Perception & Psychophysics
2001, 63 (4), 719-725

The temporal cross-capture of audition and vision

ROBERT FENDRICH and PAUL M. CORBALLIS
Dartmouth College, Hanover, New Hampshire

We report that when a flash and audible click occur in temporal proximity to each other, the perceived time of occurrence of both events is shifted in such a way as to draw them toward temporal convergence. In one experiment, observers judged when a flash occurred by reporting the clock position of a rotating marker. The flash was seen significantly earlier when it was preceded by an audible click and significantly later when it was followed by an audible click, relative to a condition in which the flash and click occurred simultaneously. In a second experiment, observers judged where the marker was when the click was heard. When a flash preceded or followed the click, similar but smaller capture effects were observed. These capture effects may reveal how temporal discrepancies in the input from different sensory modalities are reconciled and could provide a probe for examining the neural stages at which evoked responses correspond to the contents of conscious perception.

When there is a conflict between vision and another sensory modality, vision usually dominates. The best known illustration of this general finding is the *ventriloquism effect* (Howard & Templeton, 1966): One *hears* a sound coming from the mouth of a ventriloquist's dummy because it *looks* as if the dummy is speaking. Another case is the biasing of proprioceptive information by vision: When observers wear lenses that cause straight edges to look curved, those edges *feel* curved when touched with the hand (Easton & Moran, 1978; Hay, Pick, & Ikeda, 1965). Hay et al. termed this phenomenon *visual capture*. Generally, information from the conflicting modality has only a small effect or no effect on the perception of visual attributes; apparent visual locations are not altered, for example, by spatially conflicting auditory signals (Pick, Warren, & Hay, 1969).

The phenomenon of *auditory driving* reported by Gebhard and Mowbray (1959) represents a striking exception to the general finding that vision dominates other modalities. If observers are asked to judge the rate at which a light is flickering when that light is presented together with a repeating ("fluttering") sound, increasing or decreasing the flutter rate can cause the apparent flicker rate to increase or decrease in tandem. Gebhard and Mowbray found no indication of the reverse phenomenon—that is, varying flicker rates did not change the perception of concurrent flutter. Auditory driving represents a case of the auditory temporal capture of vision. Shipley (1964) attempted to ascertain the capture range of auditory driving by varying flicker rates in the presence of fixed flutter until observers reported that they were clearly different. He found that a 10-Hz flutter could entrain flickers from 7 Hz up to 22 Hz. Welch, DuttonHurt, and Warren (1986) attempted to measure the strength of auditory driving using magnitude estimation. Observers were initially presented with 2-Hz flicker and flutter and were told this rate had a value of 2. They then reported values to describe other flicker or flutter frequencies, which were presented alone or in combination. When flicker and flutter rates were discrepant, reported flicker rates shifted toward flutter rates so as to eliminate an average of 52% of the discrepancy. Flutter rates also shifted toward flicker rates but to a much smaller extent, eliminating an average of 13% of flicker–flutter discrepancy.

Welch and Warren (1980) proposed that what they term *modality appropriateness* could account for this reversal of the normal finding of visual dominance. They argued that vision is specifically designed to process spatial information, whereas audition is designed to process temporal information. In most investigations of intersensory conflict, a judgment regarding some spatial characteristic of the stimulus is required (e.g., location, tilt, or shape). Since vision is the sensory modality best suited to evaluate the desired attribute, it is given precedence and information from the other modality is adjusted to eliminate the conflict. Conversely, in the case of temporal processing, the auditory system is given precedence, and vision is adjusted to fit the auditory input.

Although this hypothesis fits with much of the data on intersensory conflicts, it does not address the stimulus conditions that lead to the yoking of information across different sensory modalities. High-level interpretations no doubt play a role in this yoking but are not required for it to occur. Although the ventriloquism effect is most effective when it is reasonable to assume that the seen stimulus is a likely source of the sound, this "assumption

---

This work was supported by NIH Grant NS17778 and by a grant from the McDonnell-Pew foundation. We thank Howard Hughes for his suggestions regarding the design of these experiments and George Wolford for his suggestions regarding their interpretation. Correspondence regarding this article should be addressed to R. Fendrich, Center for Cognitive Neuroscience, Dartmouth College, 6162 Moore Hall, Hanover, NH 03755 (e-mail: fendrich@dartmouth.edu).

of unity" (Welch & Warren, 1980) is not essential. A simple flash of light can bias the perceived location of a concurrent tone burst even when observers have no reason (other than temporal contiguity) to assume that the two are yoked (Bermant & Welch, 1976). This suggests that, in addition to any interpretive factors, low-level intersensory linking processes may be contributing to the capture phenomenon. Auditory driving seems even more likely to depend on such low level sensory linking processes; there is no evident reason for observers to assume that the flashing light and the fluttering sound are repeating at similar rates. Auditory driving could therefore provide a probe for exploring the nature of these processes.

