Auditory modulation of visual apparent motion with short spatial and temporal intervals

Hulusi Kafaligonul and
Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA

Gene R. Stoner
Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA

Abstract

Recently, E. Freeman and J. Driver (2008) reported a cross-modal temporal interaction in which brief sounds drive the perceived direction of visual apparent-motion, an effect they attributed to "temporal capture" of the visual stimuli by the sounds (S. Morein-Zamir, S. Soto-Faraco, & A. Kingstone, 2003). Freeman and Driver used "long-range" visual motion stimuli, which travel over long spatial and temporal intervals and engage high-order cortical areas (K. G. Claeys, D. T. Lindsey, E. De Schutter, & G. A. Orban, 2003; Y. Zhuo et al., 2003). We asked whether Freeman and Driver's temporal effects extended to the short-range apparent-motion stimuli that engage cortical area MT, a lower-order area with well-established spatiotemporal selectivity for visual motion (e.g. A. Mikami, 1991, 1992; A. Mikami, W. T. Newsome, & R. H. Wurtz, 1986a, 1986b; W. T. Newsome, A. Mikami, & R. H. Wurtz, 1986). Consistent with a temporal-capture account, we found that static sounds bias the perception of both the direction (Experiment 1) and the speed (Experiment 2) of short-range motion. Our results suggest that auditory timing may interact with visual spatiotemporal processing as early as cortical area MT. Examination of the neuronal responses of this well-studied area to the stimuli used in this study would provide a test and might provide insight into the neuronal representation of time.

Keywords

audio-visual interaction; temporal ventriloquism; motion processing; temporal processing; visual area MT

Introduction

The neuronal basis of time perception is a topic of much recent research (Buonomano & Karmarkar, 2002; Ivry & Schlerf, 2008; Mauk & Buonomano, 2004). A key question of that research is how temporal information from the different senses is combined (van Wassenhove, Buonomano, Shimojo, & Shams, 2008). Consistent with the fact that audition has better temporal resolution than vision, it has been found that auditory input can capture the perceived timing of visual stimuli. This phenomenon has been referred to as 'temporal ventriloquism' (Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003) due to its analogy to 'spatial ventriloquism' (Alais & Burr, 2004; Bertelson & Aschersleben, 1998; Howard & Templeton, 1966). A recent study by Freeman and Driver

Corresponding author: Hulusi Kafaligonul. hulusi.kafaligonul@gmail.com. Address: Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037-1099, USA.

suggests that temporal ventriloquism (or "temporal capture") can, in turn, bias the perceived direction of visual motion stimuli.

Freeman and Driver (2008) flashed bars sequentially to the left and right of ocular fixation, inducing a perception of visual motion. Motion perception favors smaller temporal intervals in these "apparent-motion" displays. For example, if the left–right (L–R) interval is smaller than the right–left (R–L) interval, rightward motion is perceptually dominant. In their key experiment, L–R and R–L intervals were identical but static "beeps" either lagged or led each flash. They found that the timing of the beeps determined perceived visual motion direction, even though the sounds themselves provided no motion information. Their findings agree with earlier reports that sounds can affect the quality of similarly configured apparent-motion stimuli (Getzmann, 2007; Staal & Donderi, 1983). Consistent with a temporal ventriloquism account, Freeman and Driver (2008) also found that static beeps can change the perceived interval between two flashes. The neuronal mechanisms underlying these perceptual effects are, however, uncertain (see Discussion).

Area MT is established as a key substrate of visual motion processing (for a review see Born & Bradley, 2005), but is not directionally tuned to apparent-motion stimuli with the spatial and temporal intervals (i.e. 14 deg and 300 ms, respectively) used by Freeman and Driver (Mikami, 1991, 1992; Mikami, Newsome, & Wurtz, 1986a, 1986b; Newsome, Mikami, & Wurtz, 1986). Functional imaging studies instead implicate higher-order cortical areas in the processing of such long-range motion (Claeys, Lindsey, De Schutter, & Orban, 2003; Zhuo et al., 2003). Thus while area MT's well-studied spatial and temporal tuning properties offers a good foundation for study of the mechanisms underlying temporal illusions, the stimuli devised by Freeman and Driver are not suitable for that area. There are, however, several reasons for doubting that sounds could similarly modulate the processing of the "short-range" apparent motion stimuli that engage area MT. First, long-range motion is, in general, sensitive to higher-order influences to which short-range motion is not. For example, long-range motion is sensitive to cues as to whether a given motion trajectory is biomechanically realizable (Shiffrar & Freyd, 1993). Second, neuroanatomical studies have found that convergence of sensory information increases as one ascends the cortical hierarchy (e.g. Jones & Powell, 1970) and hence auditory influences on visual motion might be restricted to the higher-order areas implicated in long-range motion.

In this study, we examined the ability of sounds to modulate the perception of short-range motion stimuli that engage area MT. We found that sounds can affect both the perceived direction (Experiment 1) and perceived speed (Experiments 2a–2c) of these stimuli. The stimuli used in these experiments thus provide a tool for investigation of area MT's role in the perception of time.

Methods

Subjects

Twelve human subjects participated in this study with ten being naïve to the purpose of the experiments. There were seven subjects in our direction discrimination (Experiment 1) and seven subjects in our primary speed discrimination experiment (Experiment 2a). Three subjects took part in both experiments. We also carried out two secondary speed discrimination experiments (Experiments 2b and 2c) and three subjects took part in these experiments. All participants had normal hearing and normal or corrected-to-normal visual acuity. Participants gave informed consent, and all procedures were in accordance with international standards (Declaration of Helsinki) and NIH guidelines.

Apparatus and stimuli

We used the CORTEX program (Laboratory of Neuropsychology, National Institute of Mental Health; http://www.cortex.salk.edu/) for stimulus presentation and data acquisition. Visual stimuli were presented on a 19″ CRT monitor (Sony Trinitron E500, 1024×768 pixel resolution and 100 Hz refresh rate) at a viewing distance of 57 cm. A PR701S photometer was used for luminance calibration and gamma correction of the display. Sounds were emitted by two speakers (ALTEC Lansing) positioned at the top of the visual display and amplitudes were measured by a sound-level meter. Consistent with Vroomen and Keetels' (2006) study of temporal ventriloquism, in pilot experiments we found that the relative positioning of auditory and visual stimuli was not critical. Accordingly, sound position was not varied in the experiments described here. Timing of visual and auditory stimuli was confirmed with a digital (Tektronix TDS 1002) oscilloscope connected to the computer soundcard and a photodiode (which detected visual stimulus onsets). Head movements were constrained by a chin rest.

