

Multisensory Congruency as a Mechanism for Attentional Control over Perceptual Selection

Raymond van Ee,¹ Jeroen J. A. van Boxtel,² Amanda L. Parker,³ and David Alais³

¹Helmholtz Institute, Utrecht University, 3584 CH, Utrecht, The Netherlands, ²Division of Biology, Caltech, Pasadena, California 91125, and ³School of Psychology, University of Sydney, New South Wales 2006, Australia

The neural mechanisms underlying attentional selection of competing neural signals for awareness remains an unresolved issue. We studied attentional selection, using perceptually ambiguous stimuli in a novel multisensory paradigm that combined competing auditory and competing visual stimuli. We demonstrate that the ability to select, and attentively hold, one of the competing alternatives in either sensory modality is greatly enhanced when there is a matching cross-modal stimulus. Intriguingly, this multimodal enhancement of attentional selection seems to require a conscious act of attention, as passively experiencing the multisensory stimuli did not enhance control over the stimulus. We also demonstrate that congruent auditory or tactile information, and combined auditory–tactile information, aids attentional control over competing visual stimuli and *visa versa*. Our data suggest a functional role for recently found neurons that combine voluntarily initiated attentional functions across sensory modalities. We argue that these units provide a mechanism for structuring multisensory inputs that are then used to selectively modulate early (unimodal) cortical processing, boosting the gain of task-relevant features for willful control over perceptual awareness.

Introduction

Although lower organisms possess a direct coupling between sensory input and behavioral output, humans are able to intervene during this sequence and influence their output (Gilbert and Sigman, 2007), not only with respect to our motor actions but, intriguingly, also for our awareness. Although we are still learning about the precise mechanisms of this voluntary control and its necessary and sufficient conditions, we do know that it operates in a top-down manner through attention. For visual stimuli there is mounting evidence (Reynolds and Chelazzi, 2004) that attention to features and spatial locations can influence neural activity at early levels of cortical processing. It is unclear, however, how attention influences perceptual selection when multisensory signals are involved. A promising way to study awareness and voluntary attentional control over perception is to expose the sensory system to an ambiguous stimulus that generates bistable perception. This provides the opportunity to study multisensory processing related to the percepts rather than to the stimulus (Leopold and Logothetis, 1999; Blake and Logothetis, 2002; Tong, 2003).

Here, we used perceptually ambiguous stimuli in a novel multimodal paradigm that combined competing auditory stimuli and competing visual stimuli. We studied whether multisensory congruency facilitates voluntary control over perceptual selec-

tion, reasoning that this would open a novel window on multisensory aspects of perceptual control and shed light on the level at which it occurs. For unisensory stimuli, quite a few reports have shown a role for attention in voluntarily selecting one perceptual interpretation in perceptually bistable stimuli. These have shown that observers can lengthen the duration that the selected percept is dominant, but they cannot exert full control over the selection process and spontaneous perceptual alternations still occur (Lack, 1978; Meng and Tong, 2004; van Ee et al., 2005). Very recently, a degree of unisensory attentional control has also been demonstrated over ambiguous stimuli in the auditory domain (Pressnitzer and Hupé, 2006). Different senses interact with each other, and it is known from audiovisual experiments that a stimulus in one modality can change perception in the other (Sekuler et al., 1997; Shimojo and Shams, 2001; Alais and Burr, 2004; Witten and Knudsen, 2005; Ichikawa and Masakura, 2006). We combine these findings to study attentional control over perceptually ambiguous stimuli in a multisensory context, focusing on the role of cross-modal congruency. Congruency may facilitate multimodal mechanisms of voluntary control, since there is more support for one of the two competing percepts when there is information from another sensory modality that is congruent with it.

Many neurons in human posterior parietal and superior prefrontal cortices are involved in voluntary attentional shifts between vision and audition (Shomstein and Yantis, 2004), and attention to audiovisual feature combinations produces stronger activity in the superior temporal cortices than does attention to only auditory or visual features (Degerman et al., 2007). It has also been shown that auditory cortex can be profoundly engaged in processing nonauditory signals, particularly when those signals are being attended (for review, see Shinn-Cunningham,

Received Feb. 20, 2009; revised Aug. 1, 2009; accepted Aug. 6, 2009.

This work was financially supported by a High potential grant of Utrecht University to R.v.E. J.v.B. was partly supported by the Rubicon grant of the Netherlands Organization for Scientific Research. We thank our colleagues and a reviewer for helpful comments.

