; Witkin et al., 1952) and to increase with the visual source's relative salience (Bertelson et al., 2000; Radeau, 1985; Radeau & Bertelson, 1987). Relevant cognitive factors include the participants' awareness of a sensory discrepancy (Warren, 1979) and their belief that the visual and auditory information come from the same source, the so-called *unity assumption* (Bermant & Welch, 1976; Bertelson & Aschersleben, 1998; Jack & Thurlow, 1973; Jackson, 1953; Radeau & Bertelson, 1987; Thomas, 1941; Warren et al., 1981). These cognitive factors

---

Bruno H. Repp and Amandine Penel, Haskins Laboratories, New Haven, Connecticut.

This research was supported by National Institutes of Health Grant MH-51230 to Bruno H. Repp and by a postdoctoral fellowship from the Fyssen Foundation to Amandine Penel. We are grateful to Yoko Hoshi for assistance, to Jack Vees for loaning us essential equipment, and to Mari Riess Jones, Ani Patel, Mira Peter, Jeff Pressing, and Paul Treffner for helpful comments on the manuscript.

Correspondence concerning this article should be addressed to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695. E-mail: repp@haskins.yale.edu

depend in part on stimulus structure and in part on instructions or prior experience (see Welch, 1999; Welch & Warren, 1980). Attention to one or to the other modality has also been found to be important (Canon, 1970, 1971). However, two recent studies have shown that the visual bias of auditory spatial localization is independent of where in space visual attention is directed (Bertelson et al., 2000; Vroomen et al., 2001). Finally, response factors, such as pointing in a direction versus making a judgment on a numeric scale, may also play a role (Warren, 1980).

To explain dominance of one modality over another in such conflict situations, researchers have proposed three hypotheses (see Welch & Warren, 1980). According to the *modality precision hypothesis*, the modality that exhibits more accurate discrimination for the kind of information presented is favored. This hypothesis is consistent with dominance of vision over audition in spatial localization, but some conflicting evidence has also been noted (Radeau & Bertelson, 1976; Warren, 1979). The *directed attention hypothesis* has been advanced by Posner, Nissen, and Klein (1976). They argued that visual stimuli do not have the strong alerting capacity of auditory stimuli, and therefore observers are predisposed to direct their attention toward visual stimulation, which in turn causes visual dominance. This hypothesis is consistent with Canon's (1970, 1971) finding that the ventriloquist effect is reduced when participants' attention is specifically directed to the sound. Finally, a *modality appropriateness hypothesis* has been proposed, according to which "the human perceptual system is cognizant of the fact that vision is a more trustworthy modality for spatial localization than is audition and proprioception and, for this reason, it is more closely attended" (Welch, 1999, p. 382). It seems to us that this hypothesis is an amalgam of the modality precision and directed attention hypotheses.

In summary, vision may dominate audition in spatial judgments because it offers more accurate sensory information, because it receives more attention, or both. There are many other situations in which multimodal information is perceptually integrated, such as audiovisual speech perception (Massaro, 1987; McGurk & Macdonald, 1976) and emotion judgment (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). However, the relevant information in these cases is usually dynamic (i.e., it involves change over time) in at least one modality, and visual dominance is not as common a finding as it is in the case of conflicting information of a purely spatial nature. Next we present evidence suggesting that audition is superior to vision in processing purely temporal information.

### Relative Sensitivity of Vision and Audition to Temporal Information

A number of studies have shown that relatively short intervals (less than 2 s) are discriminated and reproduced with greater accuracy when the stimuli are auditory than when they are visual and that this holds for both filled and empty intervals (Goldstone & Lhamon, 1971; Goodfellow, 1934; Grondin, 1993; Grondin, Meilleur-Wells, Ouellette, & Macar, 1998; Grondin, Ouellet, & Roussel, 2001; Grondin & Rousseau, 1991; Lhamon & Goldstone, 1974; Rousseau, Poirier, & Lemyre, 1983; Tanner, Patton, & Atkinson, 1965). In a similar manner, discrimination and reproduction of rhythmic patterns is superior in the auditory modality (Gault & Goodfellow, 1938). We are not aware of any studies that

have compared observers' perceptual sensitivity to small deviations from temporal regularity in extended stimulus sequences, as we have in the present study. However, it is known that people have greater difficulty synchronizing finger taps with visual than with auditory sequences, as reflected in greater variability or phase drift (Bartlett & Bartlett, 1959; Dunlap, 1910; Fraisse, 1948; Klemmer, 1967; Kolers & Brewster, 1985). Such results have led Kolers and Brewster (1985) to suggest that there are different timing mechanisms in the two modalities, and Fraisse (1948) has argued that the motor system is generally more responsive to auditory than to visual input.

The commonly found tendency for taps to precede sequence events in sensorimotor synchronization (the negative asynchrony or anticipation effect) tends to be smaller with visual than with auditory stimuli, at least when the taps are silent (Fraisse, 1948; Kolers & Brewster, 1985). This does not mean that synchronization with visual sequences is more accurate. Rather, it may be related to the fact that simple reaction times are shorter to auditory than to visual stimuli, which suggests faster neural processing of auditory stimuli (e.g., Jaśkowski, Jaroszyk, & Hojan-Jezińska, 1990). If synchrony is established at a central level on the basis of accumulated sensory information from taps and sequence events (Aschersleben, Stenneken, Cole, & Prinz, 2002; Prinz, 1997), then slower processing of taps than sequence events will lead to a negative asynchrony, and a relatively slower sensory processing time for visual than for auditory sequence events will result in a smaller negative asynchrony of taps with the former than with the latter. Paradoxically, however, an auditory stimulus must precede a visual one to be judged as simultaneous—a puzzle that has not yet been solved (Jaśkowski, 1999).

### Auditory Dominance for Temporal Information

In contrast with the large amount of research on the perception of conflicting bimodal spatial localization cues, the processing of conflicting bimodal temporal information has been much less investigated, and the results are also less consistent. Evidence for dominance of audition over vision has been found in the perception of duration, in studies that took advantage of the finding that a tone (usually lasting 1 s or more) tends to be perceived as longer than a light of equal duration (Behar & Bevan, 1961; Goldstone, Boardman, & Lhamon, 1959; Goldstone & Lhamon, 1974; Penney, Gibbon, & Meck, 2000; Walker & Scott, 1981; but see Tanner et al., 1965). When a tone and a light of equal physical duration (but presumably still unequal perceived duration) were presented simultaneously, observers' judged or reproduced duration of the light or of the composite stimulus was similar to the judged or reproduced duration of the tone alone (Goldstone et al., 1959; Walker & Scott, 1981). However, this auditory dominance was fragile, because it changed to visual dominance when the intensity of the light was raised and that of the tone was lowered (Goldstone et al., 1959). More intense tones and lights are judged as longer than less intense stimuli (Goldstone, Lhamon, & Sechzer, 1978), so it could be that the stimulus with the longer subjective duration is dominant. Walker and Scott (1981) found some support for this suggestion.

A convincing demonstration of auditory dominance in temporal judgment was provided recently by Fendrich and Corballis (2001), who asked participants to judge when a flash or click occurred by

reporting the clock position of a rotating marker. When a flash was preceded or followed by an irrelevant click, temporal judgments about the flash were strongly biased in the direction of the click. A weaker biasing effect occurred when an irrelevant flash preceded or followed a click whose temporal position was to be judged.

Auditory dominance has also been found in the perception of sequence rate at relatively fast rates (greater than 4 Hz). A change in the rate of an auditory sequence causes the perceived rate of a simultaneously presented but constant visual sequence to change as well, a phenomenon called *auditory driving* (Gebhard & Mowbray, 1959; Myers, Cotton, & Hilp, 1981; Shipley, 1964; Welch, 1999). In contrast, varying the rate of a visual sequence does not seem to change the perceived rate of a simultaneously presented constant auditory sequence.

### Purpose of the Present Study

Our aim in this study was to determine whether auditory temporal information outweighs visual temporal information not only in perception but also in the temporal control of coordinated action. The tasks that have hitherto been used to study auditory dominance in temporal processing have usually required explicit perceptual judgments or, in the case of reproduction, memory. Perceptual judgments require conscious awareness of temporal properties, differences, or relationships. Memory may rely on what has been perceived consciously. In contrast, sensorimotor synchronization relies largely on automatic, subconscious processes that temporally coordinate action with external referents (Pressing, 1999). Therefore, we investigated whether audition is dominant over vision in a sensorimotor synchronization task when temporal information is presented simultaneously in the two modalities. That is, we examined the process of *phase correction* that underlies people's ability to stay in synchrony with an isochronous sequence and that enables them to adjust to deviations from temporal regularity in such a sequence (Vorberg & Wing, 1996). Phase correction has been found to occur regardless of participants' awareness of such deviations (Repp, 2000, 2001a). Moreover, phase correction occurs involuntarily when participants try not to react to deviations, although the response tends to be reduced in magnitude (Repp, 2002a, 2002c).

We focused on this involuntary component of phase correction by using the *event onset shift* (EOS) paradigm of Repp (2002a). When the onset of a single event in an isochronous auditory sequence is shifted, the best strategy for participants is to tap regularly and not to react to the perturbation, so that the taps following the EOS are in synchrony with the following sequence events. Nevertheless, following an early event onset (a negative shift), the next tap tends to be advanced; following a late event onset (a positive shift), the next tap tends to be delayed. This immediate, obligatory response has been termed the *phase correction response* (PCR) by Repp (2002a). The average PCR tends to be larger for positive than for negative EOSs, and it increases linearly with EOS magnitude up to a point (between 10% and 20% of the sequence interonset interval) where it reaches an asymptote (see also Repp, 2002c).