Apparent-motion stimuli were adapted from Freeman and Driver (2008) except that the stimuli used in our study had spatial and temporal intervals that were similar to the range of values used in previous MT studies (Mikami, 1991, 1992). A small bright circle (0.2 deg diameter, 95.11 cd/m^2 luminance) at the center of the display served as a fixation marker. Visual stimuli consisted of "flashed" (70 ms) red bars (0.3×3 deg with a luminance of 3.71 cd/m^2) presented on a dark background (0.1 cd/m^2). As illustrated in Figure 1A, these bars were flashed up to the left ("FL") and up to the right ("FR") of fixation (bar centers were 3 deg above the fixation marker). Auditory stimuli were 10 ms beeps comprised of a rectangular windowed 480 Hz sine-wave carrier, sampled at 22 kHz with 8-bit quantization.

Procedure

Subjects sat in a dark room and fixated a bright circle at the center of the display. They were told that some visual stimuli would be accompanied by beeps but to base their responses solely on the visual stimuli. They started each trial by pressing a key after which experimental stimuli were presented.

Experiment 1: Influence of sound on perceived direction—Trials started with simultaneous presentation of left and right visual bars for a duration of 270 ms. An "event cycle" started after one bar (chosen randomly) was removed and is defined as the period from the solo onset of one bar to the reappearance of that bar (Figure 1A). We started trials with an onset of two simultaneously presented bars rather than an onset of one bar as the later sequence tends to produce a perceptual bias in the direction away from that initial onset regardless of temporal interval (Freeman & Driver, 2008). In the example shown in Figure 1A, the left bar (FL) first appears alone followed by the right bar (FR) followed by the return of FL. There were 11 such cycles per trial. As indicated by Figure 1B, there were three types of sound conditions: visual-only (no beeps), beeps in the left–right interval, beeps in the right–left interval. At the end of each trial, observers indicated, by pressing one of two keys, whether they perceived predominantly leftward or rightward visual motion (two-alternative forced choice).

For each trial, the interstimulus interval (ISI) between FL and FR (ISI_{LR}) was chosen pseudorandomly from nine values: 60, 90, 110, 130, 150, 170, 190, 210, 230, and 260 ms. The ISI_{RL} (i.e. ISI between FR and FL) covaried with ISI_{LR} and equaled 320 ms − ISI_{LR} (i.e. the sum of the two ISIs was held constant). Each solo bar presentation lasted 70 ms so that each event cycle lasted 460 ms (320 ms + 2 * 70 ms). The duration of the visual stimulus portion of every trial was thus ~5.3 seconds (simultaneous bar presentation of 270 ms + 11 event cycles of 460 ms). The ISI between flashes and beeps was held constant at 20

ms. In consequence, the ISI between the auditory stimuli ($ISI_{auditory}$) was always 60 ms less than the corresponding visual ISI (i.e. ISI_{RL} or ISI_{LR}).

A single experimental session consisted of 240 trials (3 sound conditions $\times$ 10 ISI conditions $\times$ 8 trials per condition). The spatial displacement (i.e., center-to-center separation between left and right bars) was kept constant during each experimental session and chosen from nine values: 0.2, 0.5, 0.76, 1.2, 1.8, 2.4, and 3.0 deg. Accordingly, for each subject, there were nine experimental sessions corresponding to nine different spatial displacements (SDs). The order of these sessions was randomized for each subject. Prior to these sessions, subjects were shown examples of rightward and leftward visual-only sequences after which they completed two or three practice sessions of 240 visual-only trials without any feedback. Table 1 provides a complete list of the stimulus parameters for this experiment.

Experiment 2a: Influence of sound on perceived speed—Each apparent-motion stimulus consisted of a pair of flashed bars (FL and FR) spatially separated by 0.76 deg (Figure 1C). The audiovisual "reference" stimulus had a fixed ISI between bars (ISI_{ref}) of 120 ms and a variable ISI between beeps ($ISI_{auditory}$): 20, 60, 100, 180, 260, 300, and 340 ms. The pair of beeps was always temporally centered with respect to the pair of flashed bars. The silent "test" stimulus had an ISI (ISI_{test}) that varied pseudorandomly from trial to trial: 40, 60, 80, 100, 120, 140, 160, 180, and 200 ms. Reference and test stimuli were separated by a delay of 500 ms (Figure 1C). At the end of each trial, observers indicated, by pressing one of the two keys, which apparent-motion stimuli appeared to move faster.

On a given trial, reference and test stimuli moved in the same motion direction (rightward or leftward), which varied pseudorandomly from trial to trial. Reference and test stimuli were not distinguished in the instructions to the subjects and their temporal order was randomized from trial to trial.

As indicated in Figure 1C, each experimental session had a balanced mixture of four conditions: visual-only (no beeps), $ISI_{auditory} < 120$ ms (beeps occurred between flashes), $ISI_{auditory} = 180$ ms (beeps centered on flashes), and $ISI_{auditory} > 260$ ms (flashes occurred between beeps). The $ISIs_{auditory}$ for the 3 audiovisual conditions were held constant within a session. More specifically, the audiovisual conditions (i.e. conditions with beeps) in each session had $ISIs_{auditory}$ drawn from one of three distributions with equal probability: 1) 20, 180, and 340; 2) 60, 180, and 300; 3) 100, 180, and 260. Accordingly, there were 288 trials (4 conditions $\times$ 9 ISI_{test} conditions $\times$ 8 trials per condition) in each experimental session. Each subject participated in nine such sessions (3 runs $\times$ 3 $ISI_{auditory}$ distributions). The order of these sessions was randomized. Prior to experimental sessions, each subject was shown examples of these visual apparent-motion stimuli (without sounds) followed by two or three practice sessions of 288 visual-only trials without any feedback (see Table 2 for a list of stimulus parameters).