The usefulness of auditory driving for the study of intersensory interactions has been limited, however, by problems inherent in its measurement. Direct comparisons of flicker with flutter confound auditory capture with any visual capture that may also be occurring, whereas indirect measurement methods such as magnitude estimation lack the precision of direct comparison techniques. Moreover, the temporal relationship between the flashes of the flickering light and the acoustic peaks of a fluttering sound will shift continuously when the two differ in frequency. Therefore, auditory driving cannot be used to study how this relationship influences temporal capture. We report an instance of the auditory capture of vision that overcomes these difficulties. It entails a capture effect that is akin to the phase capture of a repeating flash by a repeating sound.

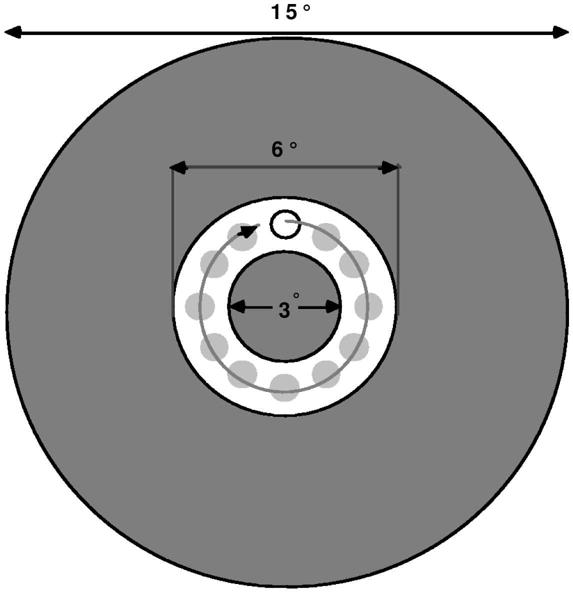

Figure 1. Experiment 1 display. The rotating marker is represented at its 12 o'clock position by the small outline circle. The small gray circles show its successive positions. If the flash (produced by the disappearance of the inner circle and outer ring) were set to occur at the 12 o'clock position, the audible click would be presented at 10, 11, 12, 1, or 2 o'clock. Experiment 2 used the same display, with the roles of the click and flash reversed.

EXPERIMENT 1

Welch and Warren (1980) observed that to obtain a "pure" measure of the temporal capture of vision by audition, one could use some unaffected sensory modality (e.g., touch) to gauge the capture effect. We thought this goal might also be achieved by the use of a spatial visual metric that would not be affected by the auditory stimulus. To do this, we asked observers to report the position of a moving visual target when a flash occurred, and then we assessed the effect of a temporally offset sound burst on that reported position. We reasoned that, since there would be no basis for associating the sound with any particular position in the target's path, the sound would not have any direct effect on the perceived location of the moving target. In addition, as noted above, visual *spatial* attributes are normally not subject to modification by auditory inputs. Therefore, if a sound alters the apparent position of the moving target at the time a flash is presented, this effect should be attributable to a change in the perceived timing of the flash.

Method

Observers. Fourteen naive observers were recruited from an introductory psychology class at Dartmouth College. They participated in this experiment in return for extra credit. All observers had normal or corrected-to-normal vision. The observers were equally distributed with respect to gender and ranged in age from 18 to 20 years.