Sensorimotor synchronization has been investigated mainly in the auditory modality (e.g., Aschersleben & Prinz, 1995, 1997; Hary & Moore, 1985, 1987; Mates, 1994a, 1994b; Michon, 1967; Pressing, 1998; Repp, 2000, 2001a, 2002a; Semjen, Schulze, &

Vorberg, 2000; Semjen, Vorberg, & Schulze, 1998; Thaut, Miller, & Schauer, 1998; Thaut, Tian, & Azimi-Sadjadi, 1998). On the basis of the few studies that compared synchronization of finger taps with auditory and visual sequences (Bartlett & Bartlett, 1959; Dunlap, 1910; Fraisse, 1948; Klemmer, 1967; Kolers & Brewster, 1985), we expected to find more variable and less negative asynchronies in a visual than in an auditory condition. However, we did not know whether PCRs would be elicited by EOSs in visual sequences. Therefore, we began by asking participants to synchronize their taps with unimodal auditory and visual sequences containing EOSs. In addition, we required participants to report after each sequence whether they had noticed an EOS, and if so, whether the shifted event had been early or late. On the basis of earlier findings (Gault & Goodfellow, 1938; Goodfellow, 1934; Klemmer, 1967; Lhamon & Goldstone, 1974; Tanner et al., 1965), we expected the detection threshold for EOSs to be higher in vision than in audition. This gave us an opportunity to investigate whether the average PCR (if any) in vision is independent of the detection threshold, as it seems to be in audition (Repp, 2002a). If so, then the average PCR should increase linearly with visual EOS magnitude from zero to some point at which an asymptote is reached, without any discontinuity at the detection threshold. We were further interested as to whether the slope of this linear increase would be the same as in the auditory modality or whether it would be more shallow, indicating less efficient phase correction.

Following these unimodal conditions, we investigated whether auditory dominance occurs when auditory and visual sequences containing EOSs are presented simultaneously and synchronously (except at the time of an EOS). We instructed participants to focus their attention on the visual sequences, to synchronize with them, and to report any EOS detected in them; the auditory sequences were to be ignored. We hypothesized that even with instructions favoring vision, auditory dominance would occur. Auditory dominance might already be evident in the magnitude and variability of asynchronies, regardless (or in the absence) of an EOS, if the asynchronies are closer to those observed in the unimodal auditory than in the unimodal visual condition. Our main criterion, however, was the response to the EOSs in the bimodal sequences.

There were two bimodal conditions with regard to EOSs: nonconflicting and conflicting. In the nonconflicting condition, an EOS occurred in one or the other modality but never in both. Here, auditory dominance would be evident in a substantial reduction of the PCRs to visual EOSs and in the occurrence of strong PCRs to auditory EOSs, despite attention to the visual modality. In the conflicting condition, EOSs of opposite direction occurred simultaneously in both modalities. Here, auditory dominance would be evident if the PCRs went in the direction of the auditory rather than of the visual EOSs, and more so than would be predicted from merely averaging the PCRs to EOSs in unimodal visual and auditory sequences (if the former are smaller than the latter). That is, our criterion for auditory dominance was that the response measure must show greater dependence on auditory than on visual information, despite the required attentional bias toward vision.

In addition to the PCRs, the detection responses to the visual EOSs in the bimodal conditions were of interest. These responses were expected to reveal whether the auditory sequences affected the perceived timing of the visual sequences, as in the Fendrich and Corballis (2001) study mentioned earlier. If shifted auditory

events induce illusory temporal shifts of coinciding visual events, this should lead to false-alarm responses in the bimodal nonconflicting condition and to difficulties in detecting and identifying the direction of the visual EOSs in the bimodal conflicting condition. If such cross-modal perceptual interactions occur, they may be directly responsible for auditory dominance observed in other response measures (asynchronies or PCRs), or they may be an unrelated phenomenon. That is, the convergence of sensory information leading to a perceptual interaction may occur before or after timing information is directed to the centers of action control. Although our study was not designed to distinguish explicitly between these two possibilities, we expected to be able to draw some conclusions from the pattern of results obtained.

## General Method

### Materials and Equipment

All sequences were produced from preassembled musical instrument digital interface (MIDI) instructions under the control of a Macintosh Quadra 660AV computer, using a program written in MAX. Because of a peculiarity of this software, all real-time temporal intervals produced or recorded were in fact 2.4% shorter than reported in this article.

The experiment was run in two versions that differed in the nature of the visual stimuli. In Version 1, the visual sequences consisted of a black X (two crossed, diagonal lines) switching back and forth between two black square frames ( $5 \times 5$  mm) separated horizontally by 2 mm. Participants viewed the computer monitor from a distance of approximately 60 cm. The X was initially in the left square; its movement to the right square (i.e., its simultaneous disappearance from the left square and appearance in the right square) constituted the first event of the sequence. The X remained in the right square until it moved back to the left square, and so on. After testing a number of participants, we realized that the 60-Hz refresh cycle of the computer screen must have introduced variable delays of up to 17 ms in the visual sequences. Because we were worried that this variability may have affected performance in the synchronization and detection tasks, we tested additional participants with Version 2, using visual sequences consisting of a flashing light. The light, a circular green light-emitting diode 3 mm in diameter, was the *out* indicator of an Opcode II MIDI Translator box that responded to MIDI messages. Because the flashes were difficult to see in bright light, all visual conditions in Version 2 were conducted in semidarkness. The lights of the windowless room were switched off, and the computer screen was dimmed and covered with a piece of cardboard, leaving a gap at the bottom that provided dim illumination for the computer keyboard, on which the detection responses were made. The box with the flashing light was placed on the rear edge of the keyboard. Under these conditions, the flashes were clearly visible and separated from each other. Instructions consisting exclusively of *note on* MIDI messages were used to activate the light.

The auditory sequences consisted of identical high-pitched (C8; 4186 Hz) synthetic piano tones (pings) of equal intensity, with no nominal duration (i.e., note offset) specified in the MIDI instructions. The tones had sharp attacks, reflecting also key impact noise, and decayed within about 100 ms. They were produced on a Roland RD-250s digital piano through a MIDI interface and were heard over Sennheiser HD540 II earphones.

Visual and auditory sequences had identical temporal specifications. The baseline interonset interval (IOI) was 500 ms. As illustrated schematically in Figure 1A, each sequence contained one time-shifted event (early or late). This EOS occurred in the 8th–12th position and was followed by five unperturbed events, so that a sequence contained between 13 and 17 events. The magnitude of the EOS ranged from  $-80$  to  $+80$  ms (including zero) in steps of 10 ms. The combination of 17 EOS magnitudes and 5 EOS positions resulted in 85 different sequences.

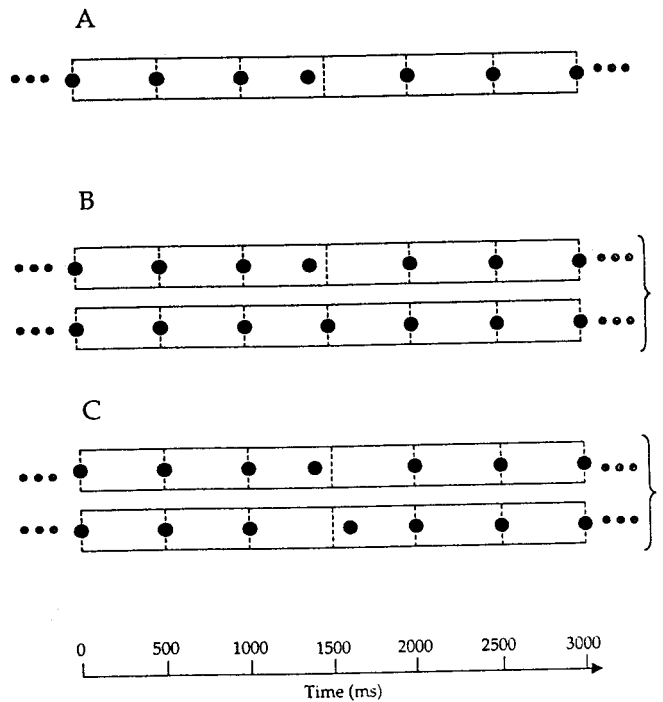


Figure 1. Schematic illustration of stimuli. A: Unimodal conditions. Each auditory or visual sequence contained one event onset shift (EOS). The event occurred either early (as shown) or late. B: Bimodal nonconflicting condition. An EOS in one modality was accompanied by an unperturbed event in the other modality. C: Bimodal conflicting condition. An EOS in one modality was accompanied by an EOS of the same magnitude but of opposite direction in the other modality.

In the unimodal conditions, visual and auditory sequences were presented in separate blocks. There were two blocks in each condition, each representing a different random order of the 85 sequences and each having a break in the middle. In the bimodal conditions, visual and auditory sequences occurred simultaneously. Simultaneous sequences always had the same length and were in synchrony, except where an EOS occurred. In the nonconflicting condition (Figure 1B), one sequence contained an EOS (including zero), whereas the other was isochronous. This resulted in 170 ( $85 \times 2$ ) different bimodal sequences, which were randomly intermixed. Two randomizations were made, each subdivided into four blocks with breaks in between. In the conflicting condition (Figure 1C), an EOS (including zero) in one modality always occurred simultaneously with an EOS of the same absolute size but of opposite direction in the other modality. This resulted in 85 different bimodal sequences. Again, two randomizations were made, each with a break in the middle.