Experiment 2b: Shorter temporal interval—In a separate set of trials, we examined the ability of sound to affect the perception of speed of apparent-motion stimuli with smaller ISIs. Specifically, the audiovisual reference stimuli had an ISI_{ref} of 60 ms and the ISIs between beeps ($ISI_{auditory}$) in audiovisual conditions were: 20, 80 and 140 ms. As in Experiment 2a, we also had one visual-only condition. For both audiovisual and visual-only conditions, the ISI of the test stimulus (ISI_{test}) varied pseudorandomly from trial to trial: 10, 30, 50, 70, 90, and 110 ms. Flash (FL and FR) duration was 30 ms (compared to 70 ms in 2a) and flash luminance was 16.19 cd/m^2. These parameters were chosen to engage direction-selective neurons with sensitivity for shorter temporal intervals than used in Experiment 2a. The visual-only and the three audiovisual conditions were run in the same experimental session. There were 288 trials per session (4 sound conditions $\times$ 6 ISI_{test}

conditions × 12 trials per condition). Each subject completed four sessions. The procedure and other stimulus parameters were the same as those in Experiment 2a (see Table 2).

Experiment 2c: Perceived speed by changing spatial displacement—In Experiments 2a and 2b, perceived speed was measured by varying the ISI between the two flashes (FL and FR) in the silent test stimulus. To confirm that subjects were following our instructions and basing their judgments on speed rather than on ISI, in this experiment we measured perceived speed by varying the spatial displacement between FL and FR. The parameters of the audiovisual reference stimulus were the same as those in Experiment 2b except that the spatial displacement between FL and FR was 1.0 deg. The silent test stimulus had a fixed ISI_{test} value of 60 ms. The spatial displacement between FL and FR (SD_{test}) was varied pseudorandomly from trial to trial: 0.30, 0.60, 0.80, 1.20, 1.40 and 1.70 deg. These values yielded different speed values for the test stimulus: 3.33, 6.66, 8.89, 13.33, 15.56 and 18.89 deg/sec. There were 288 trials (4 sound conditions × 6 SD_{test} × 12 trials per condition). Other stimulus parameters and procedure were the same as those in Experiment 2b. Table 2 provides a complete list of the stimulus parameters of Experiments 2a, 2b, and 2c.

Data analysis

For each condition, individual and group-averaged data were fitted by a Complementary Error Function (1 − Cumulative Gaussian). The 50% point on the resultant curves yields the point of subjective equality (PSE). For Experiment 1, the PSE is the ISI_{LR} (equivalently 320-ISI_{RL}) for which leftward and rightward reports are equiprobable. For Experiments 2a and 2b, the PSE is the ISI_{test} for which the test was seen as faster than the reference on 50% of the trials. For Experiment 2c, the PSE is the SD_{test} for which the test was seen as faster than the reference on 50% of the trials. In all experiments, we looked for sound-induced changes in PSE values. As described in Results, we applied repeated-measures ANOVA test for statistical analysis. When the sphericity assumption was not met, the Greenhouse–Geisser correction was applied and the epsilon values are indicated in the results.

Results

Experiment 1

As Freeman and Driver found for long-range apparent motion perception, short-range motion perception was, in the absence of sound, dominated by the motion sequence with the physically shorter ISI: When the ISI_{LR} was smaller than ISI_{RL}, leftward motion reports were more likely than rightward reports and when ISI_{RL} was smaller than ISI_{LR}, the reverse was true. For large interval asymmetries, subjects report seeing motion in one direction after which the bar simple reappears at (rather than moving to) its initial position after the longer interval. For smaller interval asymmetries, motion is seen in both directions but one direction is more salient. Figure 2A shows this effect for 0.76 spatial displacements. If the temporal-capture effects reported by Freeman and Driver extend to short-range motion, then inserting beeps between the visual flashes should result in a reduction in the perceived ISI between the flashes and hence an increase in the perceptual dominance of one direction of motion. For example, when beeps occur in the L–R interval, the perceived ISI_{LR} should be less than the actual ISI_{LR} so that the saliency of rightward motion increases. Accordingly, we predicted that beeps in the R–L interval should result in a rightward shift (relative to the visual-only condition) in the psychometric curve whereas beeps in the L–R interval should result in a leftward shift in the psychometric curve.

As shown in Figure 2A, our results support these predictions: relative to the visual-only condition, the psychometric curve for beeps in the L–R interval condition was shifted

rightward (consistent with a bias in favor of rightward motion) and the psychometric curve for beeps in the R–L interval condition was shifted leftward (consistent with a bias in favor of leftward motion).

The results of Experiment 1 extend the findings of Freeman and Driver (2008) in several ways. First and foremost, we have found that the ability of sounds to influence motion perception extends to stimuli that engage area MT. Secondly, we have shown that this cross-modal interaction is not restricted to motion sequences with ambiguous timing (i.e. when ISI_{LR} equals to ISI_{RL}) but extends to a range of ISI_{LR} and ISI_{RL} values. Our use of a range of ISIs allowed us to estimate the changes in perceived timing that would be necessary to induce the shifts in directional judgments. These changes in PSE values (see above) are shown in Figure 2B for all spatial displacements (i.e. 0.2 to 3.0 deg).

The dependency of the PSE on sound and spatial interval was analyzed with a two-way repeated-measures ANOVA with sound condition (visual-only, beeps in the L–R interval, beeps in the R–L interval) and spatial interval between flashed bars (0.2, 0.5, 0.76, 1.2, 1.8, 2.4, and 3.0) as factors. We found a significant effect of sound condition [$F(2, 12) = 38.093$, $p = 0.001$, $\varepsilon = 0.542$] but no significant effect of spatial interval [$F(6, 36) = .716$, $p = 0.543$, $\varepsilon = 0.449$]. There was also no significant interaction between sound condition and spatial interval [$F(12, 72) = 0.818$, $p = 0.486$, $\varepsilon = 0.212$]. Consistent with these statistics, the changes in PSE induced by sound (i.e. relative to the visual-only condition) were nearly constant as a function of spatial displacement with values between 40 and 50 ms.

In Experiment 1, we inferred the perceived ISI (i.e. the PSE) based on our subject's reports of whether leftward or rightward motion was dominant. While the perceptual salience of apparent-motion stimuli is known to be dependent upon ISI, other factors can affect visual motion salience and hence other explanations might be advanced for our results. For example, motion intervals with beeps might have been more salient because the beeps drew attention to that interval. The design of Experiment 2 avoided that concern and provided a more direct test of the temporal-capture account.