Correspondence should be addressed to Raymond van Ee, Helmholtz Institute, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands. E-mail: r.vanee@uu.nl.

DOI:10.1523/JNEUROSCI.0873-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2911641-09\$15.00/0

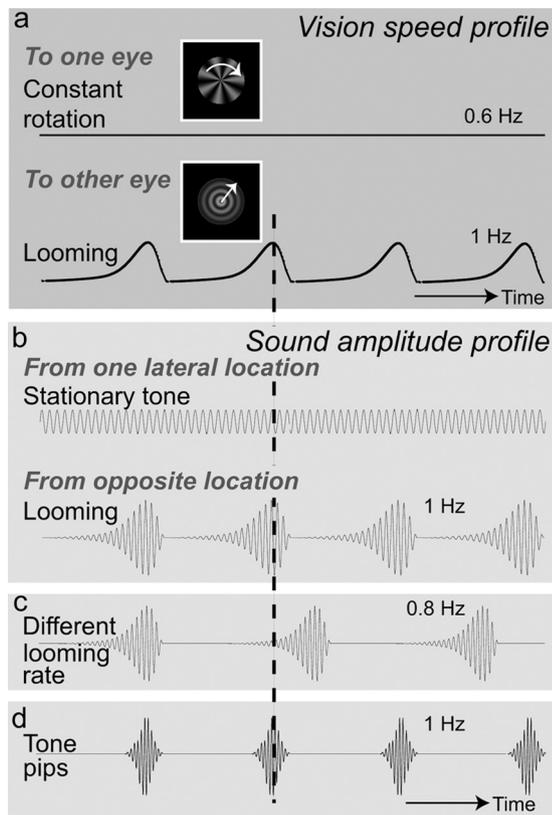


Figure 1. *a*, The vision speed profiles. The binocularly rivaling visual images consisted of a constantly rotating radial pattern and a looming concentric circle pattern. We deliberately designed the rotation rate (0.6 Hz) to be different from the looming rate (1 Hz) so that their changes over time did not match. *b*, The sound amplitude profiles. The two sounds that competed for attention consisted of a stationary tone triad (E chord) and a 1 Hz looming sound coming from opposite lateral locations (20 and -20° relative to straight ahead). Auditory and visual looming were rate matched and in phase (dashed vertical line). *c*, To examine whether it is either looming as such, or its rate, that caused the multimodal attentional control effects of the looming sound, we presented the looming sound with 0.8 Hz. *d*, To examine the role of sound rate, we presented tone pips with the same frequency (1 Hz) and phase as the looming visual pattern. Ramp and damp times were equal.

2008). What is the role of these multimodal attention-modulated neurons? None of the existing studies used competing cross-modal stimuli.

We studied whether the ability to voluntarily select one interpretation from an ambiguous visual or auditory stimulus would be enhanced when it was combined with auditory, visual, tactile, or auditory–tactile information that was congruent with that interpretation.

Materials and Methods

We presented subjects with a binocular rivalry (Levelt, 1965) stimulus consisting of a looming concentric pattern in one eye and a rotating radial pattern in the other eye (Fig. 1*a*) and a pair of auditory stimuli consisting of a looming sound and a spatially separated stationary tone triad (Fig. 1*b*). The auditory stimuli were binaurally presented over headphones (Pressnitzer and Hupé, 2006; Bidet-Caulet et al., 2007). The looming stimuli were rate matched (same frequency) in both the visual and auditory modalities. The rotating radial pattern was chosen to rival with the looming visual pattern, because it is orthogonal to the concentric looming stimulus and is symmetrical so that small eye movements in any direction would not unduly favor the visibility of one pattern over the other (Wade and de Weert, 1986; Parker and Alais, 2007). We deliberately designed the rotation rate to be different from the looming sound rate so that their changes over time did not match. Following previous

attentional studies of unisensory ambiguous perception, we examined voluntary control over visual rivalry by comparing “active” and “passive” conditions (Helmholtz, 1866; Lack, 1978; Peterson and Hochberg, 1983; Leopold and Logothetis, 1999; Suzuki and Peterson, 2000; Hol et al., 2003; Toppino, 2003; Meng and Tong, 2004; Chong et al., 2005; Slotnick and Yantis, 2005; van Ee et al., 2005, 2006; Brouwer and van Ee, 2006; Chong and Blake, 2006; Hancock and Andrews, 2007). In the passive condition, no attentional control was exerted in favor of either visual pattern. There were two types of active condition. In one, observers were instructed to “hold” the visual looming pattern dominant, and in the other they were instructed to hold the visual radial pattern. All three conditions were tested with and without the sound stimuli present, amounting to six conditions in total.