### Participants

Sixteen participants were tested with Version 1 including both authors and a research assistant who had previous experience in synchronization tasks. All others were paid volunteers from the Yale community who had responded to a campus advertisement and had no previous task experience. Synchronization with the visual sequences proved to be a difficult task for most of them. The data of 7 participants were excluded, because 6 showed large variability and phase drift in the visual condition—it would have been trivial to demonstrate auditory dominance for participants who were unable to synchronize with visual sequences—and one had similar difficulties in the auditory condition. The remaining 9 participants showed reasonable

accuracy in synchronizing with both visual and auditory sequences. Five individuals were tested with Version 2. So that we could avoid further rejection of data, these participants included three members of a regular group of participants in auditory synchronization experiments, a new research assistant, and one of us (B.R.). B.R.'s results from Version 1 were not included in the group means but are reported in tables for comparison with his Version 2 data, the only possible within-participant comparison. The group means thus were based on 13 participants, 8 from Version 1 and 5 from Version 2.

### Procedure

Each participant came for three individual sessions on different days, corresponding to the three conditions of the experiment (unimodal, bimodal nonconflicting, and bimodal conflicting). The unimodal session began with a practice block of 25 auditory sequences, 10 of which contained an early event, 10 of which contained a late event, and 5 of which were completely isochronous. The EOSs ( $\pm 100$  ms) in these sequences were larger than the EOSs in the experimental trials. The practice sequences were followed by one test block of 85 auditory sequences. Participants then received a practice block of 25 visual sequences, analogous to the auditory practice block, followed by one test block of 85 visual sequences. After a break, the second auditory and visual test blocks were presented.

Participants were asked to synchronize their finger taps with the sequence events, starting with the second event, and to report after each sequence whether they had noticed a time-shifted event, and if so, whether it occurred early or late. For synchronization, they used the index finger of their preferred hand on a white key of a Fatar Studio 37 MIDI controller (a quiet three-octave piano keyboard) that they held on their lap. Participants were instructed to keep tapping as regularly as possible and not to let any time-shifted events affect their tapping. It was explained that one tap would be out of synchrony with the time-shifted event, but that synchrony with subsequent events would be assured if tapping continued at a regular rate. Participants were informed that the shifted events occurred in unpredictable positions, but never very close to the beginning or to the end of the sequence. After a sequence had ended, a detection response was made by pressing one of three keys on the computer keyboard: the left arrow key ( $\leftarrow$ ) for an early event, the right arrow key ( $\rightarrow$ ) for a late event, and the down arrow key ( $\downarrow$ ) if no temporal irregularity had been detected. The keyboard response started the next sequence after a 3-s delay. Participants were asked to avoid blinking their eyes while synchronizing with visual sequences (maximum duration = 8 s).

The key depressions on the MIDI controller and on the computer keyboard were recorded by the MAX program that also controlled presentation of the sequences. The response key on the MIDI keyboard moved about 10 mm, and the key depression was recorded about halfway during the downward movement. This may have added up to  $-20$  ms to the measured asynchronies, depending on the force with which participants struck the response key (the greater the force, the smaller the artifact). For most participants, there was no auditory feedback from the keyboard, but a few struck the key forcefully enough to make the impact noise audible. This may have reduced their negative asynchronies (see Aschersleben & Prinz, 1995). In any case, the exact magnitude of the average asynchronies and related individual differences were not of particular interest here, and there was no reason to believe that participants changed the force of their keystrokes systematically within or across conditions.

The bimodal nonconflicting and bimodal conflicting conditions did not include practice sequences. The participants' task was the same as for unimodal visual sequences: Participants were instructed to focus their attention on the visual sequences and to synchronize with them, not to react to any irregularities in them, and to report whether an early or late event had occurred in them. The auditory sequences were to be ignored. Participants were told that the auditory sequences could also contain temporal irregularities but that they were not to be reported.

## Results and Discussion

### Detection Responses

The mean percentages of detection responses (*early, no change, or late*) are shown in Figure 2 as a function of EOS magnitude. Figures 2A and 2B show the results for the unimodal visual and unimodal auditory conditions, respectively. It is clear that deviations from temporal regularity were much more difficult to detect in the visual than in the auditory modality. If the 50% crossover points of the *no change* response functions are taken as estimates of the average detection thresholds, then the thresholds were at roughly  $\pm 30$  ms (6% of the baseline IOI) in the auditory modality and at  $\pm 60$  ms (12%) in the visual modality. If the threshold estimates are based on the percentages of correct identification of EOS direction, the modality difference would be even larger. A repeated measures analysis of variance (ANOVA) on the mean percentages of *no change* responses (averaged across EOS magnitudes excluding zero) with the variables of modality (visual or auditory) and EOS direction (negative or positive) showed the main effect of modality to be highly significant,  $F(1, 12) = 50.9$ ,  $p < .0001$ . There was also a significant main effect of EOS direction,  $F(1, 12) = 7.1$ ,  $p < .03$ , indicating that positive EOSs (late events) were somewhat easier to detect than were negative EOSs (early events) in both modalities. (The interaction was nonsignificant.) The direction of detected EOSs was also identified correctly more often in the auditory than in the visual modality. This difference was likewise significant in an ANOVA on the relevant conditional percentages,  $F(1, 12) = 59.4$ ,  $p < .0001$ . However, there was no significant asymmetry between positive and negative EOSs here,  $F(1, 12) = 0.6$ .

The poor detection of visual EOSs clearly was not an artifact of the slightly variable sequence timing caused by the computer screen refresh cycle in Version 1 of the experiment. Although this variability may have increased the difficulty of the visual detection task somewhat (B.R. actually did better in Version 2 than in Version 1, although this could also have been due to practice), 3 of the 5 participants in Version 2 were almost totally at chance in the visual condition, and all participants in both versions performed much more poorly in the visual than in the auditory condition.

Figure 2C presents the detection responses to visual EOSs in the bimodal nonconflicting condition. The results were roughly similar to those in the unimodal visual condition (Figure 2A), which indicates that participants did follow the instructions and attended to the visual sequences in the bimodal condition. However, there were also some differences. *No change* responses were more frequent in the bimodal than in the unimodal visual condition,  $F(1, 12) = 9.6$ ,  $p < .01$ , which suggests that the presence of an unperturbed auditory sequence made the visual EOSs more difficult to detect. At the same time, however, misidentification of the direction of detected EOSs was less frequent in the bimodal than in the unimodal visual condition,  $F(1, 12) = 5.0$ ,  $p < .05$ . This may have been due to the unperturbed auditory event serving as a point of reference for judging the direction of the visual EOS (i.e., a temporal order judgment).

The results for auditory EOSs in the bimodal nonconflicting condition showed unexpectedly large individual differences and therefore are presented separately for two subgroups of participants (Figures 2D and 2F). If participants had been successful in ignoring the auditory sequences, as they had been instructed to do,

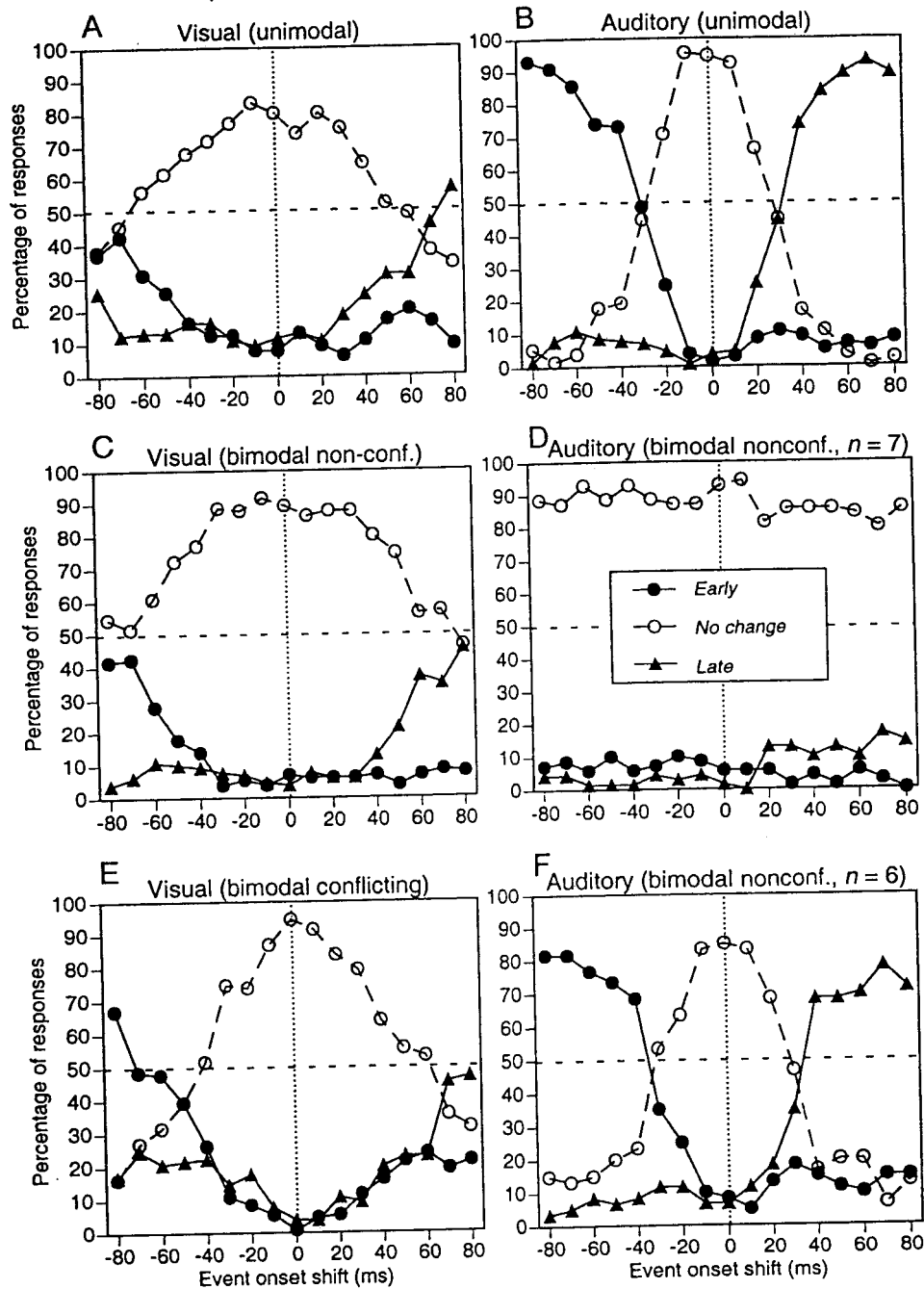


Figure 2. Mean percentages of detection responses (*early*, *no change*, and *late*) as a function of event onset shift magnitude in the five conditions, as indicated above each graph. Graphs D and F show data for different subgroups of participants in the same condition. nonconf. = nonconflicting.

they should have given only occasional detection responses when auditory EOSs occurred. Indeed, 7 participants (2 in Version 1 and all 5 in Version 2 of the experiment) showed this pattern of results (Figure 2D). The remaining 6 participants, however, responded strongly in the presence of auditory EOSs (Figure 2F). Their average results resembled those for the unimodal auditory condition (Figure 2B) but showed greater uncertainty.

