ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 44 (2004) 1907–1917

Vision Research

www.elsevier.com/locate/visres

Sound-aided recovery from and persistence against visual filling-in

Bhavin R. Sheth [a,*], Shinsuke Shimojo [a,b]

[a] *Computation and Neural Systems, California Institute of Technology, 139-74, Caltech, Pasadena, CA 91125, USA*
[b] *Human and Information Science Laboratory, NTT Communication Science Laboratories, Atsugi, Kanagawa 243-0198, Japan*

Received 6 May 2003; received in revised form 26 February 2004

Abstract

Disappearance phenomena, in which salient visual stimuli do not register consciously, have been known to occur. Recovery from such phenomena typically occurs through change in some visual attribute, such as increase in luminance contrast or stimulus duration. Thus far, there have been no reports of cross-modal modulation of disappearance phenomena. In particular, what effect a cross-modal attentional cue has on sensory suppression is unknown. Here, we show that an adapted, flickered visual target that is synchronous with a brief sound appears more vivid than a similarly adapted, otherwise identical, visual target that is offset in time by more than 200 ms from the auditory cue. We argue that the brief auditory stimuli momentarily boost the concurrent signal of the adapted visual stimulus at a site downstream of the visual adaptation, thus causing the transient recovery from the visual adaptation. Repetitive visual cues cause significantly less recovery from visual adaptation than repetitive auditory cues, implying that there are functions a cross-modal cue can perform that a cue of the same modality cannot. Moreover repetitive auditory cues selectively prevent synchronous visual targets from undergoing visual adaptation. Ours is the first report of cross-modal modulation of a disappearance phenomenon.

Keywords: Multisensory; Adaptation; Disappearance phenomena; Attention; Cross-modal

1. Introduction

Disappearance phenomena are phenomena in which information is ignored owing not from a failure to activate retinal photoreceptors, but from explicit ‘erasing’ or lack of conscious encoding later in the visual system. There are several examples of disappearance phenomena, including binocular (Blake, 1989) and monocular (Andrews & Purves, 1997; Campbell, Glinsky, Howel, Riggs, & Atkinson, 1973) rivalry, motion-induced blindness (Bonneh, Cooperman, & Sagi, 2001) and after-images that disappear and reappear (Bennet-Clark & Evans, 1963). In these phenomena, the stimulus elicits a reliable response at the earliest sensory stages, but is then suppressed at some higher level in the brain prior to reaching visual awareness. A physically sub-threshold stimulus is not consciously registered either, but, unlike disappearance phenomena, it fails to elicit a reliable response even at the earliest sensory stages (including, perhaps, retinal photoreceptors). Modulation of disappearance phenomena occurs at multiple levels of the nervous system, and sometimes it takes the form of a high-level change of a visual attribute. For instance, the abrupt appearance of a stimulus in one eye causes the stimulus presented in the other eye to disappear from consciousness (Wolfe, 1986), or the visual grouping of target items changes the dynamics of motion-induced blindness (Bonneh et al., 2001). An even clearer demonstration of a difference in the brain’s response to an above-threshold stimulus not consciously registered, and that to a sub-threshold stimulus is if a stimulus of a different modality modulates the perception of a disappearance phenomenon but fails to improve the detection of a sub-threshold stimulus.

Visual adaptation, in particular Troxler fading, is a known disappearance phenomenon. After one gazes on a flickering colored object for a minute or so, it fades from perception and is filled in perceptually by the surround, which is of a different color (Shimojo & Kamitani, 2001; Troxler, 1804). Neuronal fatigue resulting from the over-stimulation of neurons in the lateral geniculate nucleus (LGN; Clarke & Belcher, 1962) and

---

[*] Corresponding author. Tel.: +1-626-395-2358; fax: +1-626-792-8583.
E-mail address: bhavin@caltech.edu (B.R. Sheth).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.03.009

the retinal ganglia (Kotulak & Schor, 1986; Millodot, 1967), two sites that lie upstream of striate cortex (V1), is believed to underlie Troxler fading (cf. Dennett's, 1991 criticism of the idea that there is a place in the brain where consciousness happens as "vestigial Cartesianism"). Of importance, there are some projections from auditory centers to V1 (Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2002), but none from auditory centers to the LGN or the retina. Thus, if an auditory stimulus were to interact with Troxler fading, it would have to be at a site downstream of where the initial fade took place. An effect of sound on Troxler fading would, therefore, be an example of high-level modulation of a low-level disappearance phenomenon. Although an auditory cue is known to modulate visual salience (Stein, London, Wilkinson, & Price, 1996; Vroomen & de Gelder, 2000), disappearance phenomena have never been studied in this regard.

Here, we report that repetitive auditory cues selectively enhance the visibility of synchronous visual targets, preventing them from undergoing visual adaptation and causing previously adapted targets to recover. Ours is one of the first reports of cross-modal modulation of a disappearance phenomenon.

2. Materials and methods

All stimuli were presented on a SONY Trinitron monitor (75 Hz refresh; 37.5 cm × 30 cm viewing area) under control of a MAC G4 running MATLAB (Mathworks Inc.) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The observers sat comfortably in a chair in front of the computer screen at a viewing distance of 57 cm, with their heads in a chinrest (Handaya Co., Japan). Two loudspeakers (Altec Lansing Satellite 31 speakers) were positioned 25 cm on either side of the middle and at the same vertical level as the visual display. Viewing was binocular. All experiments were conducted in room illumination.