Visual stimuli. The competing binocular rivalry stimuli were a rotating radial sine wave pattern in one eye and a concentric sine wave pattern looming at 1 Hz in the other (Wade and de Weert, 1986; Parker and Alais, 2007). The visual stimuli had a mean luminance of 30 cd/m^2 , a contrast of 25%, and were presented in a Gaussian envelope ($\text{SD} = 0.6^\circ$) (Fig. 1*a*). The radial pattern consisted of seven cycles (propeller blades), rotating at $30.7^\circ/\text{s}$, producing a repetition frequency of 0.6 Hz at each visual location. The looming pattern had a spatial frequency of 3 c/degree , its motion being induced by a phase-shift that increased exponentially over a 1 s period from a baseline of 1 c/s to a maximum of 4 c/s , after which the increase was rapidly tapered off by a cosine profile. Continuous looming motion was created by repeating these profiles in a loop with 1 Hz so that it matched the looming sound rate and mismatched the rotation of the radial pattern. The stimuli were presented one on either side of a cathode ray tube monitor and viewed through a mirror stereoscope (viewing distance, 57 cm) to produce binocular rivalry. Stimuli were presented on a black square of $5 \times 5^\circ$, with a white border; the rest of the screen had a luminance of 30 cd/m^2 .

The visual stimulus used to disambiguate the ambiguous sound (experiment 6) consisted of a white flickering disk (diameter, 7.5° ; viewing distance, 57 cm) with a static frame around it. The disk flickered with on and off periods of 120 ms, equating the presentation sequence of the low tone in the ambiguous auditory stimulus.

Auditory stimuli. We used headphones to present competing stationary and looming sounds, meaning that attention needed to be used to follow the looming sound. The stationary sound was a constant, unmodulated tone triad (an “E major” chord) (Fig. 1*b*), with maximum amplitude of 76 dB sound pressure level (SPL) on average. The competing tone triad was present in all experiments in which we used competing sound to resolve ambiguity in the visual domain (thus, in all experiments, except in experiment 6). The looming sound was produced by modulating the amplitude of a pure tone (200 Hz) incremented from an amplitude of zero to a maximum amplitude designated by each subject to be comfortable (average, 74 dB SPL). The amplitude envelope had a profile identical to the phase-shift profile of the visual stimulus [1 Hz in experiment 1 (Fig. 1*b*); note that it was different, 0.82 Hz in experiment 2 (Fig. 1*c*)] and was precisely phase synchronous with the carrier sinusoid (200 Hz) to prevent readily detectable anomalies. To assign different spatial directions to the two sounds, they were both presented binaurally, with the looming sound having an interaural time difference of $+200 \mu\text{s}$ so that it was heard to originate from a location $\sim 20^\circ$ to the right (with respect to straight ahead) and the constant tone triad having an opposite phase difference of $-200 \mu\text{s}$ so that it was heard to originate from a location $\sim 20^\circ$ to the left. The “tone pips” that we presented in experiment 3 had a frequency of 1 Hz (i.e., at the visual looming frequency) (Fig. 1*d*), a duration of 280 ms, and an average maximum amplitude of 76 dB SPL.

In the experiment where we examined whether a visual stimulus can disambiguate an ambiguous sound stream (experiment 6), we followed a recent study on attentional control over auditory ambiguity (Pressnitzer and Hupé, 2006). We presented a high-frequency pure tone H alternating with a low-frequency pure tone L, in an LHL_ pattern (van Noorden, 1975). The frequency of H was 587 Hz and that of L was 440 Hz. The duration of each tone was 120 ms. The silence “_” that completed the LHL_ pattern was also 120 ms long. The sequence is perceived either as one stream (LHL-LHL, i.e., grouped

galloping rhythm) or as two streams (H-H-H-H and -L-L-, i.e., segregated Morse tones). The loudness of the tones was adjusted to a comfortable level (on average 75 dB SPL), which was kept constant during the experiment.

Tactile stimuli. To produce stationary and looming tactile stimuli, we detached a loudspeaker from its sound box (commercially available Logitech R-10 computer speakers). The vibrating speaker membrane was lightly attached to the skin of the dorsal side of the left hand by an elastic band (supplemental Fig. 6a, available at www.jneurosci.org as supplemental material). The hand was placed on the left knee underneath the stereoscope. By playing the looming sound exactly the same as in the basic experiment, the observer felt a “looming” pattern (although this was perceived as increasing pressure) that was matched to the visual looming pattern. In the no-sound conditions, observers wore earmuffs so that the sound of the vibrating membrane on the hand was not heard.