Why did the participants in the second group respond when auditory EOSs occurred, contrary to the instructions? There are two possible explanations. One is that these participants misunderstood or ignored the instructions and responded directly to auditory as well as to visual EOSs. However, one of us (A.P.) was in this group, which rules out misunderstandings in at least one case and makes it seem less likely for the other 5 participants. The

other, more interesting possibility is that auditory EOSs caused illusory EOSs in the unperturbed visual sequences and that participants responded to these illusory visual EOSs. Such a perceptual interaction would be consistent with the findings of Fendrich and Corballis (2001) and of other studies reviewed above. However, it is curious that participants were either highly susceptible or quite immune to such a cross-modal interaction.

It might be argued that the unaffected participants discovered (or knew, in the case of B.R.) that visual and auditory EOSs were mutually exclusive (a fact not mentioned in the instructions) and avoided responding to illusory visual EOSs by responding *no change* whenever they detected an auditory EOS. However, such a strategy would have been contrary to instructions, and B.R. is quite sure that he did not see any visual EOSs to which he did not respond. Moreover, how would participants have discovered that visual and auditory EOSs were mutually exclusive if there were illusory visual EOSs accompanying the auditory EOSs? It seems that these participants really did not see visual EOSs accompanying auditory EOSs. These participants tended to be more experienced in auditory synchronization tasks than the others, but this does not explain why they were more successful in ignoring the auditory sequences or evading a cross-modal temporal bias. Therefore, these individual differences remain unexplained for the time being.

The detection responses for the bimodal conflicting condition are shown in Figure 2E. The abscissa refers to the visual modality, on which attention was focused. Compared with the unimodal visual condition, there was no overall difference in either the percentage of *no change* responses,  $F(1, 12) = 0.8$ , or the conditional percentage of correct identifications of visual EOS direction,  $F(1, 12) = 2.7$ ,  $p < .13$ . However, there was a significant Modality  $\times$  Direction interaction for *no change* responses,  $F(1, 12) = 7.2$ ,  $p < .03$ , due to a reduction of these responses to negative EOSs in the bimodal condition. Thus, the conflicting auditory EOSs helped detection of negative visual EOSs and did not hinder detection of positive visual EOSs. Likewise, the conflicting auditory EOSs did not interfere significantly with the identification of visual EOS direction. When the responses were examined separately for the two groups of participants whose bimodal nonconflicting data are shown in Figures 2D and 2F, respectively, the percentages of misidentifications of visual EOS direction were found to be more frequent in the second group, especially on the positive side (i.e., when the auditory event preceded the visual one). Thus, their responses were again suggestive of a perceptual or postperceptual interaction between the auditory and visual inputs. We should note that no participant capitalized on the knowledge or discovery of the perfect negative correlation between visual and auditory EOSs by monitoring the auditory sequences and simply reversing the direction of the responses. This would have been contrary to instructions and would have led to correct response functions resembling those in the unimodal auditory condition, which was never observed.

In summary, these data show that visual EOSs are much harder to detect than are auditory EOSs. Nevertheless, participants are able to detect visual EOSs in the presence of isochronous sequences or conflicting EOSs in the auditory modality. Some participants' responses were barely affected by the auditory information, whereas the responses of others revealed a strong influence of auditory timing on the perception or judgment of visual timing. We

believe these latter participants' response patterns suggest that the auditory influence was strongest when the EOS was only in the auditory modality, smallest when the EOS was only in the visual modality, and intermediate when there were conflicting EOSs in both modalities—a pattern that was also obtained in the PCR data, as described later.

Although these cross-modal perceptual interactions and individual differences are interesting and require further investigation, it remains to be seen whether they are relevant to the confluence of temporal information for action control. If they are, then the synchronization results to be reported next should show large individual differences paralleling those seen in the detection results.

### Asynchronies

Before analyzing the asynchronies, we edited the data to eliminate trials that exhibited large phase drift or other anomalies, such as missing taps or extra taps. These problematic trials occurred primarily in the unimodal visual condition with the less experienced participants. The percentage of discarded trials per participant ranged from 0% to 12% in the unimodal visual condition and from 0% to 2% in the other conditions.

Participants' mean asynchronies were computed across all trials from the three taps preceding an EOS, so as not to include variability caused by the EOS or by the initial adaptation to the sequence tempo. The asynchronies were generally negative (i.e., taps preceded event onsets), but there were large individual differences in their magnitudes. The individual data are shown in columns 2–4 of Table 1. The mean asynchronies in the unimodal

Table 1  
Individual Participant Data: Mean Asynchronies (for the Three Pre-EOS Taps, Across all Trials) and Average Standard Deviations of Asynchronies in Isochronous Sequences in the Unimodal Visual, Unimodal Auditory, and Bimodal Conditions

Participant	Mean asynchrony (ms)			Standard deviation (ms)		
	Uni-V	Uni-A	Bimodal	Uni-V	Uni-A	Bimodal
Version 1						
1	-40	-42	-31	21.8	16.5	14.6
2	-27	-59	-25	29.4	16.8	14.0
3 (A.P.)	-111	-110	-111	21.4	10.1	15.5
4	-6	-101	-56	29.9	15.0	15.5
5	-61	-87	-43	30.4	17.5	13.8
6	-29	-75	-62	22.1	14.1	15.7
7	-12	-27	-15	31.4	25.0	15.8
8	-14	-29	-32	33.7	18.6	19.0
(B.R.) <sup>a</sup>	8	-18	-34	17.2	12.2	12.1
Version 2						
9 (B.R.)	1	-18	-16	16.4	9.6	8.7
10	-102	-136	-108	24.8	21.6	21.6
11	-51	-35	-65	23.1	19.3	17.7
12	-57	-28	-40	28.7	13.3	12.0
13	-14	-15	-14	19.1	13.6	10.6

Note. EOS = event onset shift; Uni-V = unimodal visual condition; Uni-A = unimodal auditory condition.

<sup>a</sup> Data for participant B.R. were not included in Version 1 group averages.

visual condition in Version 1 of the experiment were adjusted by subtracting 8 ms, the estimated average delay of the visual events due to the computer screen refresh cycle. The mean asynchrony in the unimodal visual condition was  $-40$  ms, whereas that in the unimodal auditory condition was  $-59$  ms. This difference, which is consistent with previous findings reviewed earlier, was shown by the majority of participants and was nearly significant,  $t(12) = 2.1$ ,  $p < .06$ . (There were 2 participants in Version 2 who showed a difference in the opposite direction.) The asynchronies in the bimodal conditions (nonconflicting and conflicting conditions combined) could be computed either relative to the visual events (i.e., corrected for the 8-ms delay in Part 1) or relative to the auditory events (without correction, as shown in Table 1). The respective mean values were  $-52$  ms and  $-47$  ms; both fell between the means for the unimodal conditions. None of the unimodal-bimodal differences were significant, however. There was no indication of any systematic differences between Versions 1 and 2 of the experiment (see B.R.'s data).

Much more consistent results were obtained with regard to the variability of the asynchronies within trials. To exclude systematic variability due to PCRs and subsequent phase correction, we computed standard deviations only within isochronous (EOS = 0) unimodal and bimodal sequences, from the last 13 taps in each sequence. The standard deviations were then averaged across the 10 isochronous trials in each condition. The results for individual participants are shown in columns 5-7 of Table 1. The average standard deviation in the unimodal visual condition was 25.6 ms, whereas that in the unimodal auditory condition was 16.2 ms. This expected difference in favor of the auditory modality was shown by every participant,  $t(12) = 6.9$ ,  $p < .0001$ . Remarkably, the average standard deviation in the bimodal conditions (nonconflicting and conflicting conditions combined) was 15.0 ms, which was not significantly different from the unimodal auditory condition but was significantly lower than the variability in the unimodal visual condition,  $t(12) = 7.7$ ,  $p < .0001$ . Thus, despite participants' attention to the visual sequences, the simultaneous auditory tones effectively controlled the variability of the asynchronies. It is clear from Table 1 that this was true for all participants but one (A.P.), whose bimodal variability was intermediate between the unimodal variabilities. Thus, it held regardless of the nature of the visual sequences.

### Phase Correction Responses

A PCR to an EOS is an unintended response and usually occurs without awareness. It was defined as the shift (in milliseconds) of the tap immediately following the EOS, relative to the time at which that tap was expected to occur, namely 500 ms after the preceding tap. Thus, the PCR in each trial was computed by subtracting 500 ms from the intertap interval. The tap following an EOS was expected to shift in the same direction as the sequence event; thus, PCRs were expected to be a positive function of EOS magnitude.