2.1. Adaptation recovery task

The fixation point (FP), located in the center of a green screen (CIE coordinates, $x = 0.446$, $y = 0.568$, 37.9 cd/m^2), was a white circular dot (CIE coordinates, $x = 0.313$, $y = 0.329$, 100.4 cd/m^2, diameter = 12.4 arcmin). A red circular target (diameter = 5.0°, CIE coordinates, $x = 0.541$, $y = 0.451$, 34.7 cd/m^2), centered 10.4° left or right of the FP, appeared at the beginning of each trial. It lasted for 27 ms before being turned off. 133–293 ms (the actual time was randomly chosen from within this range) following its offset, a second target, identical in all respects to the first, appeared on the opposite side of the FP (Fig. 1A). It too stayed on for 27 ms and then turned off, followed, as before, by a random (133–293 ms) time interval after which the first target appeared again in its location. The adaptation cycle of target alternations lasted for 4 min. Each trial of the test phase lasted 24 s (50 alternations) total, and began with a top-up adaptation phase with the same cycle of alternations between the two targets for 16.8 s (35 alternations). After this, a click (2.2 ms duration, r.m.s. intensity = 71 or 42 dBa SPL measured in fast mode by a Bruel and Kjaer Microphone Amplifier and Meter, Model 2604; sound intensity made no difference to the results, see Sheth & Shimojo, 2002), occurred on both speakers, 14 ms after the onset of one of the targets, and the trial continued for another 7.2 s (15 alternations) with the sound synchronously presented with the chosen target. On a given trial, the selection of which target for synchronous presentation with the sound was random, but was counter-balanced between left and right choices across trials. Timing of sound and target were such that the sound was presented while the target was present on the screen. At trial's end, the targets and sound were turned off, and the observer had to report, in a self-paced manner, whether the left or right target appeared more salient after the sound came on. There were 30 trials total. The task typically took 20–25 min total.

2.2. Adaptation blocking task

On each trial, two targets, identical in stimulus parameters, alternately flickered on until the participant responded (same values as in the adaptation recovery task). One of the targets was accompanied by a synchronous sound (equal, non-zero amplitudes at DC and 11 kHZ frequencies, 2.2 ms duration, r.m.s. intensity = 71 dBa SPL) from the beginning of the trial. Each target stayed on for 27 ms during each cycle of adaptation, and there was a random 133–293 ms interval between targets. The participant viewed the screen until one of the targets (left or right) disappeared from perception. The trial ended when he or she responded with a keypress (left or right). There were 10 trials total. The sound was synchronous with each (left/right) of the two targets on 5/10 trials. A typical observer took 30 min to complete the task.

3. Results

3.1. Transient auditory cues transiently focus exogenous attention on visual stimuli nearby in time

Auditory–visual synchrony is usually the optimal temporal setting for audio–visual interaction. Various cross-modal effects peak at auditory–visual synchrony (Sekuler, Sekuler, & Lau, 1997; Shimojo, Miyauchi, & Hikosaka, 1997). Absolute judgments of effect strength are susceptible to non-sensory variables such as the

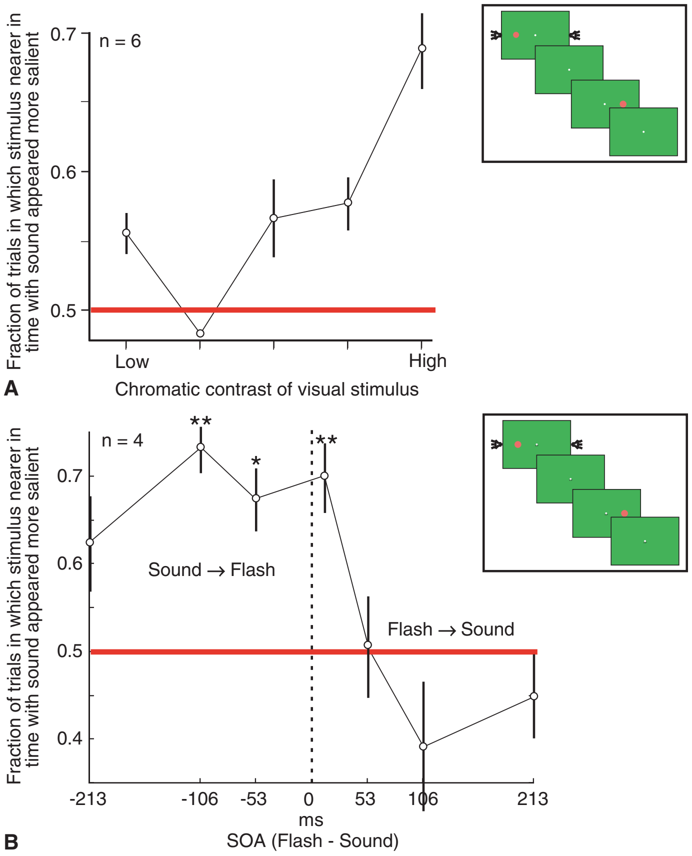

Fig. 1. Sound as an exogenous attentional cue. (A) Group reports of target salience as a function of target contrast on the two near-threshold targets experiment are shown. One of the targets (left or right, inset shows left) was synchronized with a sound from the start of each trial. Each point is the mean of 180 (30×6 observers) trials. Error bars are one between-observer standard error (s.e.m.). Color contrast $(\text{Red}/\text{Red}_{\text{max}})/(\text{Red}/\text{Red}_{\text{max}} + \text{Green}/\text{Green}_{\text{max}})$ of the red targets varied from 0.4971 (low) to 0.5421 (high) in three observers, and from 0.4762 (low) to 0.5189 (high) in the remaining three observers. (B) Group reports of relative visual target salience as a function of audio–visual target asynchrony. The inset illustrates the stimulus sequence in the synchronous sound condition, with the visual target on the left in this illustration. Negative values on the abscissa (points left of the 0 ms SOA dashed line) signify trials on which the sound physically preceded the visual target nearer in time. Each point is the mean of 120 (30×4 observers) trials. Error bars signify one s.e.m. Straight lines connect individual points. Asterisks denote significant (single asterisk: $p < 0.05$, double asterisk: $p < 0.005$) deviation from chance (50%, solid red line) as determined by a paired t-test for the given SOA. The 14 ms SOA data point was obtained in a different experimental session but on the same four observers (A, highest contrast stimulus). Although high in value (70%), the data point was not significant ($p > 0.1$).

observer's behavioral state, recent stimulus history and instruction set. Thus, a side-by-side comparison of the effect strengths of synchronous and asynchronous events *on the same trial* (called the paired-comparison method, Odgaard, Arieh, & Marks, 2003) is likely to be a better "criterion-free" test of the strength of the audiovisual interaction.