Procedure. Subjects were instructed to maintain fixation on the center of the visual pattern which is easily possible for a monocular looming pattern with a fixed reference around it and with the small size used at the distance presented (Erkelens and Regan, 1986). They pressed one of two keys when the visual looming stimulus was dominant and the other when the radial pattern was dominant. They were instructed to release both keys during instances of superimposed and piecemeal pattern perception (which averaged 13.7%), where neither pattern was exclusively dominant. We consistently compared passive and active conditions. In the passive condition, no attentional control was exerted. In one of the active attention conditions, observers were instructed to hold the visual looming pattern; in the other, observers were instructed to hold the visual radial pattern. Stimulus presentation series lasted 2 min. Between series, the stimuli were counterbalanced between the eyes, comprising sessions of 4 min per condition. All three conditions were tested with and without the sound stimuli present, amounting to six conditions in total and a duration of 24 min per experiment. Six subjects did three 24 min experiments. We established that there was no clear dependence on order and that fluctuations in mean predominance between repeated sessions were such that it was sufficient to ask the other subjects to do only one 4 min session per condition. We discarded the first 30 s of each series list for data analysis to ensure rivalry alternations had stabilized.

In the experiment where we examined whether a visual stimulus can disambiguate an ambiguous sound stream (experiment 6), the procedure was very similar. The experiment, again lasting 24 min, consisted of three 4 min sound-only and three 4 min sound plus vision sessions. In addition to the passive baseline condition, observers were instructed to hold the grouped sound (galloping) or the segregated sound stream (high and low Morse tones). The stimuli were presented using 4×1 min series per condition. In the no-vision conditions, a small marker was fixated. In the vision conditions, subjects fixated the center of the flickering disk.

In the experiment where we examined the role of congruent tactile “looming” (experiment 7), we compared the attentional gains for the sound-only, the tactile-only, and the tactile plus sound conditions. Observers were instructed to hold the visual looming pattern or to passively view the stimuli. Stimulus presentation sessions lasted again 4 min consisting of two 2 min series with stimuli counterbalanced between the eyes.

Informed written consent was obtained after the nature and possible consequences of the study were explained.

Results

Experiment 1: quantifying the influence of sound on attentional control

We first determined the baseline level of attentional control in unimodal, vision-only conditions by comparing hold versus passive conditions. Subjects ($n = 22$; 14 male, 8 female) tracked perceptual alternations in binocular rivalry. The mean perceptual durations for the looming visual pattern and the radial visual pattern (Fig. 2) are lengthened in the hold relative to the passive conditions (from 2.6 ± 0.2 s to 3.6 ± 0.3 for looming, and from 2.3 ± 0.2 s to 3.0 ± 0.3 s for radial patterns, both $p < 0.001$, paired t test), replicating previous work (Lack, 1978; Toppino, 2003; Meng and Tong, 2004; Chong et al., 2005; Slotnick and Yantis,

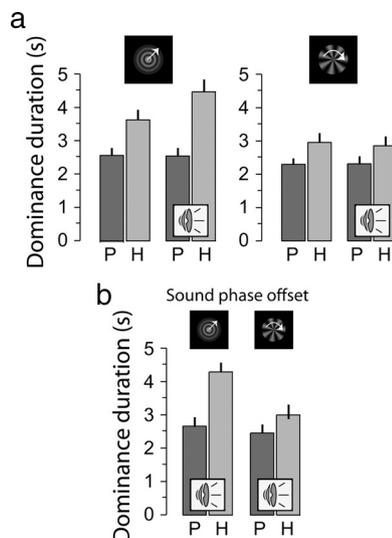


Figure 2. *a*, The mean duration for the looming (left) and the radial (right) visual patterns. Subjects were able to hold (light gray bars) the looming pattern for the no-sound condition (relative to the dark passive bar) and even more so with looming sound present (for right pair denoted with “speaker icons”). Error bars denote ± 1 SE. *b*, To test for the influence of a phase offset, five observers (the ones who participated in all 7 experiments) repeated experiment 1 but now with a sustained phase offset (a quarter of a period) of the sound stimulus relative to the visual stimulus. The data are essentially the same, as we found without phase offset.