Apparatus and Stimuli. Displays were produced with an IBM-type personal computer and presented as grayscale images on a 14-in. VGA monitor. Observers viewed the monitor screen at a distance of 57 cm in a dimly lit room with their heads positioned by a chin cup. As shown in Figure 1, a circular path was defined by surrounding a 3° gray (14 cd/m^2) circle with a concentric gray ring. A 0.5° diameter circular marker rotated around this path, completing a circuit once every 600 msec. The marker advanced clockwise through 12 equally spaced positions (arranged like the numbers on a clock face), remaining at each position for 50 msec. To observers viewing the display, the motion of this marker appeared continuous. When the marker was at 1 of its 12 positions, the gray circle and ring were switched to the background luminance (70 cd/m^2) for the 50-msec interval, producing the impression of a transient bright flash that completely encompassed the marker path. In addition to the flash, a speaker above the display screen produced an audible click when the marker was at 1 of its 12 positions. The click was produced by driving the speaker with a 1-msec pulse at the midpoint of the selected interval. It measured 76 dB at the distance of the observer's head. The position where the click occurred was one or two clock-face steps before the flash position, or at the same position as the flash, or one or two steps after the flash position. These spatial offsets corresponded to temporal offsets of −100, −50, 0, +50, or +100 msec.

Procedure. At the start of each trial, the display appeared and the marker began to rotate immediately, starting at its 12 o'clock position. The marker continued to rotate repetitively around the ring until the observers reported its apparent position when the flash occurred. They did this by pressing a key on the computer keyboard to terminate the display and then entered the position as a clock position to the closest half hour (so that there were 24 possible responses). On each trial, the click and flash occurred at the same marker positions each time the marker cycled through the positions. The observers were informed that the click and flash would not necessarily occur at the same time and that they should be careful to attend only to the flash. However, in order to provide an opportunity for the clicks to entrain the initial flash presentation, the marker completed three rotations with only the clicks presented before the flashes commenced.

Four blocks of 30 trials were run on each observer. In order to avoid longer trial blocks (which pilot observations suggested would increase the incidence of large position judgment errors), the actual flash position was limited to the even clock positions in two of the blocks and to the odd clock positions in the other two. Within each of these blocks, there were five presentations at each of the six possible flash positions—one for each of the five click–flash offsets. When asked, no observer subsequently indicated any awareness of the odd–even display constraint. The order of the presentations was randomized within each block, and the order of the blocks was counterbalanced across observers. The observers paced themselves and used the computer keyboard to initiate each trial. No fixation instructions were given. Typically, a trial block took about 15 min to complete, and there was a 5- to 10-min break between the blocks.

Predictions. When the click and flash occurred together, we expected that the observers would be reasonably accurate in judging their clock position. Previous evidence of auditory temporal capture led us to expect that when the click preceded the flash, the judged position of the marker when the flash occurred would be shifted backward to reflect the timing of the click. Auditory capture would also be evidenced by a forward shift in the perceived marker position when the click followed the flash. However, we viewed this result as more problematical because it seemed possible that the flash would be processed ahead of the click at every neural stage and reach awareness before any capture could occur. It was also difficult to predict the consequence of increasing the temporal discrepancy between the click and flash presentations: Depending on the temporal range of any capture effect, which could not be estimated from any existing data, either an increase or a decrease in the capture strength seemed possible.

Results

Results for each observer were obtained by collapsing across all 12 actual flash positions and calculating the mean difference between the actual and reported flash position for each click–flash temporal offset. However, trials in which the reported flash position differed from the actual flash position by more than three clock steps were regarded as suspicious and excluded from the calculation. This resulted in the removal of 16 trials from the total set of 1,680. In the case of the “worst” observer, 6 (out of 120) trials were eliminated. No more than 2 trials were eliminated from any other observer’s data set, and no trials were eliminated for 7 of the 14 observers. Position error judgments were converted to temporal judgment errors by counting each clock step as equivalent to 50 msec. Mean across-observer position judgment errors for each of the 5 click–flash offsets are given in Table 1.