The functions relating mean PCR magnitude to EOS magnitude in the different experimental conditions are shown in Figure 3, with between-participants double standard error bars (roughly 95% confidence intervals) and best-fitting regression lines. Each graph also shows the slope of the regression line and the proportion of variance accounted for ( $R^2$ ). The PCRs in the unimodal visual

condition (Figure 3A) were a strongly linear function of EOS magnitude. The slope of .295 means that the tap following a visual EOS shifted by 29.5% of the EOS magnitude, on average. Individual slopes are shown in column 2 of Table 2. It is clear that all participants reacted to visual EOSs and that the slopes were within a fairly narrow range, except for one individual (Participant 8) who reacted much more strongly than the others. These data demonstrate that automatic PCRs do occur in synchronization with visual sequences, despite participants' intention not to react to the EOSs. Moreover, in view of the poor detection performance in the unimodal visual condition (Figure 2A), the results in Figure 3A provide compelling new evidence that phase correction is not limited by perceptual detection thresholds.

The results for the unimodal auditory condition (Figure 3B) largely replicate earlier findings (Repp, 2002a, 2002c). The relationship between PCR and EOS magnitudes was less strikingly linear here, because of an apparent reduction in slope for negative EOSs beyond  $-40$  ms. Nevertheless, the regression line provided a good account of the average data. Its slope was significantly steeper than that obtained in the unimodal visual condition,  $t(12) = 2.9$ ,  $p < .02$ . Column 3 in Table 2 shows that most participants reacted more strongly to auditory than to visual EOSs. The exceptions were us (B.R. and A.P.) and one research assistant (Participant 13); our experience in trying not to react to auditory EOSs, acquired in earlier experiments, may have played a role. Participant 13 somehow managed to be totally unaffected by auditory EOSs, a rare feat that, interestingly, she was not able to repeat in the bimodal conditions, in which her attention was diverted from the auditory modality.

In the bimodal nonconflicting condition, there was again a linear relationship between PCRs and visual EOS magnitude (Figure 3C). However, the slope of the regression line was much more shallow here than in the unimodal visual condition (Figure 3A),  $t(12) = 5.4$ ,  $p < .001$ . Every participant showed this difference (compare columns 2 and 4 in Table 2). Thus, the simultaneously occurring isochronous auditory sequences reduced the PCRs to the visual EOSs. Nevertheless, all participants still showed a positive relationship.

Figure 3D shows that participants responded strongly to the auditory EOSs in the bimodal nonconflicting condition, even though they were trying to ignore the auditory sequences. The slope of the regression line was only slightly less steep than the slope found in the unimodal auditory condition,  $t(12) = 1.8$ ,  $p < .10$ , in part because of a more pronounced leveling off for EOSs beyond  $-40$  ms (see column 5 of Table 2 for the individual slopes). Thus, the isochronous visual sequences had only a slight effect on the involuntary PCRs to the supposedly unattended auditory EOSs.

Finally, Figures 3E and 3F present the results of the bimodal conflicting condition. These two graphs show the same data as a function of visual and auditory EOS magnitudes, respectively. It is clear that the PCRs went in the direction of the auditory rather than the visual EOSs. The leveling off for auditory EOSs beyond  $-40$  ms was also pronounced here. The positive slope of the regression line (Figure 3F) was significantly more shallow than the slope obtained in the unimodal auditory condition (Figure 3B),  $t(12) = 4.1$ ,  $p < .001$ . This indicates that both modalities affected the PCRs to conflicting EOSs. The individual data (negative slopes)



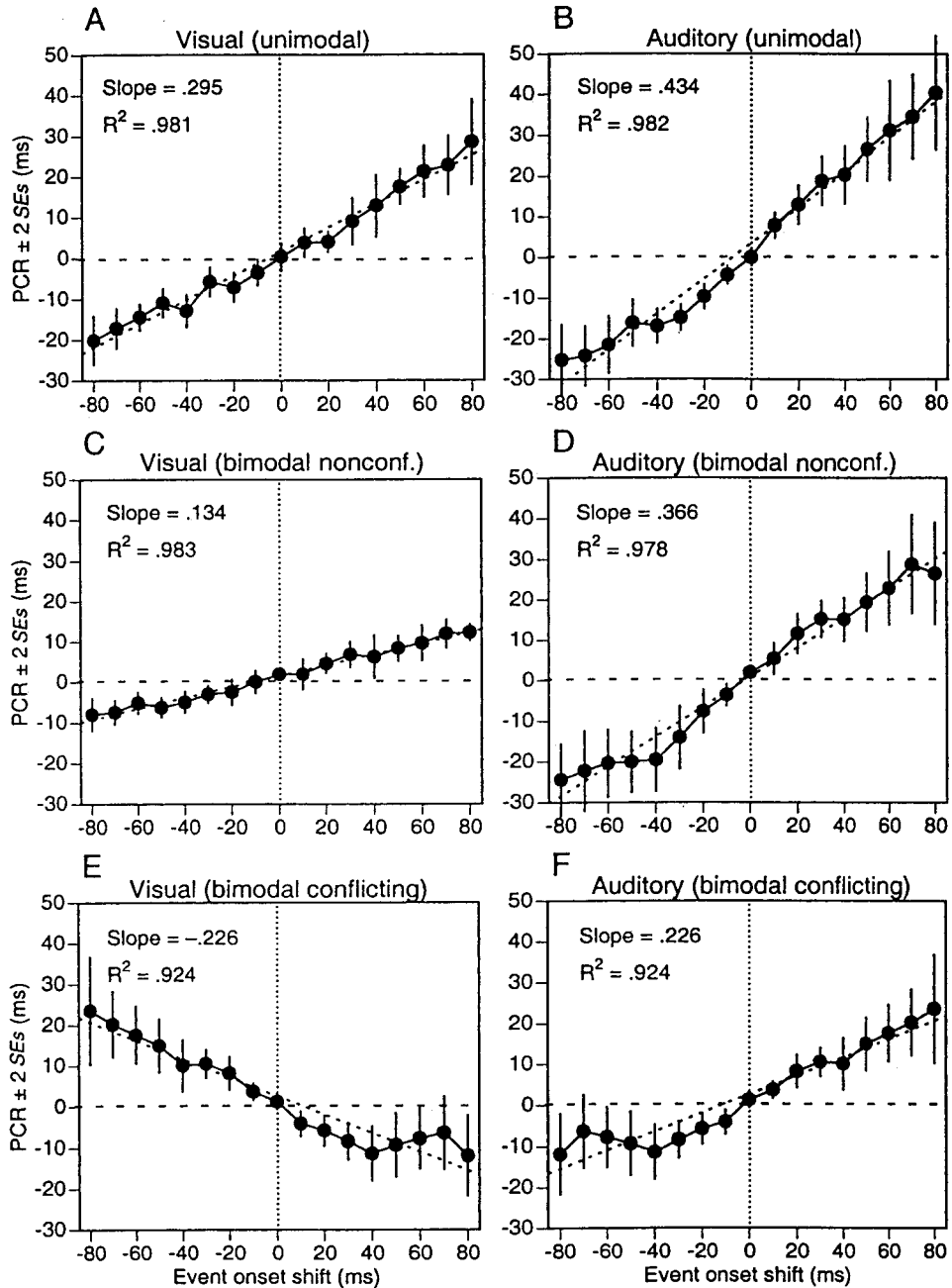


Figure 3. Mean phase correction responses (PCRs) as a function of event onset shift magnitude in the various conditions, as indicated above each graph. (Graphs E and F show the same data as a function of visual and auditory event onset shift magnitudes, respectively.) Also included are double standard error bars, best-fitting regression lines, their slopes, and the proportions of variance accounted for ( $R^2$  values).

are shown in column 6 of Table 2. Only one participant (Participant 4) did not have a negative slope.

The relative influences of the two modalities on the PCRs in the bimodal conditions can be quantified by expressing the slope of the bimodal PCR function as a percentage of the range spanned by the slopes of the relevant unimodal visual and unimodal auditory conditions. Thus, a bimodal slope halfway between the unimodal slopes (50%) would indicate equal influences of each modality on

the PCR, whereas a bimodal slope significantly closer to the unimodal auditory slope than to the unimodal visual slope (greater than 50%) would indicate auditory dominance. If a bimodal slope was steeper than the unimodal auditory slope, as was the case for a few participants, auditory dominance was considered to be 100%.

From the slopes of the mean PCR functions (Figure 3), the average auditory influence in the bimodal nonconflicting condition

Table 2  
*Individual Participant Data: Slopes of Regression Lines Fitted to PCR Functions and Percentages of Auditory (vs. Visual) Influence on the PCR*

Participant	Mean slope of PCR function (ms)					Auditory influence (%)		
	Uni-V	Uni-A	Bim-V	Bim-A	Con(V)	Bim-V	Bim-A	Con
Version 1								
1	.293	.492	.139	.427	-.362	52.6	86.8	83.4
2	.310	.773	.163	.862	-.468	47.4	100.0	71.8
3 (A.P.)	.333	.186	.183	.183	-.144	45.0	98.4	91.9
4	.333	.457	.159	.044	.078	52.3	9.6	32.3
5	.234	.389	.180	.257	-.094	23.1	66.1	52.6
6	.219	.411	.098	.326	-.156	55.3	79.3	59.5
7	.324	.584	.165	.439	-.402	49.1	75.2	80.0
8	.673	.741	.181	.732	-.399	73.1	98.8	75.8
(B.R.) <sup>a</sup>	.356	.200	.160	.242	-.141	55.1	100.0	89.4
Version 2								
9 (B.R.)	.291	.250	.125	.152	-.095	57.0	60.8	71.3
10	.249	.494	.097	.479	-.247	61.0	97.0	66.8
11	.219	.378	.108	.308	-.122	50.7	81.5	57.1
12	.201	.472	.046	.378	-.338	77.1	80.1	80.1
13	.153	.019	.102	.175	-.186	33.3	100.0	100.0

Note. Conditions were unimodal visual (Uni-V), unimodal auditory (Uni-A), bimodal nonconflicting with visual EOSs (Bim-V), bimodal nonconflicting with auditory EOSs (Bim-A), bimodal conflicting [Con; or Con(V) if regarded from the visual perspective]. PCR = phase correction response; EOS = event onset shift.  
<sup>a</sup> Data for participant B.R. were not included in Version 1 group averages.

when the EOS was in the visual modality was found to be  $100 \times (1 - .134/.295) = 54.6\%$ . This average value, which is not significantly different from 50%,  $t(12) = 0.5$ , indicates approximately equal influences of the two modalities on the PCR. The individual coefficients of auditory influence are shown in column 7 of Table 2.