2005; van Ee et al., 2005). In total, superimposed or piecemeal pattern perception averaged 13.7% of the observation period. Further details of the influence of attentional control over perception in unimodal conditions are presented in supplemental Results (supplemental Fig. 1; supplemental text, available at www.jneurosci.org as supplemental material). Interestingly, in the sound present conditions, the data suggest that the presence of an attended and matched sound enhances a subjects’ ability to select and hold a looming visual pattern (4.5 vs 3.6 s, in sound and no-sound conditions, respectively) (Fig. 2) but slightly impairs their ability to select and hold an unmatched (radial) visual pattern (2.9 vs 3.0 s). To quantify this multimodal attentional effect, we calculated the increase in perceptual duration for the hold task relative to the passive task and compared these values between sound-present condition and the no-sound condition. We defined the gain of multimodal attentional control as “hold-dependent increase in sound condition”/“hold-dependent increase in no-sound condition.” These gains implicitly normalize differences in attentional control across subjects and isolate the multimodal aspects of attentional control. They are plotted in rank order for all subjects in supplemental Fig. 1*b*, available at www.jneurosci.org as supplemental material.

For the hold-looming condition, the mean multimodal attentional gain amounted to $29.3 \pm 5.8\%$ ($p < 0.001$, t test) (Fig. 3*a*), indicating that subjects were more successful in holding the visual looming pattern when the matched looming sound was present than when it was absent. (Alternatively, this gain can be denoted as a ratio of 1.293; we will use the ratio and percentage notations interchangeably for ease of discussion. Since the metric for the statistic analysis is a linear transformation of the ratio used to denote multimodal attentional gain, it has no bearing on the results.) The same attention-related change in perceptual duration was calculated for the rotating radial pattern. The effect of voluntarily holding the visual radial pattern with the unmatched looming sound present was on average $-3.8 \pm 3.8\%$ (Fig. 3*a*), indicating that the presence of the looming sound decreased the