Experiment 2

If the beeps do indeed change the perceived timing of the flashes, then when the $ISI_{auditory}$ is smaller than the ISI_{ref}, the ISI_{ref} should appear to contract leading to an increase in the perceived speed of the reference stimulus. Conversely, when the $ISI_{auditory}$ is larger than the ISI_{ref}, the ISI_{ref} should appear to lengthen leading to a decrease in the perceived speed of the reference stimulus. On the other hand, if the beeps induced changes in salience but not duration, we would not expect the timing of beeps to elicit a systematic change in perceived speed.

Figure 3A shows our results for experimental sessions that included the smallest (20 ms) and largest (340 ms) $ISI_{auditory}$ conditions used in the experiment (see Methods). These results support the temporal-capture account. For example, for trials in which the ISI_{test} and the ISI_{ref} were physically identical at 120 ms, an $ISI_{auditory}$ of 20 ms lead subjects to judge the reference stimulus to be faster than the test stimulus. Conversely, an $ISI_{auditory}$ of 340 ms lead subjects to judge the reference stimulus to be slower than the test stimulus. A decrease in the perceived speed of the reference stimulus was seen for the $ISI_{auditory} = 180$ condition, consistent with perceptual expansion of the actual ISI_{ref} (i.e. 120 ms).

Analysis of these data over the full range of ISI_{test} values yields the PSE, which is an estimate of the sound-induced perceptual expansion or contraction of ISI_{ref}. Figure 3B shows the PSEs for all $ISI_{auditory}$ values. As can be seen, we found a marked dependence of the PSE on the $ISI_{auditory}$ (one-way repeated measures ANOVA: $F(6, 36) = 25.197$, $p =$

0.001, $\varepsilon = 0.210$). This dependency had limits however as the shifts in the PSE never exceeded 50 ms. As shown in Figure 4, in Experiment 2b we found that the PSE's dependency on beep timing extends to a smaller visual ISI ($ISI_{ref} = 60$ ms) and flash duration (30 ms): the visual PSE was very significantly dependent on the $ISI_{auditory}$ (one-way repeated measures ANOVA: $F(2, 4) = 25.847$, $p = 0.005$).

In these speed discrimination experiments, subjects were instructed to compare the speed of the two apparent-motion stimuli. Nevertheless, their judgments might have been influenced by the perceived temporal interval rather than speed *per se*. To rule out this possibility, we carried out an additional experiment (Experiment 2c) in which the speed of the silent test stimulus was varied by varying the spatial displacement rather than the ISI (which was fixed at 60 ms). As was found in Experiments 2a and 2b, perceived speed was highly dependent on the $ISI_{auditory}$ (one-way repeated measures ANOVA: $F(2, 4) = 23.171$, $p = 0.006$) with larger $ISI_{auditory}$ values yielding smaller speed PSEs (Figure 5).

In their study of long-range apparent motion, Freeman and Driver (2008) provided evidence that visual intervals with intervening beeps are perceived to be perceptually shorter than intervals without beeps (Supplementary Methods). Our findings extend their results by: 1) demonstrating temporal expansion in addition to temporal contraction, 2) confirming that temporal capture elicits corresponding changes in perceived speed, and 3) demonstrating that temporal capture applies to short-range motion stimuli known to activate area MT.

Discussion

Freeman and Driver (2008) reported that static sounds could drive the perceived direction of long-range apparent-motion stimuli, an effect they attributed to temporal capture of the visual stimuli by the sounds. Their findings imply that cross-modal effects can occur in the higher-order visual areas implicated in the processing of long-range apparent motion. We have discovered that static sounds can also drive the perceived direction and speed of short-range motion stimuli that engage area MT. We discuss the relationship of our new findings to previous studies of both temporal and cross-modal processing. We end with a discussion of what our findings suggest about the future study of the neuronal basis of temporal processing.

Different types of visual motion

It is well established that there is not a single substrate for visual motion processing. Anstis (1980) and Braddick (1974, 1980) offered a distinction between short- and long-range motion, specialized for short and long spatio-temporal intervals respectively. Our use of these terms in this study is not meant, however, to precisely mesh with previous distinctions, nor do we mean to embrace a two-process visual motion scheme. Instead, we are simply distinguishing between apparent-motion stimuli that engage area MT and those that do not. Our findings in conjunction with those of Freeman and Driver (2008) demonstrate that static sounds can modulate the perception of the former class of motion as well as the later. Apparent motion stimuli with parameters like those used in this study have been shown to engage direction-selective neurons in area MT of the macaque (Mikami, 1991, 1992). However, the larger spatial displacements may have been beyond the optimal range of area MT neurons and accordingly may have preferentially activated other, higher-order, cortical areas with direction-selective neurons. Moreover, direction-selective neurons in other cortical areas undoubtedly also responded to the stimuli that did engage area MT. Our study thus does not identify which cortical areas are involved in the cross-modal illusion documented here but does provide new tools for identifying those areas.

Auditory influences within the visual processing hierarchy

Early neuroanatomical and functional studies found little evidence of cross-modal interactions at low-level stages of sensory processing, suggesting that sensory convergence in the cortical hierarchy occurs in higher-order "association areas" (for reviews see Driver & Noesselt, 2008; Kayser & Logothetis, 2007). From that viewpoint, Freeman and Driver's finding that sounds can bias long-range motion is perhaps not that surprising since the areas implicated in that type of motion are within higher-order (and potentially polysensory) cortex (Claeys et al., 2003; Zhuo et al., 2003). We have found, however, that static sounds also affect the processing of short-range motion for which the presumed neuronal substrate is area MT. Area MT is low in the visual processing hierarchy (indeed it receives direct input from the LGN, Sincich, Park, Wohlgemuth, & Horton, 2004) and has traditionally been thought of as a purely visual area. Our findings suggest the possibility that cross-modal temporal interactions might occur as early as MT.