Table 1

Errors in Judgments of the Time (in Milliseconds) the Flash Occurred for the Five Temporal Offsets of the Click (Experiment 1) and the Time the Click Occurred For the Five Temporal Offsets of the Flash (Experiment 2)

	Error in Judged Flash/Click Time			
	Experiment 1		Experiment 2	
Click/Flash Offset	*M*	*SE*	*M*	*SE*
−100	−57.5	6.68	−17.0	7.35
−50	−40.0	6.94	−6.0	7.34
0	−20.0	5.37	9.0	6.04
50	2.5	6.82	23.0	6.14
100	1.0	7.88	24.5	7.82

Eleven of the 14 observers had a negative bias in their estimates of the marker position when the click and flash occurred in the same clock position, which resulted in a mean error of −20 msec (−0.4 clock steps) in the 0-offset condition. This negative bias is significantly different from 0 [$t(13) = 3.74$, $p < .01$]. We do not know the source of the bias but presume it was present in all the conditions, so that the 0-offset condition is the appropriate reference point for the evaluation of capture effects. We note, in this regard, that in subsequent testing we found a comparable negative bias when only the flash was presented (see Experiment 3). Results are graphed in Figure 2 with the bias in the 0-offset condition subtracted from all the condition means.

As expected, the observers reported the rotating marker at an earlier position (relative to the 0-offset condition) when the click preceded the flash. They also reported the marker at a later position when the click followed the flash, demonstrating that the auditory capture effect of the click could act backward in time. It can be seen that click–flash offsets ranging from −100 to +50 msec (−2 to +1 clock steps) produced a linear shift in the judged position of the marker when the flash occurred. However, increasing the temporal offset of the click from +50 to +100 msec had essentially no effect. A one-way repeated measures analysis of variance (ANOVA) indicates that the overall shift in the marker position judgments is highly significant [$F(4,52) = 32.31$, $p < .001$]. We conducted (student Newman-Keuls) pairwise comparisons between the mean reported marker positions for each click–flash temporal offset). These revealed that the marker was reported at a significantly earlier position when the offset was −100 versus −50 msec ($p < .05$), −50 versus 0 msec ($p < .01$), and 0 versus +50 msec ($p < .01$). There was no significant difference in the reported marker position for +50- versus +100-msec offsets. The slope of the regression line for the reported marker positions as a function of the click–flash offset is .32 if one considers all five offsets and .40 if only the −100- to +50-msec subset of offsets is considered.

EXPERIMENT 2

Experiment 2 was designed to assess whether there would be any temporal visual capture of a click by a temporally offset flash.

Method

Observers. Fourteen naive observers who had not participated in Experiment 1 were recruited from an introductory psychology course at Dartmouth College. The observers participated in return for either course credit or payment of $10. They ranged in age from 18 to 20 years, and both genders were equally represented. All observers had normal or corrected-to-normal vision.

Stimuli and Procedure. The method in Experiment 2 was identical to that in Experiment 1, save that the roles of the click and the

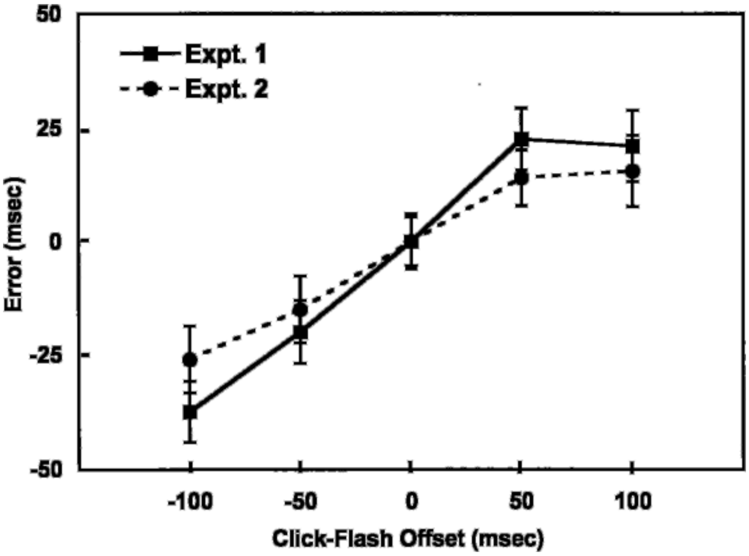

Figure 2. Errors in judgments of the time the flash was presented (Experiment 1) and the time the click was presented (Experiment 2) for each temporal offset between the flash and click. Each 50-msec step on the abscissa represents one clock position step of the rotating marker. In each experiment the temporal bias in the 0-offset condition has subtracted from all the condition means. Plotted values therefore show the difference between the mean of each condition and the mean of the 0-offset condition (which is fixed at 0). Error bars show between-observer standard errors.

flash were reversed. The click position was held constant and the flash was offset from the click by −2, −1, 0, 1, or 2 clock positions (−100, −50, 0, 50, or 100 msec). Observers reported the clock position of the revolving marker when they heard the click. They were cautioned that the click and flash might not occur at the same time and that they should be careful to attend to the click. Because the flash was essentially omnidirectional with respect to the marker, it conveyed no directional information that could influence the perception of the marker's position. Any change in the apparent position of the marker produced by the flash should therefore be attributable to its effect on the perceived timing of the click.