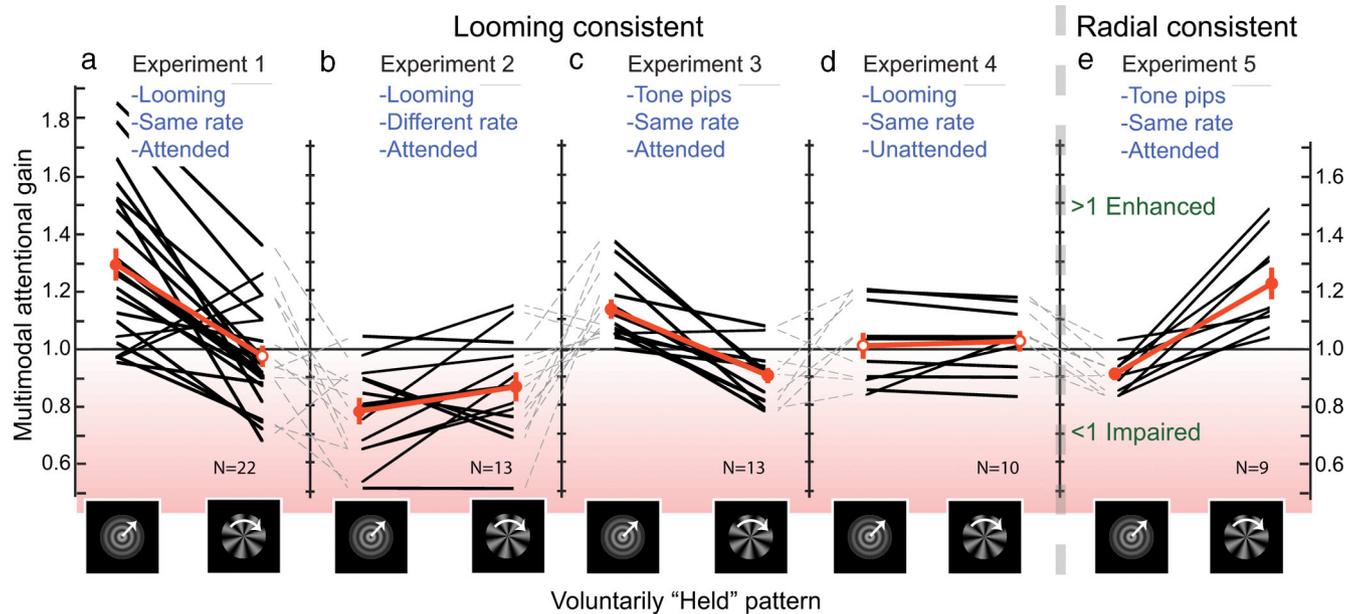


Figure 3. *a*, Results from experiment 1. Black lines connect the data of a particular subject. *b*, A different rate of the looming sound impairs the multimodal attentional gain (experiment 2). *c*, Tone pips with the same rate as the looming visual pattern enhance multimodal attentional control over the looming pattern, indicating that rate (or rhythm) is important (experiment 3). *d*, Paying attention to the congruent looming sound is required to enhance holding the visual looming pattern (experiment 4). *e*, The benefit of congruent sound also holds for the radial pattern and is not specific to the looming pattern (experiment 5). The dashed lines between the panels connect data of identical subjects. Filled circles indicate significance (*t* test, see Results); error bars denote ± 1 SE.

ability to attentionally hold the radial visual pattern, although not significantly ($p=0.3$, *t* test) (supplemental Fig. 1*b*, available at www.jneurosci.org as supplemental material). The attentional gains for the looming and radial patterns were uncorrelated (supplemental Fig. 1*c*, available at www.jneurosci.org as supplemental material) (linear regression: $r^2 = 0.055$, $p = 0.29$). This suggests that response bias did not cause the pattern of results, as there is no reason for the subject to assume that sound would facilitate holding the looming pattern but not the radial pattern. Results from the next experiment also add evidence against response bias.

It is worth noting that on average we did not find an influence of sound on mean percept durations for passive viewing. It has, nevertheless, been reported that concurrently presented looming sounds can increase perceptual dominance of a looming image in binocular rivalry even in passive viewing (Alais and Parker, 2006). However, this 2006 study was different in two ways: First, it did not include a comparison between attention and no-attention conditions. Including a passive “no-attention” condition in the current experiments may have mitigated the effect of involuntary automatic attention. Second, the previous report involved only a single auditory stimulus (a looming sound), whereas our study involved a looming sound and a second competing sound in the form of a constant tone triad. Even though the looming sound was clearly audible over the tone triad, it is possible that the requirement of attention for the cross-modal effect only applies when the critical sound is accompanied by a competing sound. That is, in cases where there are competing auditory stimuli, selectively attending to the relevant sound may be necessary. The absence of a competing stimulus could then explain why Parker and Alais (2006) got their cross-modal effect in the passive condition. Therefore, in all the following experiments, we present competing information in each of the sensory modalities. This, in turn, enables us to study multisensory processing related to the attentively selected percepts.

In sum, congruent sound aids attentional control over visual ambiguity. In this experiment, we started with a high level of congruency between auditory and visual information. In the next experiments, we systematically manipulate the congruency to determine the importance of aspects of congruency for multimodal attentional control. Our experiments capitalize on congruency in frequency (rhythm). Pilot experiments indicated that changing the phase (offset in time) of the sound relative to the visual pattern did not significantly affect the influence of sound. As experimenters, we noted during the programming of our stimuli that without objective measures it was hard to validate the phase offset in any of our conditions; even a phase offset of 1/4 of a period between the looming sound and the looming visual pattern went subjectively unnoticed. Our observers confirmed this, as they could readily match the perceived offset in timing between the two patterns, particularly when attention to the two sensory modalities was involved (Kanai et al., 2007). This happens in the real world, as when experiencing the periodicity of pile driving at a close distance or at a farther distance: the different transmission times for visual and auditory stimuli produce different offsets. The brain is able to deal with this by constantly calibrating the point of synchrony, as shown by adaptation to artificial temporal delays (Fujisaki et al., 2004). Figure 2*b* depicts objective data of the five subjects (who all participated in all experiments yet to be presented), showing that the mean percept durations for a 1/4 period phase offset (between looming sound and vision) was very similar to the mean percept durations with zero phase offset. Although it is possible that there may be a systematic temporal offset effect (such as in the recently reported enhanced perception of visual change by a coincident auditory tone pip) (van der Burg et al., 2008), from our pilot work we expect that in our setting it must be much smaller in magnitude than the frequency effect. Thus, the next set of experiments capitalizes on congruency in frequency (rhythm).