Two low contrast, barely visible targets, one located left and another right of fixation, alternately flickered on and off. One of the targets was synchronized with a sound from the start (Fig. 1A, inset). So as to keep adaptation to a minimum, this cycle was repeated only five times, with the sound present and synchronous with the same target all five times. Across trials, we varied target chromatic contrast within a small range around visibility. The observer's task was to report whether the left or right target appeared more salient, namely which target appeared more distinct from the background. By using two targets, the observer had to report whether the left or right target appeared more salient *after the sound came on*. By using two targets, one on either side of fixation, we minimized eye movement artifacts. As Fig. 1A depicts, over a range of chromatic contrasts, the observer predominantly judged the target synchronous with the sound as more salient. Thus, sound enhances the salience or perceptibility of a concurrent visual stimulus, and can be thought of as a temporal cue that focuses attention on a specific instant in time for preferential processing. Past studies have used sound to guide the observer's attention to a discrete moment in time (Stein et al., 1996; Vroomen & de Gelder, 2000).

To examine how the sound directed visual attention in this paradigm, we varied the SOA between the auditory cues and the visual targets (Fig. 1B, inset).

According to a stimulus-driven or exogenous attention account based on studies of visual attention (Jonides, 1981), an exogenous cue has maximum impact on the target if it precedes the target by no more than 50–100 ms. Assuming that exogenous cues of various modalities all have maximum impact at similar timing, a auditory cue preceding the visual target by 200 ms or less should enhance visual salience the most (McDonald, Teder-Sälejärvi, & Hillyard, 2000). On a second account, summation of auditory and visual signals enhances visual salience; so, the sound should have maximum impact if it is synchronous with the visual target (0 SOA), or, because auditory neural delays are known to be 50 ms shorter than visual (Goldstone & Lhamon, 1972; Keele, 1986), if the sound follows the target by 50 ms. On a third account, the sound is an endogenous, cognitive cue. Endogenous cues are typically slower because top-down signals are slow, so the sound should have maximum impact if it precedes the target by 200–400 ms (Posner, 1978; Posner, Nissen, & Klein, 1976; Posner, Nissen, & Ogden, 1978). Thus, SOA dependency provides a behavioral "fingerprint" for how auditory attentional cues impact visual perception.

The data were consistent with the first account: the auditory cue maximally enhanced visual salience if the sound preceded the visual target by 50–100 ms (Fig. 1B). Contrary to the second account, if the sound followed the target by 50–100 ms, it had little impact. Contrary to the third account, if the sound preceded the target by 250 ms (or more), it decreased visual salience, if anything. In sum, the sound behaved as an attentional cue that exogenously cues attention to the visual target appearing 50–100 ms later (McDonald et al., 2000; Vroomen & de Gelder, 2000; Watanabe & Shimojo, 1998; see Driver & Spence, 1998 for a review).

3.2. Auditory cues cause previously adapted visual stimuli synchronous with them to recover

Can an auditory stimulus that behaves as an exogenous cue focusing visual attention on certain moments in time modulate visual adaptation? Specifically, can repetitive auditory cues cause a previously adapted visual stimulus to recover from Troxler fading (i.e. deadapt the visual stimulus)? Endogenous visual attentional cues are known to enhance Troxler fading (Lou, 1999) but the effects of exogenous cues on Troxler fading have not been explored.

From a multisensory perspective, it is known that an auditory cue modulates visual salience. A visual target interspersed among non-targets was more detectable if a salient sound was played in approximate synchrony (SOA < 250 ms) with it (Vroomen & de Gelder, 2000). The target in Vroomen and de Gelder (2000) was an object (four spots arranged at the corners of an imaginary diamond); thus the auditory cue in their study probably affected visual processing in high-level cortical areas that process objects. In Stein et al. (1996), sound enhanced the perceived brightness of a synchronous foveal light; no comparison was made of the effects of synchronous and non-synchronous sounds, however. V1 is a known neural correlate of the brightness percept (Kinoshita & Komatsu, 2001; Rossi & Paradiso, 1999). Thus, it seems that sound can modulate visual percepts arising in cortex, including perhaps those that arise in V1 (Odgaard et al., 2003 claimed that the enhancement is eliminated if a paired-comparison method test is used and if the sound accompanies the light on 25% and not 50% of trials; they proposed an alternative account based on response bias). This claim is not surprising in light of recent evidence for connections from auditory centers to V1 (Bhattacharya, Shams, & Shimojo, 2002; Falchier et al., 2002; Giard & Peronnet, 1999; Rockland & Ojima, 2002; Shams, Kamitani, Thompson, & Shimojo, 2001). These findings beg the question of whether sound can affect the processing in even lower levels of the visual pathway. On the one hand, there are no known projections from the auditory brain areas to the LGN or the retina. On the other hand, the LGN receives profuse feedback projections from V1 (Erisir, Van Horn, & Sherman, 1997; Montero, 1991; Murphy, Duckett, & Sillito, 1999; Wilson, Friedlander, & Sherman, 1984), and V1 receives projections from auditory areas. Thus, the auditory areas → V1 → LGN polysynaptic pathway could be one route, albeit indirect, for sound to affect visual processing in the LGN. Because the LGN is a known neural correlate of Troxler fading (Clarke & Belcher, 1962), studying the effects of sound on Troxler fading may help answer the question of whether sound can effect processing in the LGN.

We employed near-threshold visual targets in an experimental paradigm and tested for the effects of sound on their perception. A target appeared momentarily to the left or right of fixation (FP in Fig. 2A). A random time (133–293 ms) following its offset, another target appeared on the opposite side of fixation, followed, between 133 ms to 293 ms later, by the first target. This adaptation cycle continued for 4 min, long enough for the perception of the two targets to be blurred and to be filled in by the background (Shimojo & Kamitani, 2001). There was a seamless transition into the test phase. Each trial of the test phase began with a 28 s long top-up adaptation phase. In the subsequent "sound on" phase of the trial, a clearly audible sound nearly synchronous with one of the two targets, was played. This was repeated five times, with the sound present and synchronous with the same target all five times. The random nature of the time interval (266–486 ms) between two successive sounds prevented the observer from ever acquiring prior knowledge of exactly when the sound would be played. At trial's end, the observer had to report whether the left or right target