Predictions. Since, in the auditory driving literature, only a small effect or no effect of flicker rates on flutter rates has been reported, we expected that there would be little, if any, temporal capture of the click by the flash.

Results

Results were analyzed in the same manner as those of Experiment 1. Overall, 16 trials were eliminated because position judgment errors were greater than three clock positions: 4 from the data set of each of the 2 worst observers and 3 or less from 5 additional observers. Mean position judgment errors across observers for each of the five click–flash offsets are given in Table 1. With a 0 temporal offset, the bias in the reported marker position was +9 msec (0.18 clock steps). The bias is not significantly different from 0 [$t(13) = 1.53$, n.s.]. However, in order to facilitate a direct comparison of the results of Experiments 1 and 2, the results of Experiment 2 are graphed in Figure 2 with the offset subtracted from all the condition means.

Contrary to our expectations, there was a temporal visual capture of the click by the flash. Relative to the 0-temporal-offset condition, the click was heard earlier (based on the reported marker position) if the flash preceded the click and later if the flash followed it. This visual capture effect is significant with a one-way repeated measures ANOVA [$F(4,52) = 38.04$, $p < .001$], but smaller than the auditory capture obtained in Experiment 1. As in Experiment 1, the capture effect appears linear for offsets from −100 to +50 msec, but does not increase with the +100-msec offset relative to the +50-msec offset. We again performed pairwise comparisons to evaluate the effect of each of the successive click–flash offset steps. These comparisons revealed that the marker was reported at a significantly earlier position when the offset was –100 msec versus –50 msec ($p < .05$), –50 versus 0 msec ($p < .05$), and 0 versus +50 msec ($p < .05$). Again, there was no significant difference in the reported marker position for +50- versus +100-msec offsets. The slope of the regression line for the reported marker position as a function of the flash–click offset is .22 if one considers all five offset conditions and .27 if only the −100- to +50-msec subset of offsets is considered.

Visual Versus Auditory Capture

The temporal capture of the flash by the click in Experiment 1 appeared to be stronger than the capture of the click by the flash in Experiment 2. To determine the statistical reliability of this difference, we combined the data from the two experiments and performed a two-way ANOVA, treating the temporal offset between the click and flash as a repeated measures factor (*offset*) and the type of capture effect as an independent groups factor (*capture type*). In such an analysis, any difference between the strength of the auditory and visual capture effects will be reflected by an offset × capture type interaction. When all levels of offset were considered, there was a significant main effect of both offset [$F(4,104) = 65.71, p < .001$] and capture type [$F(1,26) = 12.01, p < .01$]. The latter reflects the tendency for the flash to be judged as occurring earlier than the click, but the interaction of these factors falls short of significance [$F(4,104) = 2.23, p \approx .12$]. However, with a full data set, the difference between auditory and visual capture data is diluted by the absence in both cases of a difference between the +50- and +100-msec offset conditions. We therefore performed a second ANOVA in which we analyzed only the −100- to +50-msec offset conditions. In this analysis, the main effects of offset [$F(3,78) = 84.35, p < .001$] and capture direction [$F(1,26) = 13.22, p < .001$] do interact significantly [$F(3,78) = 3.22, p < .05$]. These data suggest that for the range of temporal offsets in which the capture processes operated effectively, the auditory temporal capture of vision was stronger than the visual temporal capture of audition.

EXPERIMENT 3

In Experiment 1, we found a negative bias in judgments of the time of the flash, even when the flash and click were presented in the same marker interval. In order to determine whether this bias was tied in some way to the presentation of the click, we ran an additional experiment in which only the flash was presented. For completeness, we also included a *click-only* condition. We also ran a *brief-flash* condition to test the hypothesis that the bias occurred because the observers were reporting the apparent position of the rotating marker at the onset of the 50-msec flash rather than at its temporal midpoint (see the General Discussion section).