Although the idea that audio-visual interactions occur early within the visual hierarchy conflicts with the traditional view of sensory convergence being restricted to higher-order cortex, it is consistent with some of the findings from more recent studies of cross-modal interactions. For example, recent neuroanatomical studies have provided evidence of sparse connections between primary auditory and visual cortices (Cappe & Barone, 2005; Clavagnier, Falchier, & Kennedy, 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002). Of more direct relevance to our current study, several functional imaging studies have found auditory influences on area MT (e.g. Alink, Singer, & Muckli, 2008; Calvert et al., 1999; Ciaramitaro, Buracas, & Boynton, 2007; Scheef et al., 2009).

Area MT and temporal processing

As indicated above, there is accumulating evidence of auditory modulation of area MT. While there is yet no direct evidence that area MT is involved in the audiovisual temporal illusion documented here, there is evidence that area MT plays a more general role in the temporal processing of visual stimuli. It is well-established that area MT precisely encodes the fine temporal structure of visual stimuli (Bair & Koch, 1996; Buracas, Zador, DeWeese, & Albright, 1998) and thus could support the perception of visual timing. Moreover, area MT has been found to be activated when subjects engage in visual timing tasks such as the timing of rhythmic visual stimuli (Jantzen, Steinberg, & Kelso, 2005) and estimating the time of visual interception (Bosco, Carrozzo, & Lacquaniti, 2008). While the above evidence is suggestive, Bueti, Bahrami, and Walsh (2008) have provided direct evidence that area MT plays a role in temporal perception. In particular, Bueti et al. (2008) found that transcranial magnetic stimulus (TMS) targeted at area MT impaired the ability of human subjects to discriminate short temporal intervals (i.e. of hundreds of milliseconds). Conversely, the discrimination of longer temporal intervals likely involves higher-order cortical areas (for a review see Battelli, Walsh, Pascual-Leone, & Cavanagh, 2008).

Bueti et al.'s findings are generally consistent with psychophysical evidence that the perceived timing of visual events relies on visually responsive neurons with relatively small receptive fields (Ayhan, Bruno, Nishida, & Johnston, 2009; Burr, Tozzi, & Morrone, 2007; Johnston, Arnold, & Nishida, 2006). These studies have demonstrated that the perception of duration can be altered by prior adaptation to dynamic stimuli in a spatially specific manner. Moreover, Kaneko and Murakami (2009) report that the perception of duration depends on mechanisms tuned to speed (rather than temporal frequency). Since neurons tuned to speed are rare in area V1 but are reported in area MT (Perrone & Thiele, 2001; Priebe, Lisberger, & Movshon, 2006), Kaneko and Murakami conclude that this change in perceived duration is primarily mediated by higher level motion areas (such as area MT) in the dorsal pathway. Taken together, the evidence thus suggests that area MT may play a role in the purely

temporal processing of visual events in addition to its well-established role in spatiotemporal (i.e. motion) processing.

Conclusions

Given the evidence reviewed above, we suggest that single-unit examination of area MT might provide useful insight into the neuronal mechanisms underlying the perception of short temporal intervals. Specifically, we suggest that the cross-modal temporal illusion demonstrated in Experiment 2 could be used to examine the underlying neuronal mechanisms. These mechanisms might take various forms, which might only be distinguished at the level of the single neuron. To illustrate, the observed temporal capture illusion could result from shifts in the onset and/or offset of the responses to the individual flashes. This would suggest that temporal capture occurs prior to MT and operates on the timing of the events themselves. This scenario would account for the ability of sounds to affect the processing of both short- and long-range apparent-motion as the cross-modal interaction would occur prior to the computation of both. Alternatively, temporal capture could operate on the interval independent of response timing and hence be manifest by changes in response magnitude without changes in the dynamics of those responses. Such a result would be consistent with evidence suggesting that perceived duration can be altered without any apparent change in the perceived timing of the onset and offset of the stimuli defining that interval (Kaneko & Murakami, 2009). In conclusion, our findings, in conjunction with a variety of related converging findings, suggest that the well-documented temporal tuning of area MT (Mikami, 1991, 1992; Mikami et al., 1986a, 1986b; Newsome et al., 1986) may offer a solid foundation upon which to investigate the neuronal basis of the perceived timing of visual events.

Acknowledgments

References

Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. Current Biology. 2004; 14:257–262. [PubMed: 14761661]

Alink A, Singer W, Muckli L. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. Journal of Neuroscience. 2008; 28:2690–2697. [PubMed: 18337398]

Anstis SM. The perception of apparent movement. Philosophical Transactions of Royal Society London B: Biological Sciences. 1980; 290:153–168.

Ayhan, I.; Bruno, A.; Nishida, S.; Johnston, A. The spatial tuning of adaptation-based time compression; Journal of Vision. 2009. p. 2p. 1-12.http://www.journalofvision.org/content/9/11/2

Bair W, Koch C. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Computation. 1996; 8:1185–1202. [PubMed: 8768391]

Battelli L, Walsh V, Pascual-Leone A, Cavanagh P. The 'when' parietal pathway explored by lesion studies. Current Opinion in Neurobiology. 2008; 18:120–126. [PubMed: 18708141]

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

Born RT, Bradley DC. Structure and function of visual area MT. Annual Review of Neuroscience. 2005; 28:157–189.

Bosco G, Carrozzo M, Lacquaniti F. Contributions of the human temporoparietal junction and MT/V5+ to the timing of interception revealed by transcranial magnetic stimulation. Journal of Neuroscience. 2008; 28:12071–12084. [PubMed: 19005072]

Braddick O. A short-range process in apparent motion. Vision Research. 1974; 14:519–527. [PubMed: 4423193]

Braddick OJ. Low-level and high level processes in apparent motion. Philosophical Transactions of Royal Society London B: Biological Sciences. 1980; 290:137–151.

Bueti D, Bahrami B, Walsh V. Sensory and association cortex in time perception. Journal of Cognitive Neuroscience. 2008; 20:1054–1062. [PubMed: 18211231]

Buonomano DV, Karmarkar UR. How do we tell time? Neuroscientist. 2002; 8:42–51. [PubMed: 11843098]

Buracas GT, Zador AM, DeWeese MR, Albright TD. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. Neuron. 1998; 20:959–969. [PubMed: 9620700]

Burr D, Tozzi A, Morrone MC. Neural mechanisms for timing visual events are spatially selective in real-world coordinates. Nature Neuroscience. 2007; 10:423–425.