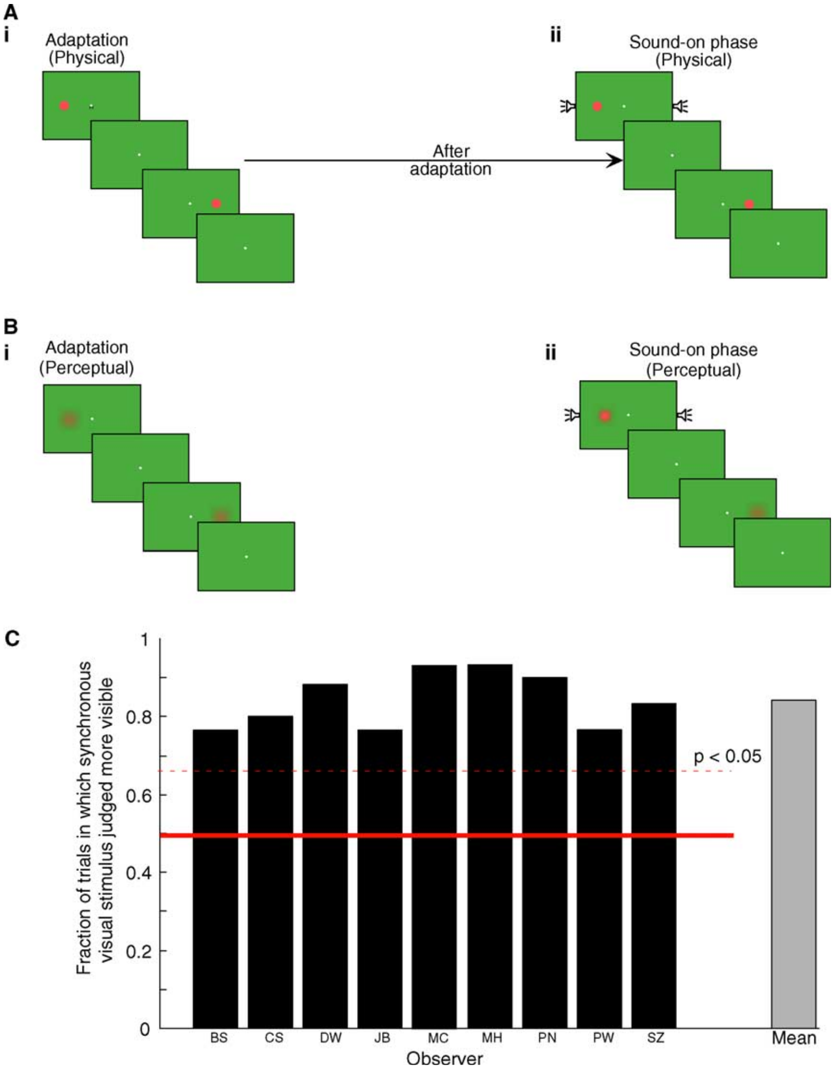

Fig. 2. Sound-induced recovery from visual adaptation (de-adaptation). (A) The sequence of stimuli and how the stimuli physically appear in the adaptation experiment are shown. Two red targets—one on either side of fixation (FP, shown in white), flickered back and forth on a green background in the adaptation phase (i). In the "sound on" phase that was at the end of each trial (ii), a sound was synchronously presented with one of the targets (left target in the figure). (B) How the stimuli typically appear to the observer at the end of adaptation (i; the green background appears to fill in), and in the "sound on" phase (ii; the left target that was accompanied in time by a sound appears more salient after the sound comes on). Observers claimed that, on some trials, the targets disappeared entirely from their perception. (C) Individual reports of target salience for each of the nine observers and the group mean in gray on the right is shown. The solid red line indicates chance performance (target synchronous with the sound appears salient on 50% of trials), and the red dashed line indicates significance at the 0.05 level (sign test, 70% of trials).

was more salient *after the sound came on* (paired-comparison method). Salience was defined subjectively for each observer—which target stood out more from the background (i.e. distinctive in space and time from the background) in terms of its flicker and red color. Having two targets on opposite sides of fixation helped minimize deviations of the eye from fixation.

Our observers typically found that, over the course of adaptation, both targets gradually took on the color of the background i.e. the background filled in (Fig. 2B, i). After the sound was turned on, perception changed: One or both targets gradually became more salient from the background (Fig. 2B, ii). Observers had to judge which of the two targets appeared more salient from the background once the sound came on. They tended to choose the target that was concurrent with the sound (Fig. 2C, mean ± one across-subject standard error: 84.3 ± 2.2% of all trials; also see Stein et al., 1996). This preference was significantly above chance (Fig. 2C, 50%, solid line) for each of the nine observers ($p < 0.05$, sign test for each. We nickname our effect the "beating heart effect", as the sound is critical and the phenomenology is akin to that of a beating heart). Anecdotally, observers reported that sometimes, only the synchronous flicker was visible, and other times, both targets stood out from the background but the one synchronous with the sound stood out first and more sharply. Still, the recovery from filling-in was partial and transient. It never equaled the visually guided recovery usually observed following an overt shift in eye position and consequent refreshing of the retinal image.

3.3. Response bias?

A critical issue is whether the de-adaptation was a genuine perceptual effect or whether it was due to response bias, namely that the sound did not affect perception but rather biased observers to respond for the light concurrent with the sound. That is to say, the sound merely affected post-perceptual decision processes, and not attention or perception.

It is necessary—but not sufficient—that the targets be clearly visible and temporally distinct so that the observer's post-perceptual mechanisms can match the sound with the visible target synchronous with it. We therefore conducted a new experiment to minimize the effect of response bias by ensuring that the targets were not visible following the adaptation. In order to achieve this, one major change was made in experimental design. Adaptation on a given trial did not terminate until the observer reported, via a keypress, that both targets were invisible. Following the keypress, the sound came on synchronous with one of the targets, and stayed on until the observer ($n = 4$) responded 'left' or 'right'. Observers ($n = 4$) responded 'left' or 'right'. The critical new element was that the targets were not visible when the sound was played, thus response bias could not play a major role. There was a significant preference ($60.0 \pm 5.1\%$ of all trials, $p < 0.05$, sign test) for the target synchronous with the sound, nonetheless. To minimize further the influence of response bias, we did an analogous experiment but now allowed the observer ($n = 5$) three choices—'left', 'right' or 'both'. Therefore, if the observer did not perceive either target to be more visible after the sound began playing, he or she could now report 'both'. Observers still reported 'left' or 'right', and not 'both', on 88% of all trials. Of the trials in which the observers did not report 'both', they reported the target synchronous with the sound on 65.2% of them. The preference was significantly above chance (50%, $p < 0.01$, sign test). The overall preference for the target synchronous with the sound was smaller (57.3% = 65.2% of 88% of all trials) than in the main experiment (84.3%, $p < 0.001$, Wilcoxon rank sum test). In sum, the experiments show that although response bias was not a non-factor, it was not the key factor.