The auditory influence in the bimodal nonconflicting condition when the EOS was in the auditory modality, again estimated from the mean slopes, was  $100 \times (.366/.434) = 84.3\%$ . This clearly indicates auditory dominance. All but one participant (Participant 4) showed percentages greater than 50% in this condition (see column 8 of Table 2),  $t(12) = 4.3$ ,  $p < .001$ . (Participant 4 actually reacted to the auditory EOSs on the second rather than on the first tap, a rare case of delayed PCRs.) The auditory influence was significantly greater here than when the EOS was in the visual modality,  $t(12) = 3.6$ ,  $p < .005$ . We should especially note that the participants who were immune to auditory influences in the detection task (Participants 1, 5, and 9-13) showed auditory dominance comparable with that of the other participants (apart from Participant 4). This suggests that the auditory influence on the PCRs is independent of the auditory influence on the detection responses.

Finally, the average auditory influence in the bimodal conflicting condition was estimated as  $100 \times [(.295 + .226)/(.295 + .434)] = 71.4\%$ , again indicating auditory dominance. All participants but Participant 4 showed percentages greater than 50% (see column 9 in Table 2),  $t(12) = 4.3$ ,  $p < .001$ . The degree of auditory dominance observed was greater here than for visual EOSs in the bimodal nonconflicting condition,  $t(12) = 3.0$ ,  $p < .02$ , but tended to be smaller than for auditory EOSs in the bimodal nonconflicting condition,  $t(12) = 1.9$ ,  $p < .09$ .

It is interesting that the average of the auditory influence estimates obtained for visual and auditory EOSs in the nonconflicting condition (69.4%) is quite similar to the single percentage obtained in the conflicting condition (71.4%). Thus, it seems that there were two independent factors affecting the relative influence of the two modalities on PCRs: an inherent dominance of audition over vision in temporal processing and the modality in which the EOS occurred in the nonconflicting condition. The second effect indicates that a temporal change is given more weight in temporal processing than is the absence of a change, perhaps because it captures the participant's attention. Whether the two factors are truly additive needs to be investigated further.

#### Phase Correction Following the PCR

Each PCR created an asynchrony that had to be corrected during subsequent taps. When the relative shifts of these taps were calculated in the same way as the PCR on the first tap (i.e., relative to the expected phase extrapolated from the tap coinciding with the EOS) and plotted as a function of EOS magnitude (as in Figure 3), the functions were approximately linear and had slopes that decreased and approached zero as the taps were increasingly removed from the PCR. A slope of zero indicates complete phase correction; that is, the relative phase of the tap is no longer affected by the EOS. (Systematic phase drift independent of the EOS was sometimes reflected in the intercept of the regression line but does not concern us here.) Although the time course of phase correction was not of primary interest in this study, we discuss it briefly to illustrate that it, too, was strongly influenced by the auditory input.

Figure 4 shows the slopes of the regression lines for the six taps following an EOS in the various experimental conditions. The

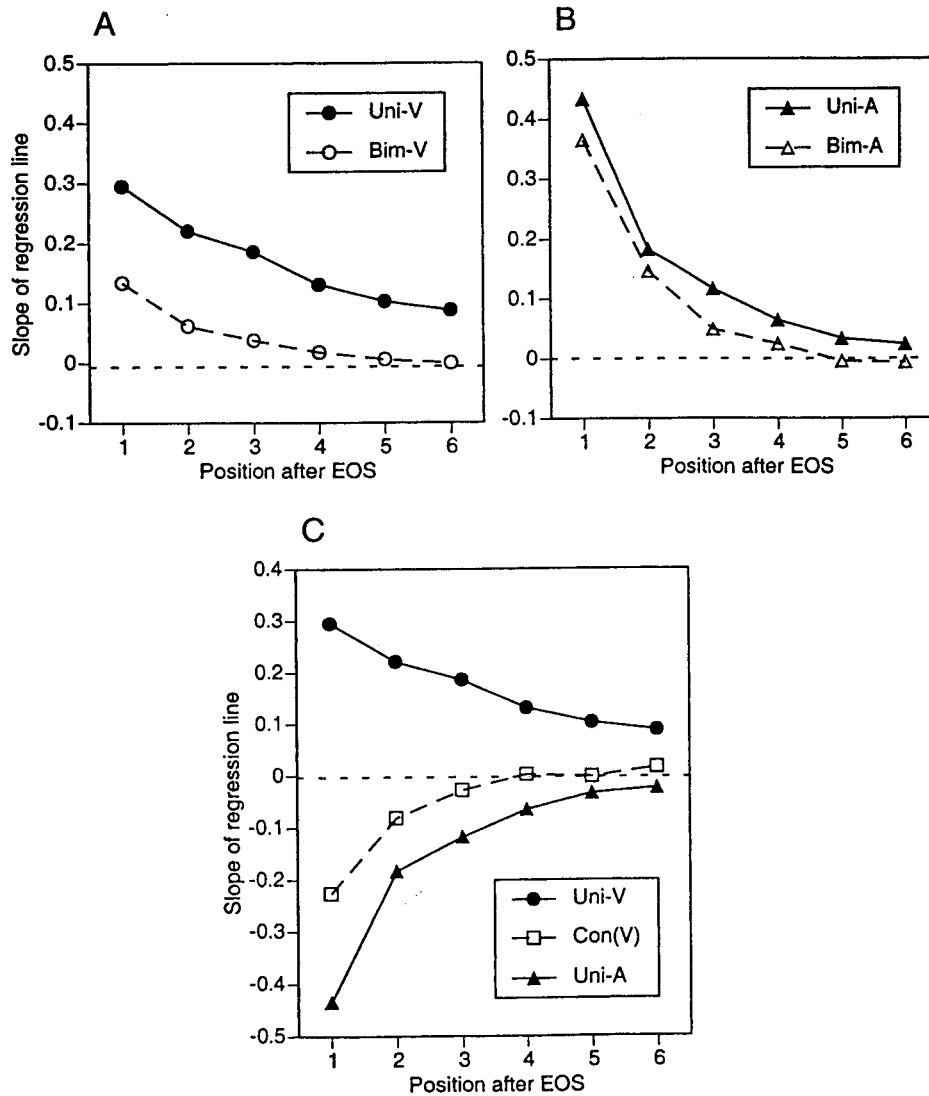


Figure 4. Average time courses of phase correction following event onset shifts (EOSs): Slopes of regression lines as a function of position in the various conditions. The conditions were unimodal visual (Uni-V), unimodal auditory (Uni-A), bimodal nonconflicting with visual EOSs (Bim-V), bimodal nonconflicting with auditory EOSs (Bim-A), and bimodal conflicting regarded from the visual perspective [Con(V)].

slopes in Position 1 are those of the PCR functions (Figure 3). Functions of this general shape are predicted by a linear phase correction model (Mates, 1994a, 1994b; Pressing, 1998; Vorberg & Schulze, 2002; Vorberg & Wing, 1996). Figure 4A shows the unimodal visual condition and the bimodal nonconflicting condition with EOSs in the visual modality. It is evident that phase correction in the unimodal visual condition was very gradual and still incomplete after six taps. (This was mainly true following positive EOSs; we gloss over some asymmetries in the data here.) Phase correction in the bimodal nonconflicting condition was gradual also, which seems to reflect the influence of the visual modality. However, phase correction did result in complete restoration of the original phase by the sixth tap, which must have been due to the auditory influence. Figure 4B shows the unimodal auditory condition and the bimodal nonconflicting condition with

EOSs in the auditory modality. These functions are very similar and show fast initial correction followed by a gradual approximation of the zero baseline. The visual modality made but a slight contribution here. Figure 4C compares the two unimodal conditions with the bimodal conflicting condition. The signs of the unimodal auditory slopes have been reversed here to represent the negative correlation between visual and auditory EOSs (i.e., the functions are drawn from the perspective of the visual modality). The time course of phase correction in the bimodal conflicting condition resembled that in the unimodal auditory condition, except for being more shallow. These data suggest that synchronous auditory and visual sequences constituted a stronger phase attractor than did either sequence alone, so that in each bimodal condition, the original relative phase of the taps was restored after 4–5 taps.

## General Discussion

### *Unimodal Conditions*

The unimodal conditions of our experiment replicate the results of several previous studies that have shown that asynchronies between taps and sequence events are less negative and more variable in synchronization with visual sequences than with auditory sequences (Bartlett & Bartlett, 1959; Dunlap, 1910; Fraisse, 1948; Klemmer, 1967; Kolers & Brewster, 1985). According to the hypothesis that synchronization requires the temporal alignment of central codes based on cumulative sensory evidence (Aschersleben & Prinz, 1995, 1997; Aschersleben et al., 2002; Prinz, 1997), the difference in mean asynchronies suggests a slower growth of sensory evidence for visual than for auditory stimulation. The central sensory evidence for the tap, which in our experiment was mostly tactile and kinesthetic in nature, presumably accumulated at an even slower rate, hence the generally negative asynchronies. However, the considerable individual differences observed, which included some reversals of the modality difference, suggest that this may not be the whole story.