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS. Response amplification in sensory-specific cortices during cross-modal binding. Neuroreport. 1999; 10:2619–2623. [PubMed: 10574380]

Cappe C, Barone P. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. European Journal of Neuroscience. 2005; 22:2886–2902. [PubMed: 16324124]

Ciaramitaro VM, Buracas GT, Boynton GM. Spatial and cross-modal attention alter responses to unattended sensory information in early visual and auditory human cortex. Journal of Neurophysiology. 2007; 98:2399–2413. [PubMed: 17715196]

Claeys KG, Lindsey DT, De Schutter E, Orban GA. A higher order motion region in human inferior parietal lobule: Evidence from fMRI. Neuron. 2003; 40:451–452. [PubMed: 14642268]

Clavagnier S, Falchier A, Kennedy H. Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. Cognitive, Affective and Behavioral Neuroscience. 2004; 4:117–126.

Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on sensory-specific-brain regions, neural responses, and judgments. Neuron. 2008; 57:11–23. [PubMed: 18184561]

Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical evidence of multimodal integration in primate striate cortex. Journal of Neuroscience. 2002; 22:5749–5759. [PubMed: 12097528]

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

Freeman E, Driver J. Direction of visual apparent motion driven solely by timing of a static sound. Current Biology. 2008; 18:1262–1266. [PubMed: 18718760]

Getzmann S. The effect of brief auditory stimuli on visual apparent motion. Perception. 2007; 36:1089–1103. [PubMed: 17844974]

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

Ivry RB, Schlerf JE. Dedicated and intrinsic models of time perception. Trends in Cognitive Sciences. 2008; 12:273–80. [PubMed: 18539519]

Jantzen K, Steinberg F, Kelso J. Functional MRI reveals the existence of modality and coordination-dependent timing networks. Neuroimage. 2005; 25:1031–1042. [PubMed: 15850722]

Johnston A, Arnold DH, Nishida S. Spatially localized distortions of event time. Current Biology. 2006; 16:472–479. [PubMed: 16527741]

Jones EG, Powell TP. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain. 1970; 93:793–820. [PubMed: 4992433]

Kaneko, S.; Murakami, I. Perceived duration of visual motion increases with speed; Journal of Vision. 2009. p. 14p. 1-12.http://www.journalofvision.org/content/9/7/14

Kayser C, Logothetis NK. Do early sensory cortices integrate cross-modal information? Brain Structure and Function. 2007; 212:121–132. [PubMed: 17717687]

Mauk MD, Buonomano DV. The neural basis of temporal processing. Annual Review of Neuroscience. 2004; 27:307–340.

Mikami A. Direction selective neurons respond to short-range and long-range apparent motion stimuli in macaque visual area MT. International Journal of Neuroscience. 1991; 61:101–112. [PubMed: 1809728]

Mikami A. Spatiotemporal characteristics of direction-selective neurons in the middle temporal visual area of the macaque monkeys. Experimental Brain Research. 1992; 90:40–46.

Mikami A, Newsome WT, Wurtz RH. Motion selectivity in macaque visual cortex: I. Mechanisms of direction and speed selectivity in extrastriate area MT. Journal of Neurophysiology. 1986a; 55:1308–1327. [PubMed: 3016210]

Mikami A, Newsome WT, Wurtz RH. Motion selectivity in macaque visual cortex: II. Spatiotemporal range of directional interactions in MT and V1. Journal of Neurophysiology. 1986b; 55:1328–1339. [PubMed: 3734858]

Morein-Zamir S, Soto-Faraco S, Kingstone A. Auditory capture of vision: Examining temporal ventriloquism. Cognitive Brain Research. 2003; 17:154–163. [PubMed: 12763201]

Newsome WT, Mikami A, Wurtz RH. Motion selectivity in macaque visual cortex: III. Psychophysics and physiology of apparent motion. Journal of Neurophysiology. 1986; 55:1340–1351. [PubMed: 3734859]

Perrone JA, Thiele A. Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. Nature Neuroscience. 2001; 4:526–532.

Priebe NJ, Lisberger SG, Movshon JA. Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. Journal of Neuroscience. 2006; 26:2941–2950. [PubMed: 16540571]

Scheef L, Boecker H, Daamen M, Fehse U, Landsberg MW, Granath DO, et al. Multimodal motion processing in area V5/MT: Evidence from an artificial class of audio visual events. Brain Research. 2009; 1252:94–104. [PubMed: 19083992]

Shiffrar M, Freyd J. Timing and apparent motion path choice with human body photographs. Psychological Science. 1993; 4:379–384.

Sincich LC, Park KF, Wohlgemuth MJ, Horton JC. Bypassing V1: A direct geniculate input to area MT. Nature Neuroscience. 2004; 7:1123–1128.

Staal HE, Donderi DC. The effect of sound on visual apparent movement. American Journal of Psychology. 1983; 96:95–105. [PubMed: 6859347]

van Wassenhove V, Buonomano DV, Shimojo S, Shams L. Distortions of subjective time perception within and across senses. PLoS One. 2008; 3:e1437. [PubMed: 18197248]

Vroomen J, Keetels M. The spatial constraint in intersensory pairing: No role in temporal ventriloquism. Journal of Experimental Psychology: Human Perception and Performance. 2006; 32:1063–1071. [PubMed: 16846297]

Zhuo Y, Zhou TG, Rao HY, Wang JJ, Meng M, Chen M, et al. Contributions of the visual ventral pathway to long-range apparent motion. Science. 2003; 299:417–420. [PubMed: 12532023]
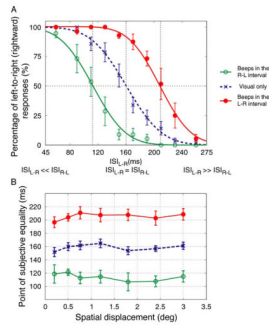
Figure 1.
(A) Stimulus configuration. Red bars were first flashed simultaneously to upper-left (FL) and right (FR) of a central fixation circle, then either FL or FR was extinguished after which bars were flashed sequentially. Auditory beeps were emitted by speakers positioned above display. (B) Timing diagram for event cycle in Experiment 1. In this example, FL was followed by FR with a delay of ISI_{LR}. The delay between FR and the subsequent FL was ISI_{RL}. ISI_{LR} plus ISI_{RL} was held constant at 320 ms. The visual-only condition had no sounds. For audiovisual conditions, beeps were introduced in either the LR or RL interval. The ISI between individual beeps and flashes was 20 ms. (C) Timing diagram for Experiment 2. Reference and test stimuli consisted of two visual flashes. Beeps were only introduced during the reference stimulus. Each experimental session consisted of three audiovisual conditions and one visual-only condition. In B and C, timing of bars in audiovisual trials is indicated by dashed lines aligned with bar presentation in visual-only condition.