In visual cortex, attentional cues have an effect akin to an increase in stimulus contrast (Martinez-Trujillo & Treue, 2002). The auditory cue that focuses attention on the concurrent visual target in our study may therefore be able to compensate for an enhancement in chromatic contrast in the non-concurrent target. To obtain a sensory measure of the effect of sound synchrony, we pitted the effect of sound synchrony with the sensory effect of an increase in chromatic contrast following the adaptation. There were two kinds of trials—sound following adaptation, and no sound following adaptation. Trials of both conditions were randomly intermixed. As in the previous experiment, observers ($n = 5$) waited until both targets were invisible before proceeding to the next stage of the trial. Immediately after the keypress (the FP changed color from white to black upon keypress on all trials) in the "no sound following adaptation" trials, the chromatic contrast of one target was increased ($0.207 \rightarrow 0.285$ or 0.361, depending on the observer). It was perceptibly redder—in the absence of prior adaptation-than before. On the "sound following adaptation" trials, simultaneous with the contrast increase, a sound began playing in synchrony with the remaining lower contrast target. The observer's task was to judge whether the target synchronous with the auditory cue or the target asynchronous with the auditory cue but of higher chromatic contrast appeared more salient. The experiment thus provided a quantitative albeit crude, *sensory* or *perceptual* measure of the impact of sound synchrony on visual perception. Observers did not know prior to the experiment that the contrast of one target was increased, or that the other was paired with a sound, nor did they acquire conscious knowledge of it later, as post-task questioning revealed. All observers favored the high chromatic contrast target on the "no sound following adaptation" trials ($65.0 \pm 4.0\%$ of all trials; observers reported 'both' on 1% of trials). All favored it substantially less on the "sound following adaptation" trials ($46 \pm 6.2\%$; observers reported 'both' on 0% of trials) when the target paired with a synchronous sound was in competition with it ($p < 0.05$, sign test). This suggests that the auditory stimulus had a measurable *perceptual* impact on vision.

3.4. Eye movements?

One account of sound-induced de-adaptation or recovery from adaptation is based on ventriloquism. Attraction of the perceived location of an unlocalizable sound causes reflexive eye movements to be more frequent toward the cued light. Studies of ventriloquized sound (Spence & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001) have shown that ventriloquism is automatic. Thus, the ventriloquism account indirectly supports our claim that the sound in our study was an exogenous cue that automatically drew attention to the concurrent light. Reflexive eye movements can thus be thought of as an overt end-result of attentional redirection toward the synchronous light by the unlocalizable sound, and not the cause of the de-adaptation.

Our observers were naïve to the purpose of the task, but were trained psychophysicists, so eye movements to the "cued" target synchronous with the sound were not likely. We monitored the eye movements (Eyelink II, SR Research) of two observers and observed almost no saccadic eye movements in any direction even after the sound came on. Moreover, the small drifts in eye position that occurred during viewing did not appear to correlate with their final choice (data not shown).

3.5. Attentional auditory cues prevent synchronous visual stimuli from undergoing adaptation

We have shown that sound causes previously adapted concurrent visual targets to recover. Could the enhanced salience imparted by the sound to the concurrent light buffer it from visual adaptation and help it fade slower than a second, identical light, but offset in time from the sound? We devised a new experiment (adaptation blocking task) to address the question. It differed from previous experiments in two ways. From the trial's start, there was a sound concurrent with one of the targets, and the targets alternately flickered back and forth until one or both disappeared from perception. The observer had to report whether the left or the right target faded first. If sound selectively prevents or slows down the adaptation of the co-occurring visual target, it should fade after the target not synchronous with the sound fades. Indeed, the co-occurring target was reported to fade first on 21.4 ± 5.5% of all trials. This was significantly less than chance (50%) for the group of observers ($n = 7$; $\chi^2_c = 21.73$, corrected using Yates correction for continuity, $p < 0.001$ [1]). In this experiment, it is possible—in spite of the high visual presentation frequency (2.1 Hz)—that observers knew which target the sound was concurrent with, so cognitive penetration may have contributed partly to the preference. However, observers were explicitly warned to judge target salience, not audio–visual synchrony, and none stated to have explicitly used synchrony as a proxy for salience. An account based on response bias would be hard-pressed to explain why observers were biased to respond for the synchronous target in the de-adaptation task, but for the non-synchronous target in the blocking task. Further, an endogenous attentional cue enhances Troxler fading (Lou, 1999). This is in contrast to our finding that auditory cues delayed fading, which implies that the sound in our study was an *exogenous*, and not an endogenous, cue (see Jonides, 1981 for differences between exogenous and endogenous cueing). Summing up, repetitive auditory cues both prevent lights synchronized with them from undergoing adaptation and cause previously adapted lights synchronous with them to recover.

3.6. Can sound improve the detection of a synchronous subthreshold visual stimulus

We have shown that auditory stimuli can modulate the perception of a visual stimulus that is not consciously registered but does elicit some response in the earliest sensory stages. We now ask whether sound can modulate the detection of a visual stimulus. Detection is related to adaptation, because adaptation to a visual stimulus generally impairs its detection. There are, however, important differences. The degree of adaptation to a visual stimulus can be modulated in multiple levels of the visual pathway, but the detection of a visual stimulus may not.

To explore this contrast further, we investigated whether repetitive auditory cues improve the detection of visual stimulus. A single target, located either left or right of fixation, flickered five times over the course of each trial (Fig. 3, inset). On a given trial, there was either a brief sound synchronous with the target, a sound not synchronous with the target, or no sound at all. The observer ($n = 6$, five naïve) had to report whether the target was left or right of fixation. The values of color contrast were the same as those used in the first experiment (see Fig. 1). The results, graphed in Fig. 3, demonstrate that sound, regardless of whether or not it was synchronous with the visual target, failed to improve detection performance (see Stein et al., 1996; Wuerger, Hofbauer, & Meyer, 2003 for converging evidence).