Method

Observers. Sixteen naive observers who had not participated in either of the previous experiments were recruited from an introductory psychology course at Dartmouth College. They received course credit in return for their participation. Both genders were equally represented. The observers ranged in age from 18 to 20 years. All had normal or corrected-to-normal vision.

Stimuli and Procedure. In a flash-only condition, the display and procedure in each trial were the same as in Experiment 1, but no click was presented. A single block of 36 trials was run, with the flash presented three times at each of the 12 marker positions. In the click-only condition, the procedure was identical, but the click was substituted for the flash. The brief-flash condition was identical to the flash-only condition, but the flash duration was reduced to a single 16.7-msec video frame centered in the 50-msec interval during the time the marker was at the selected position.

All observers participated in the flash-only condition. Eight of them were also run in the click-only condition and 12 in the brief-flash condition. The order of the click-only and flash-only conditions was counterbalanced across observers. The brief-flash condition was always run last.

Predictions. We had no reason to think that the negative bias in the 0-offset condition of Experiment 1 was the result of an auditory–visual interaction. We therefore predicted that this bias would still be found when only the flash was presented. We also thought that centering a brief flash in the middle of the marker-dwell interval might reduce this bias. Finally, since reports of the marker position when the click occurred were not biased in the 0-offset condition of Experiment 2, we expected there would also be no bias when the click was presented by itself.

Results

In the flash-only condition, 10 of the 16 observers showed a negative bias in their judgments of marker position when the flash occurred. The mean size of this bias across observers was 0.23 marker positions, which is equivalent to a temporal offset of −11.5 msec. This bias is significantly different from 0 [$t(15) = 2.13, p < .05$], but not significantly different from the 20-msec bias obtained in the trials of Experiment 1 in which the sound and flash were temporally aligned [$t(28) = 1.04$]. As expected, there was no bias in judgments of the position of the marker in the click-only condition; the mean error for the 8 observers run was only −0.4 msec, which was not significantly different from 0 [$t(7) = .06$]. Finally, in the brief-flash condition, 11 of the 12 observers showed a negative bias, which has a mean value −24.8 msec [$t(11) = 4.17, p < .01$]. If only these 12 observers are considered, the mean bias in the flash-only condition is −9 msec, and the difference between the flash-only and brief-flash conditions is highly significant [$t(11) = 4.57, p < .001$].

GENERAL DISCUSSION

The present experiments demonstrate that when a flash and brief sound burst (e.g., a click) are temporally proximal, there can be cross-capture between these stimuli, drawing them toward temporal correspondence. In our data, the auditory capture of the flash by the click was more pronounced than the visual capture of the click by the flash. This contrasts with the normal dominance of vision that occurs when there is conflicting spatial visual and auditory information. The temporal capture effect we report seems likely to be related to the temporal capture that occurs in the case of auditory driving. However, whereas auditory driving represents a case of frequency capture, the effect we report might be better likened to a case of temporal phase capture. We note, though, that this characterization is really appropriate if the effect depends only on the presentation of repeating stimuli like those we used. The possibility that it might be observable with a single flash and a burst of sound remains to be evaluated. For generality, we will refer to the temporal cross-capture of vision and audition as *intersensory temporal locking* (ITL). Following Welch and Warren (1980), we will designate auditory biasing of vision as A(V) and visual biasing of audition as V(A). Informal phenomenal observations indicate that ITL, at least in the case of V(A), is genuinely perceptual. This was especially evident when we fixed the actual flash position at 12 o'clock during preliminary testing. Presenting the click at 11 o'clock made the marker *look* like it had not reached 12 when the flash occurred; presenting the click at 1 o'clock made the marker *look* like it had moved past 12 when the flash occurred. Just as visual capture appears to be a mechanism for resolving intermodal spatial discrepancies, ITL seems likely to be a mechanism for resolving intermodal temporal discrepancies. A recent report by Scheier, Nijhawan, and Shimojo (1999) suggests that a mechanism like ITL could also serve to sharpen the temporal boundaries of visual events.