Experiment 2: the rate (rhythm) of the sound is key to enhance attentional control

We asked whether the synchronized periodicity in experiment 1 was necessary or whether non-synchronized looming sounds would be equally effective in promoting multimodal attentional control. We slightly changed the rate of the looming sound envelope to 0.82 Hz (Fig. 1c) and repeated the measurements with 13 subjects (9 male, 4 female) from experiment 1. In Figure 3b, we show individual subjects' multimodal attention gain. The slight change in auditory looming rate dramatically changed the effect of sound from one of enhancing attentional control of bistable visual perception to one of impairing attentional control, as shown by the impaired multimodal attentional gain (Fig. 3b, left) ($-20.3 \pm 4.4\%$; $p < 0.001$, t test) (see supplemental Fig. 2a, available at www.jneurosci.org as supplemental material, for percept durations). Thus, subjects were less able to hold the looming visual pattern when it was accompanied by a looming sound of a different rate than when sound was absent. After debriefing, observers reported that the mismatched looming sound was annoying and distracting, which may explain part of the impairing effects and might point to an automatic and obligatory component in cross-modal integration (Guttman et al., 2005; van der Burg et al., 2008). In turn, this might mean that the cross-modal effect obtained may have been partly motivational; subjects might have made more effort to control bistability when the sound matched the to-be-held visual motion, whereas they might have been less motivated when the sound "annoyingly" mismatched the to-be-held visual motion. However, that would likely only cause the disappearance of the attentional effect measured in experiment 1, not the observed decrease in dominance durations. Supporting an automatic component is the reported presence of cortical activity specifically related to coincident visual and auditory looming stimuli (Maier and Ghazanfar, 2007) and points to the functional significance of looming (approaching) stimuli (Neuhoff, 2001; Parker and Alais, 2007).

This experiment underscores the importance of temporal congruency, versus the looming character of the sound, in enhancing attentional control. Because subjects were still explicitly instructed to pay attention to the looming sound, this finding supports the conclusion that the results of experiment 1 were not a bias in response to instructions. Also, consistent with the findings of experiment 1, the ability to voluntarily hold the visual radial pattern was impaired (relative to no sound) when a looming sound was present ($-12.2 \pm 4.8\%$; $p < 0.03$) (Fig. 3b, right; supplemental Fig. 2a, available at www.jneurosci.org as supplemental material).

Experiment 3: rhythmic tone pips also enhance attentional control

If congruent rate is the key factor as indicated by experiment 2, would another sound with a rate identical to the visual looming stimulus be sufficient to enhance attentional control? We tested this on the same 13 subjects (9 male, 4 female) using discrete tone pips presented at 1 Hz (i.e., at the visual looming frequency) (Fig. 1d) and found significant multimodal gain in holding the looming pattern dominant ($13.8 \pm 3.2\%$, $p < 0.001$) (Fig. 3c; see supplemental Fig. 2b, available at www.jneurosci.org as supplemental material, for percept durations). Although there is a significant decrease in effect compared with experiment 1 ($p < 0.05$, t test), this difference disappears when comparing only the subjects that participated in both experiments ($p > 0.2$, paired t test), suggesting again that the looming character of the sound was not a cardinal factor of congruency. Furthermore, there was

a small impairment in the ability to hold the radial pattern relative to the no-sound condition of $-8.4 \pm 2.5\%$ ($p < 0.006$) (Fig. 3c; supplemental Fig. 2b, available at www.jneurosci.org as supplemental material). Therefore, the rate of the auditory signal is the factor that governs multimodal control of visual ambiguity, rather than the sound's looming-like envelope.

Experiment 4: paying attention to the sound is essential to enhance attentional control

In the experiments above, subjects were explicitly instructed to pay attention to the sound. We wished to determine whether paying attention to the sound was essential for multimodal control to occur. A group of 10 subjects (7 male, 3 female) who had not participated in any of the previous conditions performed an additional experiment before experiment 1. They were given the instruction that the sound was not relevant to their task, although no explicit instruction was given to attend or to disregard the sound. Interestingly, we found for this group that multimodal gain was not significantly different from zero [$1.6 \pm 4.3\%$, $p > 0.70$; and $-3.1 \pm 3.4\%$, $p > 0.40$, for the looming and radial visual patterns, respectively (Fig. 3d); and see supplemental Fig. 3a, available at www.jneurosci.org as supplemental material, for percept durations], meaning that the mere presence of a matched looming sound did not automatically trigger the ability to select and hold the looming visual stimulus. Instead, control over the visual stimulus requires an explicit act of attention to the sound stimulus. We then let this group of subjects do the previously described experiments, for which they showed average behavior (other panels, Fig. 3). Supplemental Fig. 3b, available at www.jneurosci.org as supplemental material, directly compares the data of experiments 1 and 4 for each individual of this group of subjects and emphasizes that the subjects' ability to hold the visual stimulus was profoundly enhanced once the sound was attended. One could argue that the absence of an effect for these subjects could be attributable to being unpracticed at the task. However, when comparing these subjects' results in experiment 1 (their second experiment) to the subjects whose first experiment was experiment 1, there was no significant difference ($p > 0.8$, t test) (supplemental Fig. 3c, available at www.jneurosci.org as supplemental material).