3.7. Can visual cues effectively substitute for auditory cues?

Is the effect of the auditory attentional cue on the recovery from and resistance to visual adaptation because the cue was an auditory stimulus or because it helped guide attention to the visual target (or both)? Replacing the auditory cues with visual cues will help dissociate between the two possibilities. As a visual analog to the auditory cue, a bright, horizontal line, bilaterally symmetric about fixation, appeared for 14 ms while either the left or the right target was on the screen in the "cue on" phase (Fig. 4A, right). Thus, the line, like the auditory cues of previous experiments, was a temporal cue, and not a spatial one. Observers ($n = 7$) weakly preferred the target simultaneous with the line (Fig. 4B) on 57.6 ± 5.1% of all trials, 26% less than the preference observed when auditory cues were used (cf. Fig. 2).

The weaker effect of the line cue compared with the sound could be because the line was not as salient as the sound. However, the line was visually salient. We do not think this was the case because the line was colored white on a green background and was brighter than the background. Nevertheless, in a second experiment, we used an even more salient visual cue—a large, bright,

---

[1] Heterogeneity $\chi_2 = 5.94$, $p > 0.25$, which justifies the pooling of data.

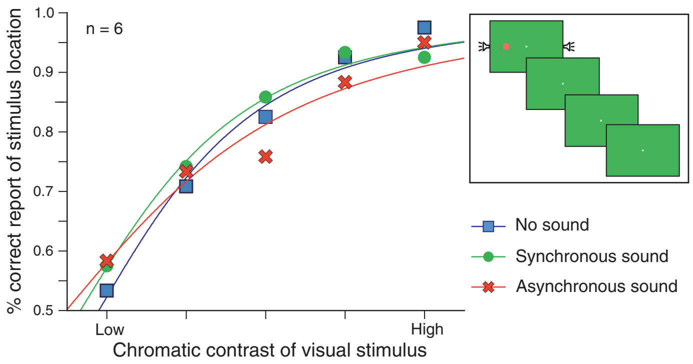

Fig. 3. Effect of sound on visual detection. Correct group reports of target location (left or right) as a function of target contrast on the single near-threshold target experiment is shown. The same values of color contrast as in Fig. 1A were used for the set of the same six observers. Each point is the mean of 180 (30×6 observers) trials. On a given trial, a sound was synchronous with the target (green circles and curve, see inset), the sound was present but was not synchronous with the target (red crosses and curve), or there was no sound at all (blue squares and curve). Error bars are not shown for purposes of clarity. The tiny shifts in the curves are accounted for solely by a large shift in author data. None of the naïve observers showed this trend.

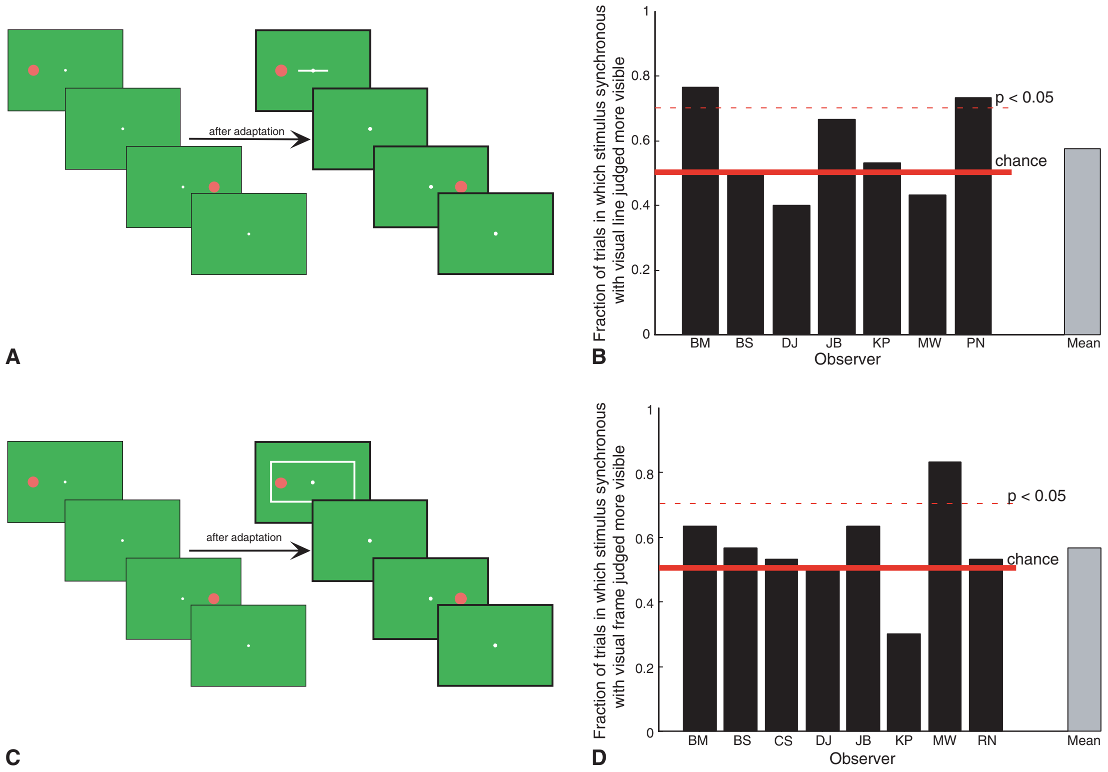

Fig. 4. The effects of a visual attentional cue on recovery from visual adaptation (de-adaptation). (A) The sequence of stimuli and how the stimuli physically appear in the adaptation (left) and post-adaptation "cue on" test (right) phases of each trial of the "line cue" experiment are shown. A horizontally symmetric, bright, white line centered on the FP is flashed in synchrony with one of the targets (left in the figure) in the "cue on" phase. (B) Individual and group mean (gray bar on right) reports of target salience on the "line cue" experiment are shown. (C) The sequence of stimuli and how the stimuli physically appear in the adaptation (left) and post-adaptation "cue on" test (right) phases of each trial of the "line cue" experiment are shown. A large, bright, white frame is flashed in synchrony with one of the targets (left in the figure) in the "cue on" phase. (D) Individual and group mean (gray bar on right) reports of target salience on the "frame cue" experiment are shown.

highly visible frame circumscribing both targets (Fig. 4C). This change did not enhance the preference, however: observers ($n = 8$) preferred but weakly the target simultaneous with the frame (mean = 56.7 ± 5.0% of all trials; Fig. 4D). In sum, in stark contrast to repetitive auditory cues, repetitive visual cues failed to enhance the saliance of the concurrent visual target.