Our paradigm permitted us to manipulate the direction and magnitude of the temporal offset between the auditory and visual stimuli. We found that in the case of both A(V) and V(A), with a 50-msec temporal offset capture, effects were equally effective in a forward and a backward temporal direction. To account for backward ITL (the ability of a stimulus to alter the apparent time of occurrence of a prior stimulus), one could posit that at the neural stage at which ITL is implemented, sensory representations are buffered and can be modified by subsequent inputs before becoming accessible to consciousness. This idea is consistent with Libet's (e.g., Libet, 1985; Libet, Wright, Feinstein, & Pearl, 1979) argument that sensory events lay down temporal markers that serve as references for determining when they are perceived as occurring. If we adopt this perspective, our data suggest that markers from different modalities may combine to establish a single temporal reference for a multisensory event.

In any case, the backward range of ITL appears to be more limited than its forward range. The forward acting capture effect grew linearly in magnitude for both A(V) and V(A) when we increased the temporal offset between the click and flash from 50 to 100 msec. The backward acting capture effect, on the other hand, remained essentially unchanged when we increased the offset from 50 to 100 msec. We think the simplest way to account for this is to assume that the backward range of ITL does not exceed 50 msec, so temporal offsets larger than 50 msec were simply not subject to the capture effect. Presumably, forward acting ITL would also cease to grow if a sufficiently large temporal offset was employed, and, with large enough temporal offsets, both forward and backward capture effects would diminish and disappear.

Even a 50-msec temporal offset may have been too large for maximal ITL to have occurred. We observed only partial capture effects. The summed effects of A(V) and V(A) were not large enough to have completely canceled the click–flash temporal offsets that we tested. We estimated the combined effect of A(V) and V(A) by adding the mean forward A(V) with the backward V(A) and vice versa, and taking the mean of these sums. With a 50-msec offset between the click and flash, the summed capture effects cancel 72% of this temporal discrepancy. With a 100-msec offset, 50% of the discrepancy is canceled. It is possible that ITL might completely cancel temporal offsets shorter than 50 msec. The systematic mapping of the build-up and drop-off of ITL as a function of the temporal offset between stimuli remains a matter for future investigation. Nevertheless, it does not appear that ITL, by itself, can completely account for the insensitivity of observers to auditory–visual asynchrony. Dixon and Spitz (1980) report that the detection of this asynchrony sometimes requires temporal offsets larger than 200 msec. However, the stimuli employed by these investigators were videos of a person speaking or of a hammer hitting a peg. With such stimuli, powerful top-down factors are likely to have reinforced the perceptual yoking of the visual and auditory stimuli, decreasing sensitivity to temporal mismatches.

In auditory driving paradigms, observers generally report that flutter rates remain phenomenally steady as concurrent flicker rates are varied. Using magnitude estimation, Welch et al. (1986) did find a small effect of flicker on flutter, but the effect of flutter on flicker was four times larger. These findings are in accord with the modality appropriateness hypothesis (Welch & Warren, 1980). In our ITL paradigm, on the other hand, auditory and visual capture effects were largely reciprocal. Both A(V) and V(A) were highly significant, and although A(V) was significantly larger than V(A), the actual difference between the two was relatively modest. This relative reciprocity between V(A) and A(V) raises the possibility that ITL and auditory driving are mediated at least in part by different mechanisms. However, there is an alternative way to account for the relatively strong effect of V(A) that we observed. It has been proposed that attentional factors might contribute to the resolution of intersensory conflicts (e.g., Canon, 1970; Kelso, Cook, Olson, & Epstein, 1975), and Welch and Warren (1980) have specifically argued that attention might mediate the effects of modality appropriateness. With our displays, irrespective of whether the observers were judging the time that the flash or click occurred, they were required to attend to the *visually* displayed revolving marker. This paradigmatic emphasis on visual attention could have weighted intersensory capture processes in favor of the visual modality. If this was in fact occurring, even the modest dominance of A(V) over V(A) that we found could indicate a substantial predisposition toward V(A) in the neural architecture mediating ITL.

Even when the click and flash were temporally aligned in Experiment 1, there was a negative bias in judgments of marker position when the flash occurred. Since a similar bias was observed in Experiment 3 when only the flash was presented, this bias does not appear to be the consequence of an auditory–visual interaction. No corresponding bias was found for judgments of the marker position when the click occurred, so it appears to be a specifically visual phenomenon. The finding is surprising since other investigators have found that a moving target is normally perceived as being *ahead* of a spatially aligned flashed target (e.g., Nijhawan, 1994; Whitney & Murakami, 1998). If this "flash-lag" effect were present in our displays, it would have produced a bias in the opposite direction to the one that we found.