The greater variability of asynchronies in synchronization with visual than with auditory sequences was a highly reliable result, holding for every individual tested and regardless of the type of visual stimulus. It suggests that purely temporal information is conveyed less accurately through the visual than through the auditory modality, and it is also consistent with Fraisse's (1948) claim that auditory rhythms are more strongly coupled to the motor system than are visual rhythms. The former interpretation is further supported by the poor performance in the visual detection task. Detection thresholds for visual EOSs were uniformly higher than those for auditory EOSs, again regardless of the nature of the visual stimulation. A number of participants were unable to detect any visual EOSs within the range presented, just as others (rejected from the experiment) were unable to stay in synchrony with a visual sequence having IOIs of 500 ms.

Of primary interest to us were the involuntary PCRs elicited by the EOSs. Our results for unimodal auditory sequences largely replicate earlier findings (Repp, 2002a, 2002c). In these earlier studies, which likewise used IOIs of 500 ms, the function relating PCR magnitude to auditory EOS magnitude was found to be linear until somewhere between  $\pm 50$  and  $\pm 100$  ms of EOS magnitude, and it then reached an asymptote that extended to EOSs of  $\pm 250$  ms. In one study (Repp, 2002a), but not in its follow-up (Repp, 2002c), there was also a strong asymmetry, with larger PCRs to positive than to negative EOSs, especially near zero. The present data show only a slight asymmetry in that direction, and farther away from zero: PCRs to negative EOSs approached an asymptote toward the end of the range of EOS magnitudes ( $\pm 80$  ms), whereas PCRs to positive EOSs did not. Some differences in average results across experiments may be due to the fact that many of the present participants were novices, some of whom reacted very strongly to auditory EOSs, whereas the previous studies used only practiced participants.

One major new finding is that PCRs also occur to EOSs in visual sequences. The slope of the linear function relating PCR magnitude to EOS magnitude was more shallow than that in the auditory condition, but the relationship was just as systematic and linear. Thus, both auditory and visual EOSs elicit an involuntary, automatic response from the neural system responsible for senso-

rimotor coordination, but the response to visual EOSs is smaller, which indicates a weaker sensorimotor coupling in the visual than in the auditory modality (cf. Fraisse, 1948). The slower and less complete phase correction following PCRs to visual EOSs, compared with the unimodal auditory condition, is also consistent with this interpretation.

The theoretical significance of the PCRs in the visual condition is that they occurred largely without participants' awareness of the EOSs that elicited them. Although we did not test this directly, we feel safe in assuming that the participants were also generally unaware of the asynchronies between their taps and the shifted visual stimuli (cf. Repp, 2000), which presumably triggered the phase correction process (Vorberg & Wing, 1996). Therefore, these results provide support for the claim that sensorimotor coordination is not mediated by conscious perception of temporal information (Repp, 2000, 2001a). Rather, the processes involved in action control seem to have direct (or at least preconscious) access to temporal information. This may be considered a form of *direct parameter specification* (Neumann, 1990). The visual PCR results are stronger than the auditory results in that regard because there is a much wider range over which visual EOSs (and perhaps asynchronies as well) remain subliminal. Whereas special data analyses were necessary to prove that the auditory PCR function is linear in the narrow zone between the detection thresholds for negative and positive EOSs (Repp, 2000), the linearity of the average visual PCR function seems to us to be *prima facie* evidence for the sensitivity of action control processes to subliminal temporal differences.

### *Bimodal Conditions*

In the bimodal conditions, we examined the relative influences of visual and auditory sequence timing on synchronization behavior. We instructed our participants to focus their attention on the visual sequences, thus biasing them in favor of what we expected to be the weaker modality for temporal processing. Several results indicate that these instructions were followed: There were visual influences on detection responses, mean asynchronies, and PCRs. However, the auditory sequences obviously could not be ignored.

The mean asynchronies seemed to be influenced by both the visual and the auditory sequences, but there were substantial individual differences in that regard, and the group results were unclear. In contrast, the variability of the asynchronies exhibited complete auditory dominance overall and in nearly every participant: that is, variability was virtually the same as that for unimodal auditory sequences. In view of other evidence suggesting that participants did not ignore the visual input entirely, this result indicates that the variability of the taps was governed by the modality providing the more reliable temporal information. We believe that this complete auditory dominance with regard to variability was mediated by an internal representation of the sequence period or, equivalently, by entrainment of an internal periodicity (Large & Jones, 1999). The less precise temporal information in the visual modality thus led to a less stable internal representation of the sequence period, which in turn caused more variable tapping and tempo drift. A simultaneous auditory sequence stabilized the internal period. Because it was virtually complete and showed little individual variation, this form of au-

ditory dominance is likely to be independent of the relative influences of the two modalities on the phase correction process.

Influences on phase correction were determined from the PCRs to EOSs in the bimodal conditions. On average, the magnitude of these involuntary and largely subconscious responses reflected about 70% of the auditory and 30% of the visual temporal information. This indicates moderate auditory dominance in a situation in which attention was biased toward vision as much as possible. Undoubtedly, equal or stronger auditory dominance would have been observed had we asked participants to attend to the auditory sequences and to ignore the visual sequences. (Whether PCRs are sensitive to attentional manipulations at all remains to be investigated.) A second factor that influenced the degree of auditory dominance over PCRs was the modality in which an EOS occurred in the bimodal nonconflicting condition. A visual EOS reduced auditory dominance, whereas an auditory EOS increased it, relative to the bimodal conflicting condition. This suggests a role of attention, not of the sustained kind assumed by the directed attention hypothesis (see the introduction), but rather of attention captured momentarily by the shifted event in either modality. This attentional capture would have to be a subconscious process, for it did not depend on, or necessarily result in, conscious detection of EOSs. Visual EOSs in particular were rarely detected; nevertheless, they decreased auditory dominance when there was no competing auditory EOS, compared with the situation in which a conflicting auditory EOS was present. This form of attentional capture seems similar to exogenous covert orienting in space following a cue in either modality (e.g., Spence & Driver, 1997; Vroomen et al., 2001), but here it was a temporal change that seemed to cause participants to orient more toward the modality in which the change occurred.

Our findings are consistent with all three hypotheses discussed in the literature reviewed in the introduction: The findings are consistent with the modality precision hypothesis, because audition provides more precise temporal information than does vision; they are consistent with the directed attention hypothesis, because auditory rhythms may automatically attract more attention than visual rhythms (contrary to participants' intentions); and even with the modality appropriateness hypothesis, because there may be an inherently stronger coupling between auditory rhythms and motor behavior (Fraisse, 1948). It was not our aim to rule out any of these hypotheses here, which are by no means mutually exclusive. What we can conclude is that auditory dominance for temporal information occurs even when participants do not have to make explicit judgments of timing or tempo, which have been required in most previous research on this subject.

An important theoretical question is whether the auditory dominance reflected in PCRs arises from different weightings of independently registered auditory and visual temporal information at the level of action control, or whether it is mediated by a perceptual interaction, particularly by an influence of auditory timing on perceived visual timing. Our study did not address this question head-on, but several results suggest that action control is independent of perceptual judgment. First, although detection responses in the bimodal nonconflicting condition revealed an auditory influence on responses that, according to instructions, pertained to the visual modality, this influence was exhibited by only 6 of 13 participants, and for them it was very large. These striking individual differences make it doubtful that the influence is truly

perceptual. If it represents a postperceptual interaction, it would seem even less likely to mediate action control. Second, those participants who did not show any substantial auditory influence on detection responses nevertheless exhibited auditory dominance in their PCRs. In other words, the PCRs were independent of the perceptual judgments. Third, the finding that phase correction is independent of conscious detection of phase shifts or asynchronies (Repp, 2000, 2001a, 2002a), which was extended to visual EOSs in the present study, as well as a recent demonstration that perception of timing is more context sensitive than sensorimotor synchronization (Repp, 2002b), suggest a dissociation between perception and action with regard to the processing of temporal information. Therefore, any audiovisual perceptual interactions revealed in temporal judgments (Fendrich & Corballis, 2001) are probably irrelevant to action control at the level of sensorimotor coordination. This conclusion is consistent with a host of other recent results suggesting dissociations between perception and action, although usually in the visuospatial domain (e.g., Creem & Proffitt, 1998; Goodale & Humphrey, 1998; Klotz & Neumann, 1999; Neumann, 1990).

In conclusion, we have presented new evidence that audition is superior to vision in the precision of temporal information conveyed and that auditory temporal information is given a correspondingly greater weight by the action control system. Our results extend earlier perceptual findings by showing that auditory dominance over vision in temporal processing extends to preperceptual, subconscious domains of action control that make direct use of temporal information from multiple modalities. In fact, we believe we have uncovered two different and probably independent effects at that level: one on the precision of the internal representation of the period of a rhythm (i.e., on central timekeeper variability) and the other on the reactivity of the phase correction process (which makes local adjustments to timed intervals). This distinction is consistent with a two-process model of timing control in synchronization (Large & Jones, 1999; Mates, 1994a; Repp, 2001b). In addition, we have found that some (but not all) participants' perceptual judgments are strongly influenced by supposedly unattended auditory information. This is a third kind of audiovisual interaction that we believe to be independent of the two interactions found in sensorimotor synchronization and whose precise nature (perceptual or postperceptual) remains to be clarified.