This result is not too surprising in theory. The main difference between visual and auditory cues—besides the fact that they are of different modalities—is that visual cues, unlike auditory ones, must occupy discrete locations in space, which, in our experiments, are different from the locations of the targets. Thus, repetitive visual cues pull spatial attention away from targets located at different positions from the cues, which reduces their power to transiently enhance the signals of the targets synchronous with them (Spence & Driver, 2000; Vroomen et al., 2001). Moreover, visual–visual suppression is more potent than audio–visual suppression in masking and inhibition of return (Spence, Lloyd, McGlone, Nicholls, & Driver, 2000), which too would explain the lower potency of the visual cue in enhancing attention to the synchronous light. Regardless of cause, our experiments are among the first, to our knowledge, to find that an attentional cue of a different modality has a stronger effect on perception than a cue of the same modality. This implies there are functions a cross-modal cue can perform that a cue of the same modality cannot.

4. Discussion

We have shown that repetitive auditory cues selectively enhance the visibility of synchronous visual targets, preventing them from undergoing visual adaptation and causing previously adapted targets to recover. Thus, sound enhances visual perception by momentarily enhancing the visual signal that is near simultaneous with it. We offer a tentative account of our findings. The sensory signal in response to the visual stimulus competes with other signals from other visual stimuli for awareness. Depending on a variety of factors, including relative signal strength and the observer's locus of attention, the higher-level cortical mechanisms that gate or modulate conscious perception at the level of objects will allow the stimulus to be consciously perceived. If a cue helps focus one's attention to a target stimulus, the stimulus will be more likely, as compared to its competitors, to be consciously perceived (Kastner, De Weerd, Desimone, & Ungerleider, 1998). Audition interacts with vision at the level of cortex. The brain signal from the transient light that is synchronous with the transient auditory cue is selectively enhanced. This selective enhancement explains why a brief sound temporarily causes previously adapted visual targets synchronous with it to recover and prevents visual targets synchronous with it from undergoing visual adaptation.

There is evidence for high-level influence on filling-in, and Troxler fading, in particular. For instance, intermediate-level surface- and object-based fading has been observed (Ramachandran & Gregory, 1991). Prolonged observation of filled-in motion including the blind spot of one eye causes a motion aftereffect to be perceived in the other eye, again arguing for high-level influence over filling-in (Murakami, 1995). Patients with parietal lesions report accelerated Troxler fading (Holliday, Kennard, & Ruddock, 1985; Mennemeier et al., 1994) and fading of moving peripheral stimuli contralateral to their brain lesion (Mennemeier et al., 1994); in contrast, patients with frontal lesions rarely report Troxler fading (Mennemeier et al., 1994). These neurological results implicate high-level parietal and frontal areas in Troxler fading. Thus, Troxler fading can be modulated at higher levels of processing. Our study may be understood in this context: a cue of a different modality modulates filling-in at some stage beyond visual detection. In the context of multisensory integration, our study adds to the mounting evidence for high-level modulation of low-level percepts in the multimodal (Calvert et al., 1997; Giard & Peronnet, 1999; Laurienti et al., 2002; Shams et al., 2001; Taylor-Clarke, Kennett, & Haggard, 2002) and unimodal domains (Chauduri, 1990; Rees, Frith, & Lavie, 1997).

On a different, and arguably, more important note, this study is the first to show that an attentional cue of a different modality (audition) can modulate the perception of a disappearance phenomenon (Troxler fading). Our data further indicate what effects a cross-modal attention cue has on perception—it can enhance perception of an above-threshold stimulus, but cannot improve detection of a sub-threshold stimulus. Two accounts have been provided of disappearance phenomena—sensory suppression (Blake, 1989; Burbeck & Kelly, 1984) and higher-level selection (Logothetis, 1998; MacKay, 1986). Undoubtedly, Troxler fading results from local adaptation—a form of sensory suppression (Clarke & Belcher, 1962; Kotulak & Schor, 1986; Millodot, 1967). On the other hand, by all indications, the sound in our study was an attentional or selectional cue. Our study is among the first, therefore, to demonstrate that one mechanism, namely selection, can compensate, transiently and in part, for another, namely sensory suppression, thereby implying that the two mechanisms are inter-related.

Influential studies have claimed that filling-in is not an active process, but one based on mid-level long-range disinhibition or post-inhibitory rebound. Hardage & Tyler (1995) showed that the twinkle aftereffect observed in a visual patch following stimulation by a dynamic noise stimulus of a surrounding patch was dissociable from the filling-in that occured during stimulation, which supports the inhibition–disinhibition account of filling-in. Tyler & Hardage (1998) further showed that the disinhibition was restricted to the magnocellular system (no aftereffect in equiluminant noise or by noise below a temporal frequency of 5 Hz). In the context of this evidence, perhaps the sound in our study momentarily disinhibited the synchronous visual target—a transient disinhibition that lived and died with the sound—and, thereby provided a brief recovery from the visual adaptation. The fast temporal characteristics of the magno system render it suitable as a basis for the fast disinhibition. We believe neurophysiological studies will furnish the most rigorous test of our account.

Acknowledgements

We thank Dr. Chris Tyler and the two anonymous reviewers for their insightful suggestions and criticisms. We are grateful to Daw-An Wu for discussions. Special thanks to Prof. Mark Konishi for the generous loan of his sound pressure meter, and to the volunteers who kindly participated.

References

Andrews, T. J., & Purves, D. (1997). Similarities in normal and binocularly rivalrous viewing. *Proceedings of the National Academy of Sciences USA, 94*, 9905–9908.

Bennet-Clark, H., & Evans, C. (1963). Fragmentation of patterned targets when viewed as prolonged after-images. *Nature, 199*, 1215–1216.

Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport, 13*, 727–730.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review, 96*, 145–167.

Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature, 411*, 798–801.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 443–446.

Burbeck, C., & Kelly, D. (1984). Role of local adaptation in the fading of stabilized images. *Journal of Optical Society of America, 1*, 216–220.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. (1997). Activity of auditory cortex during silent lipreading. *Science, 276*, 593–596.