We speculated that the negative bias might have been due to the fact that our marker was in apparent motion, stepping though a sequence of 12 discrete positions. Since the perceptual system interpreted the marker's motion as continuous, the marker presumably had to be seen as traversing some 30 degrees of arc while the flash was occurring. The onset of the flash was coincident with the marker's advance to the position where the flash occurred. If subjects tended to report the perceived position of the marker at the onset of the flash, that position would be biased backward by a half step from the marker's actual position. Because the click was briefer than the flash and was presented at the midpoint of the marker-dwell interval, it would not be subject to this backward bias. The brief-flash condition in Experiment 3 was designed to test this speculation. We thought that limiting the flash to the center of a marker interval might reduce the negative bias. In fact, the reverse occurred: The brief temporally centered flash was seen to have an increased negative bias. The source of the bias therefore remains uncertain. It does not, however, appear to bear directly on the intermodal capture effects that are our primary concern.

Several cortical regions in the primate brain respond to both visual and auditory inputs (e.g., Benvento, Fallon, Davis, & Rezak, 1977; Wollenberg & Sela, 1980), but the specific regions mediating intersensory capture are unknown. Evoked potential recording methods are well suited for detecting changes in temporal patterns of neural activity, so the neural correlates of ITL and auditory driving might be amenable to investigation with such methods. There has been only one attempt to investigate auditory driving electrophysiologically. Regan and Spekreijse (1977) found no change in the frequency of flash-induced potentials that corresponded to phenomenal auditory driving. However, these investigators recorded from a single electrode site close to the occipital pole (H. Spekreijse, personal communication, 1999), so responses from the early visual processing stages are likely to have dominated their records. Temporal capture effects might occur in polymodal areas functionally subsequent to these early sensory stages. Modern recording techniques that employ large electrode arrays might allow for the identification of processing stages at which temporal capture occurs. If successful, such investigations would reveal the cortical processing stages at which temporal patterns of activity first correspond to the contents of conscious perception.

REFERENCES

Benvento, L. A., Fallon, J., Davis, B. J., & Rezak, M. (1977). Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and orbital frontal cortex of the Macaque monkey. *Experimental Neurology*, **57**, 849-872.

Bermant, R. I., & Welch, R. B. (1976). The effect of degree of visual-auditory stimulus separation and eye position upon the spatial interaction of vision and audition. *Perceptual & Motor Skills*, **43**, 487-493.

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.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, **9**, 719-721.

Easton, R. D., & Moran, P. W. (1978). A quantitative confirmation of the visual capture of curvature. *Journal of General Psychology*, **98**, 105-112.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology*, **72**, 521-528.

Hay, J. C., Pick, H. L., Jr., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, **2**, 215-216.

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. New York: Wiley.

Kelso, J. A. S., Cook, E., Olson, M. E., & Epstein, W. (1975). Allocation of attention and the locus of adaptation to displaced vision. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 237-245.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral & Brain Sciences*, **8**, 529-566.

Libet, B., Wright, E. W., Feinstein, B., & Pearl, D. K. (1979). Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in man. *Brain*, **102**, 193-224.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, **370**, 256-257.

Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflicts in judgments of spatial direction. *Perception & Psychophysics*, **6**, 203-205.

Regan, D., & Spekreijse, H. (1977). Auditory-visual interactions and the correspondence between perceived auditory space and perceived visual space. *Perception*, **6**, 133-138.

Scheier, C. R., Nijhawan, R., & Shimojo, S. (1999). Sound alters visual temporal resolution [Abstract]. *Investigative Ophthalmology & Visual Science*, **40**, S792.

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

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

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

Whitney, D., & Murakami, I. (1998). Latency difference, not spatial extrapolation. *Nature Neuroscience*, **1**, 656-657.

Wollenberg, Z., & Sela, J. (1980). Frontal cortex of the awake squirrel monkey: Responses of single cells to visual and auditory stimuli. *Brain Research*, **198**, 216-220.

(Manuscript received April 13, 1999;
revision accepted for publication September 5, 2000.)