Figure 2.
Results of Experiment 1. (A) Group averaged data (N = 7, five were naïve observers) for 0.76 deg spatial displacement. The plot indicates proportion of trials in which the direction of visual motion was judged to be rightward as a function of ISI_{LR}. In both *A* and *B* of this figure, filled and open symbols represent beeps in the LR and RL intervals, respectively. Error bars correspond to ± *SEM*. In all cases, the data were well fit by complementary error functions and the intersections of the 50% point with the vertical lines gave an estimate of PSEs. (B) The averaged PSE values across observers as a function of spatial displacement. The dashed line represents average PSE values for the visual-only condition. Error bars correspond to ±*SEM*.
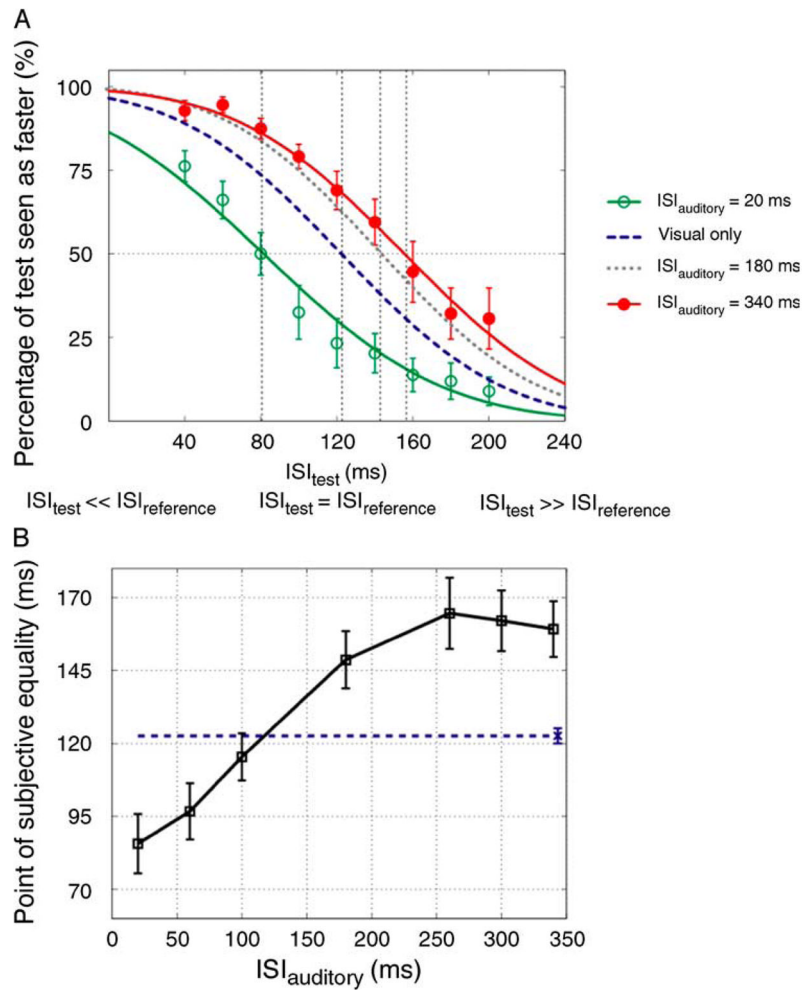
Figure 3.
Results of Experiment 2a. (A) Group averaged data ($N = 7$, six were naïve observers) for different $ISI_{auditory}$ conditions. The plot indicates the proportion of trials in which the test stimulus was judged to move faster than the reference stimulus as a function of ISI_{test}. The open and filled symbols represent 20 and 340 ms $ISI_{auditory}$ conditions, respectively. Error bars indicate $\pm SEM$. The error bars for the visual-only and 180 ms $ISI_{auditory}$ conditions were similar to those in the other conditions and were omitted to avoid clutter. The dotted and dashed curves indicate visual-only and 180 ms $ISI_{auditory}$ psychometric fits, respectively. The intersection of the 50% point with the vertical line gave an estimate of the PSE for each condition. (B) The averaged PSEs of all observers as a function of $ISI_{auditory}$. Error bars correspond to $\pm SEM$. The dashed line indicates the PSE for the visual-only condition and the error-bar placed over the symbol at the end of this line represents $\pm SEM$.
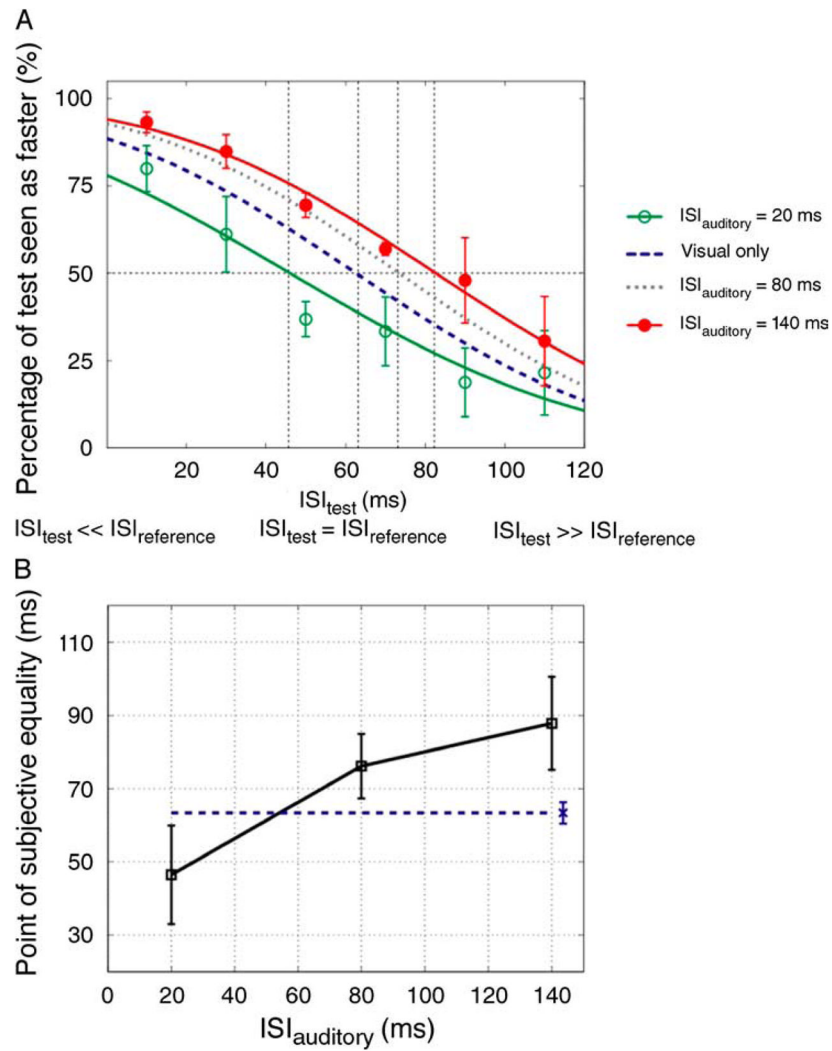
Figure 4.