Campbell, F. W., Glinsky, A. S., Howel, E. R., Riggs, L. A., & Atkinson, J. (1973). The dependence of monocular rivalry on orientation. *Perception, 2*, 123–125.

Chaudhuri, A. (1990). Modulation of the motion aftereffect by selective attention. *Nature, 344*, 60–62.

Clarke, F. J. J., & Belcher, S. J. (1962). On the localization of Troxler's effect in the visual pathway. *Vision Research, 2*, 53–68.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: BackBay Books.

Driver, J., & Spence, C. (1998). Cross-modal attention. *Current Opinion in Neurobiology, 8*, 245–253.

Erisir, A., Van Horn, S. C., & Sherman, S. M. (1997). Relative numbers of cortical and brainstem inputs to the lateral geniculate nucleus. *Proceedings of the National Academy of Sciences USA, 94*, 1517–1520.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*, 5749–5759.

Giard, M. H., & Peronnet, M. (1999). Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*, 473–490.

Goldstone, S., & Lhamon, W. T. (1972). Auditory–visual differences in human temporal judgment. *Perceptual & Motor Skills, 34*, 623–633.

Hardage, L., & Tyler, C. W. (1995). Induced twinkle aftereffect as a probe of dynamic visual processing mechanisms. *Vision Research, 35*, 757–766.

Holliday, I. E., Kennard, C., & Ruddock, K. H. (1985). Rapid fading of visual sensations in a subject with a parietal–occipital tumour. *Ophthalmic and Physiological Optics, 5*, 149–156.

Jonides, J. (1981). In J. Long & A. D. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 187–204). Hillsdale, NJ: Erlbaum.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). De Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282*, 108–111.

Keele, S. W. (1986). In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. II. Cognitive processes and performance* (pp. 30.1–30.60). New York: Wiley and Sons.

Kinoshita, M., & Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neurophysiology, 86*, 2559–2570.

Kotulak, J. C., & Schor, C. M. (1986). The accommodative response to sub-threshold blur and to perceptual fading during the Troxler phenomenon. *Perception, 15*, 7–15.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience, 14*, 420–429.

Logothetis, N. K. (1998). Single units and conscious vision. *Philosophical Transactions of the Royal Society of London B, 353*, 1801–1818.

Lou, L. (1999). Selective peripheral fading: evidence for inhibitory sensory effect of attention. *Perception, 28*, 519–526.

MacKay, D. M. (1986). In J. Pettigrew, K. Sanderson, & W. Levick (Eds.), *Visual neuroscience* (pp. 365–373). Cambridge, UK: Cambridge University Press.

Martinez-Trujillo, J., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron, 35*, 365–370.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*, 906–908.

Mennemeier, M. S., Chatterjee, A., Watson, R. T., Wertman, E., Carter, L. P., & Heilman, K. M. (1994). Contributions of the parietal and frontal lobes to sustained attention and habituation. *Neuropsychologia, 32*, 703–716.

Millodot, M. (1967). Extra-foveal variations of the phenomenon of Troxler. *Psychologie Francaise, 12*, 190–196.

Montero, V. M. (1991). A quantitative study of synaptic contacts on interneurons and relay cells of the cat lateral geniculate nucleus. *Experimental Brain Research, 86*, 257–270.

Murakami, I. (1995). Motion aftereffect after monocular adaptation to filled-in motion at the blind spot. *Vision Research, 35*, 1041–1045.

Murphy, P. C., Duckett, S. G., & Sillito, A. M. (1999). Feedback connections to the lateral geniculate nucleus and cortical response properties. *Science, 286*, 1552–1554.

Odgaard, C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Perception & Psychophysics, 65*, 123–132.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Posner, M. I. (1978). *Chronometric explorations of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. *Psychological Review, 83*, 157–171.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). In H. L. Pick & B. J. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137–158). Hillsdale, NJ: Erlbaum.

Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature, 350*, 699–702.

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science, 278*, 1616–1619.

Rockland, K. S., & Ojima, H. (2002). Multimodal convergence in calcarine visual areas in macaque monkey, paper presented at the 3rd annual meeting of the International Multisensory Research Forum, Geneva, Switzerland.

Rossi, A. F., & Paradiso, M. A. (1999). Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. *Journal of Neuroscience, 19*, 6145–6156.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*, 308.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport, 12*, 3849–3852.

Sheth, B. R., & Shimojo, S. (2002). Recovery of visual perception from adaptation by sound: The cross-modal "beating-heart" effect, poster presented at the 2nd annual meeting of the Vision Sciences Society, Sarasota, FL.

Shimojo, S., & Kamitani, Y. (2001). Filling-in induced by high-contrast edge adaptation. *Journal of Vision, 1*, 53a.

Shimojo, S., Miyauchi, S., & Hikosaka, O. (1997). Visual motion sensation yielded by non-visually driven attention. *Vision Research, 37*, 1575–1580.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport, 11*, 2057–2061.

Spence, C., Lloyd, D., McGlone, F., Nicholls, M. E., & Driver, J. (2000). Inhibition of return is supramodal: a demonstration between all possible pairings of vision, touch and audition. *Experimental Brain Research, 134*, 42–48.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of Cognitive Neuroscience, 8*, 497–506.

Taylor-Clarke, M., Kennett, S., & Haggard, P. (2002). Vision modulates somatosensory cortical processing. *Current Biology, 12*, 233–236.

Troxler, D. (1804). In K. Himley & J. A. Schmidt (Eds.), *Ophthalmologisches bibliothek* (pp. 51–53). Fromman, Germany: Jena.

Tyler, C. W., & Hardage, L. (1998). Long-range twinkle induction: an achromatic rebound effect in the magnocellular processing system? *Perception, 27*, 203–214.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica, 108*, 21–33.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception & Performance, 26*, 1583–1590.

Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception, 27*, 1041–1054.

Wilson, J. R., Friedlander, M. J., & Sherman, S. M. (1984). Fine structural morphology of identified X- and Y-cells in the cat's lateral geniculate nucleus. *Proceedings of the Royal Society of London Series B—Biological Science, 221*, 411–436.

Wolfe, J. M. (1986). Briefly presented stimuli can disrupt constant suppression and binocular rivalry suppression. *Perception, 15*, 413–417.

Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics, 65*, 1188–1196.