We need to qualify our conclusions concerning auditory dominance by admitting that we have not yet begun to explore a variety of sensory and cognitive factors that may affect the extent of the auditory and visual influences at different levels of temporal processing. It may well be possible to contrive a task situation in which vision dominates audition with regard to temporal processing, but we suspect that auditory dominance will prove to be the typical result in most situations.

## References

- Aschersleben, G., & Prinz, W. (1995). Synchronizing actions with events: The role of sensory information. *Perception & Psychophysics*, *57*, 305-317.
- Aschersleben, G., & Prinz, W. (1997). Delayed auditory feedback in synchronization. *Journal of Motor Behavior*, *29*, 35-46.
- Aschersleben, G., Stenken, P., Cole, J., & Prinz, W. (2002). Timing

- mechanisms in sensorimotor synchronization. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 227–244). Oxford, England: Oxford University Press.
- Bartlett, N. R., & Bartlett, S. C. (1959). Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, *66*, 203–218.
- Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *American Journal of Psychology*, *74*, 17–26.
- Bermant, R. I., & Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills*, *43*, 487–493.
- Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*, 482–489.
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*, 578–584.
- Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321–332.
- Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, *84*, 141–147.
- Canon, L. K. (1971). Directed attention and maladaptive adaptation to displacement of the visual field. *Journal of Experimental Psychology*, *88*, 403–408.
- Creem, S. H., & Proffitt, D. R. (1998). Two memories for geographical slant: Separation and interdependence of action and awareness. *Psychonomic Bulletin & Review*, *5*, 22–36.
- de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*, 289–311.
- Dunlap, K. (1910). Reactions to rhythmic stimuli, with attempt to synchronize. *Psychological Review*, *17*, 399–416.
- Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*, 719–725.
- Fraisse, P. (1948). Rythmes auditifs et rythmes visuels [Auditory rhythms and visual rhythms]. *L'Année Psychologique*, *49*, 21–41.
- Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, *81*, 284–310.
- Gault, R. H., & Goodfellow, L. D. (1938). An empirical comparison of audition, vision, and touch in the discrimination of temporal patterns and ability to reproduce them. *Journal of General Psychology*, *18*, 41–47.
- Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology*, *72*, 521–528.
- Geldard, F. A. (1970). Vision, audition, and beyond. In W. D. Neff (Ed.), *Contributions to sensory physiology* (pp. 1–17). New York: Academic.
- Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The study of time* (Vol. 2, pp. 29–58). Berlin, Germany: Springer.
- Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1959). Intersensory comparisons of temporal judgments. *Journal of Experimental Psychology*, *57*, 243–248.
- Goldstone, S., & Lhamon, W. T. (1971). Levels of cognitive functioning and the auditory-visual differences in human timing behavior. In M. H. Appley (Ed.), *Adaptation level theory (symposium)* (pp. 263–280). New York: Academic Press.
- Goldstone, S., & Lhamon, W. T. (1974). Studies of auditory-visual differences in human time judgments: 1. Sounds are judged longer than lights. *Perceptual and Motor Skills*, *39*, 63–82.
- Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. *Bulletin of the Psychonomic Society*, *12*, 83–84.
- Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, *67*, 181–207.
- Goodfellow, L. D. (1934). An empirical comparison of audition, vision, and touch in the discrimination of short intervals of time. *American Journal of Psychology*, *46*, 243–258.
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, *54*, 383–394.
- Grondin, S., Meilleur-Wells, G., Ouellette, C., & Macar, F. (1998). Sensory effects on judgments of short time-intervals. *Psychological Research*, *61*, 261–268.
- Grondin, S., Ouellet, B., & Roussel, M.-E. (2001). About optimal timing and stability of Weber fraction for duration discrimination. *Acoustical Science and Technology*, *22*, 370–372.
- Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception & Psychophysics*, *49*, 245–256.
- Hary, D., & Moore, G. P. (1985). Temporal tracking and synchronization strategies. *Human Neurobiology*, *4*, 73–77.
- Hary, D., & Moore, G. P. (1987). Synchronizing human movement with an external clock source. *Biological Cybernetics*, *56*, 305–311.
- 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*, 967–969.
- Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*, 52–65.
- Jaśkowski, P. (1999). Reaction time and temporal-order judgment as measures of perceptual latency: The problem of dissociations. In G. Aschersleben, T. Bachmann, & J. Müseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 265–291). Amsterdam: Elsevier.
- Jaśkowski, P., Jaroszyk, F., & Hojan-Jezińska, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, *52*, 35–38.
- Klemmer, E. T. (1967). Sequences of responses to signals encoded in time only. *Acta Psychologica*, *27*, 197–203.
- Klotz, W., & Neumann, O. (1999). Motor activation without conscious discrimination in metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 976–992.
- Kolers, P. A., & Brewster, J. M. (1985). Rhythms and responses. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 150–167.
- Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 318–320.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time-varying events. *Psychological Review*, *106*, 119–159.
- Lhamon, W. T., & Goldstone, S. (1974). Studies in auditory-visual differences in human time judgment: 2. More transmitted information with sounds than lights. *Perceptual and Motor Skills*, *39*, 295–307.
- Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*, 215–221.
- Mates, J. (1994a). A model of synchronization of motor acts to a stimulus sequence: I. Timing and error corrections. *Biological Cybernetics*, *70*, 463–473.
- Mates, J. (1994b). A model of synchronization of motor acts to a stimulus sequence: II. Stability analysis, error estimation, and simulations. *Biological Cybernetics*, *70*, 475–484.
- McGurk, H., & MacDonald, J. (1976, December 23/30). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Michon, J. A. (1967). *Timing in temporal tracking*. Assen, the Netherlands: van Gorcum.

- Myers, A. K., Cotton, B., & Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance. *Perception & Psychophysics*, *30*, 33–38.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859.
- Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research*, *52*, 207–215.
- O'Connor, N., & Hermelin, B. (1972). Seeing and hearing and space and time. *Perception & Psychophysics*, *11*, 46–48.
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1770–1787.
- Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflicts in judgments of spatial direction. *Perception & Psychophysics*, *6*, 203–205.
- Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origin and significance. *Psychological Review*, *83*, 157–171.
- Pressing, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology*, *42*, 63–101.
- Pressing, J. (1999). The referential dynamics of cognition and action. *Psychological Review*, *106*, 714–747.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*, 129–154.
- Radeau, M. (1985). Signal intensity, task context, and auditory–visual interaction. *Perception*, *14*, 517–577.
- Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics*, *20*, 227–235.
- Radeau, M., & Bertelson, P. (1977). Adaptation to auditory–visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, *22*, 137–146.
- Radeau, M., & Bertelson, P. (1987). Auditory–visual interaction and the timing of inputs: Thomas (1941) revisited. *Psychological Research*, *49*, 17–22.
- Repp, B. H. (2000). Compensation for subliminal timing perturbations in perceptual–motor synchronization. *Psychological Research*, *63*, 106–128.
- Repp, B. H. (2001a). Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 600–621.
- Repp, B. H. (2001b). Processes underlying adaptation to tempo changes in sensorimotor synchronization. *Human Movement Science*, *20*, 277–312.
- Repp, B. H. (2002a). Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 410–430.
- Repp, B. H. (2002b). Perception of timing is more context-sensitive than sensorimotor synchronization. *Perception & Psychophysics*, *64*, 703–716.
- Repp, B. H. (2002c). Phase correction in sensorimotor synchronization: Nonlinearities in voluntary and involuntary responses to perturbations. *Human Movement Science*, *21*, 1–37.
- Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception & Psychophysics*, *34*, 541–548.
- Semjen, A., Schulze, H.-H., & Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological Research*, *63*, 137–147.
- Semjen, A., Vorberg, D., & Schulze, H.-H. (1998). Getting synchronized with the metronome: Comparisons between phase and period correction. *Psychological Research*, *61*, 44–55.
- Shipley, T. (1964, September). Auditory flutter-driving of visual flicker. *Science*, *145*, 1328–1330.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*, 1–22.
- Stratton, G. M. (1897). Vision without inversion of the retinal image. *Psychological Review*, *4*, 341–360 and 463–481.
- Tanner, T. A., Patton, R. M., & Atkinson, R. C. (1965). Intermodality judgments of signal duration. *Psychonomic Science*, *2*, 271–272.
- Thaut, M. H., Miller, R. A., & Schauer, L. M. (1998). Multiple synchronization strategies in rhythmic sensorimotor tasks: Phase vs. period correction. *Biological Cybernetics*, *79*, 241–250.
- Thaut, M. H., Tian, B., & Azimi-Sadjadi, M. R. (1998). Rhythmic finger tapping to cosine-wave modulated metronome sequences: Evidence of subliminal entrainment. *Human Movement Science*, *17*, 839–863.
- Thomas, G. J. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, *28*, 163–177.
- Vorberg, D., & Schulze, H.-H. (2002). A two-level timing model for synchronization. *Journal of Mathematical Psychology*, *46*, 56–87.
- Vorberg, D., & Wing, A. (1996). Modelling variability and dependence in timing. In H. Heuer & S. W. Keele (Eds.), *Handbook of perception and action* (Vol. 2, pp. 181–262). London: Academic Press.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*, 651–659.
- 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.
- Warren, D. H. (1979). Spatial localization under conflict conditions: Is there a single explanation? *Perception*, *8*, 323–337.
- Warren, D. H. (1980). Response factors in intermodality localization under conflict situation. *Perception & Psychophysics*, *27*, 28–32.
- Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual–auditory compellingness in the ventriloquist effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*, 557–564.
- Welch, R. B. (1999). Meaning, attention, and the unity assumption in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, G. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam: Elsevier.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638–667.
- Witkin, H. A., Wapner, S., & Leventhal, T. (1952). Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology*, *43*, 58–67.

Received June 22, 2001

Revision received October 16, 2001

Accepted February 12, 2002 ■