Results of Experiment 2b. (A) Group averaged data (N = 3, two were naïve observers) for different $ISI_{auditory}$ conditions. The open and filled symbols represent 20 and 140 ms $ISI_{auditory}$ conditions, respectively. The dotted and dashed curves indicate visual-only and 80 ms $ISI_{auditory}$ psychometric fits, respectively. (B) The averaged PSEs of all observers as a function of $ISI_{auditory}$. Other conventions are the same as those in Figure 3.
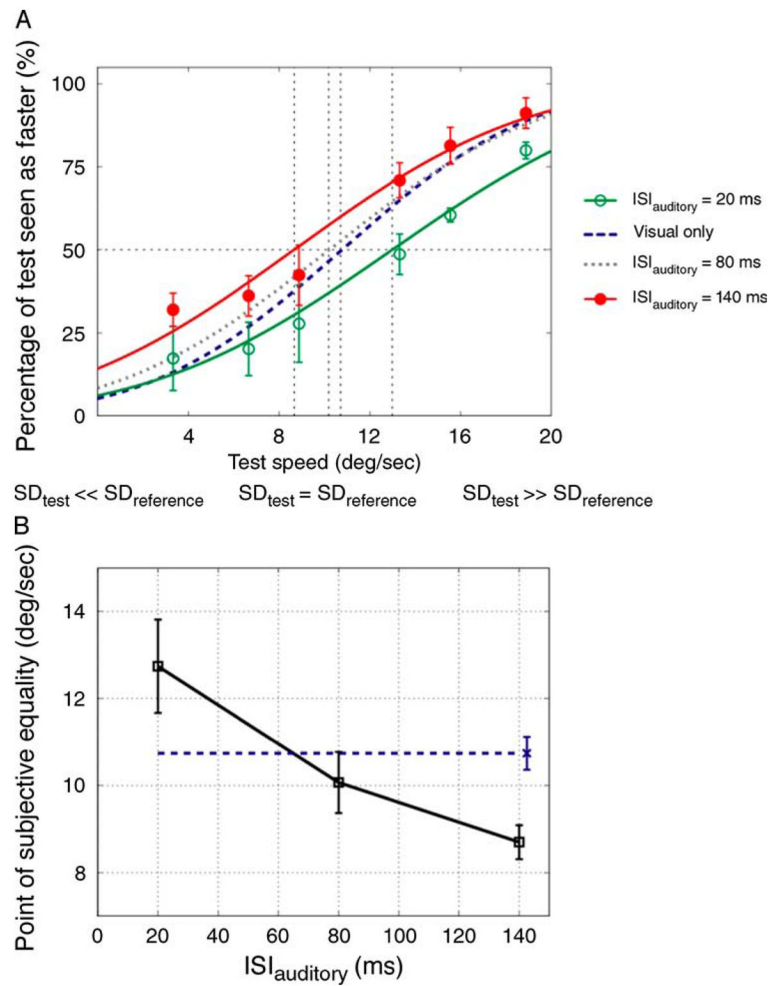
Figure 5.
Results of Experiment 2c. (A) Group averaged data ($N = 3$, two were naïve observers) for different $ISI_{auditory}$ conditions. The plot indicates the proportion of trials in which the test stimulus was judged to move faster than the reference stimulus as a function of test speed (deg/sec). The open and filled symbols represent 20 and 140 ms $ISI_{auditory}$ conditions, respectively. The dotted and dashed curves indicate visual-only and 80 ms $ISI_{auditory}$ psychometric fits, respectively. The error bars for the visual-only and 80 ms $ISI_{auditory}$ conditions were similar to those in the other conditions and were omitted to avoid clutter. (B) The averaged PSEs of all observers as a function of $ISI_{auditory}$. Other conventions are the same as those in Figure 4.

Table 1

Stimulus parameters for Experiment 1.

	Visual Parameters		Auditory Parameters
	SD (deg)	ISI_{LR} (ms)	Beep Timing
Experiment 1	0.2, 0.5, 0.76, 1.0, 1.2, 1.8, 2.4, 3.0	60, 90, 110, 130, 150, 170, 190, 210, 230, 260	beeps in the LR interval beeps in the RL interval silent[*]

Note:

[*] Silent (visual-only) conditions had no beeps.

Table 2

Stimulus parameters for Experiments 2a, 2b and 2c.

		Visual Parameters		Auditory Parameters
		SD (deg)	ISI (ms)	ISI (ms)
Experiment 2a	Reference	0.76	120	20, 60, 100, 180, 260, 300, 340 silent[*]
	Test	0.76	40, 60, 80, 100, 120, 140, 160, 180, 200	silent[*]
Experiment 2b	Reference	0.76	60	20, 80, 140 silent[*]
	Test	0.76	10, 30, 50, 70, 90, 110	silent[*]
Experiment 2c	Reference	1	60	20, 80, 140 silent[*]
	Test	0.30, 0.60, 0.80, 1.20, 1.40, 1.70	60	silent[*]

Note:

[*] Visual-only reference condition had no beeps.