Experiment 5: generalization to other visual patterns—rhythmic tone pips enhance attentional control over the radial pattern

Does ambiguity resolution hold for the radial pattern as well? Discrete tone pips (the same as used above in experiment 3) were presented at 0.6 Hz to match the rotational frequency of the radial pattern. The tone pips were timed to occur each time a spoke pointed exactly downward, and this was explicitly indicated to the subjects ($n = 9$, 6 male, 3 female; participated in all previous experiments). We found significant multimodal gain in holding the radial pattern dominant ($22.4 \pm 5.3\%$, $p < 0.006$) (Fig. 3e; see supplemental Fig. 4, available at www.jneurosci.org as supplemental material, for percept durations) when this train of congruent tone pips was present and attended. Conversely, the ability to voluntarily hold the looming pattern was impaired (relative to the no-sound condition) when attending the tone pips ($-7.8 \pm 2.1\%$, $p < 0.003$) (Fig. 3e; supplemental Fig. 4, available at www.jneurosci.org as supplemental material).

Together with experiments 1 and 3, these results reveal that the resolution of visual ambiguity by congruent sound is not specific to looming visual stimuli as it also occurs when the auditory stimulus is temporally congruent with radial visual stimuli.

Experiment 6: a congruent visual pattern aids in control over ambiguous sounds

Thus far, our experiments have involved the resolution of ambiguity in the visual domain by a congruent auditory stimulus. Would congruent vision also facilitate control of ambiguous auditory signals? To test this, we presented subjects with alternating high and low tones (van Noorden, 1975), where observers either hear segregated tone streams (Morse) or a grouped (galloping) pattern (supplemental Fig. 5*a*, available at www.jneurosci.org as supplemental material). This stimulus has become a standard way to study auditory scene analysis, and in the only extant study that addressed attentional control of this ambiguous auditory stimulus, observers were able to lengthen the duration of the dominance of one of the alternatives (Pressnitzer and Hupé, 2006). Interestingly, there is evidence that perception of ambiguous stimuli in the auditory domain can be biased by a visual stimulus (O'Leary and Rhodes, 1984), but this study did not specifically address the role of attentional control. As an unambiguous visual stimulus, we used a disk (diameter, 7.5°; see Materials and Methods) flickering at the low tone frequency. As a competing visual stimulus, we presented a static frame around the disk. We tested seven subjects (4 male, 3 female) who had all participated in the other experiments. We explicitly asked observers to pay attention to the flicker frequency of the disk, as pilot experiments made it readily obvious that without actively viewing the disk there is no effect of the presence of the disk (see supplemental Fig. 5*d*, available at www.jneurosci.org as supplemental material, for pilot data). Thus, even though subjects had already participated in experiments one to five in which they matched the frequency of a sound stimulus to a visual stimulus, in this experiment they did not automatically match the frequency of the visual stimulus to the sound stimulus. Interestingly, this was even the case when we presented the disk (and the frame) on a large projector screen with a diameter subtending a visual angle of 80° horizontal by 60° vertical. It must be said though that these pilot experiments using the whole field projection were ran for 1 min only. It could be the case that prolonged subjection to a whole field visual stimulus with the same frequency as the auditory stimulus might lead to automatic cross-modal effects.

The results, using the 7.5° disk, showed a significant multimodal gain in holding the segregated Morse-like percept dominant ($23.9 \pm 9.1\%$, $p < 0.039$) when the flickering visual disk was present and attended (supplemental Fig. 5*c*, available at www.jneurosci.org as supplemental material; Fig. 4, fourth bar). Conversely, there was an insignificant impairment in the ability to voluntarily hold the grouped (galloping) percept dominant when attending to the visual pattern ($-3.4 \pm 9.9\%$, $p > 0.7$) (supplemental Fig. 5*c*, available at www.jneurosci.org as supplemental material). These results reveal that the resolution of ambiguous perceptual signals by congruent stimuli is not limited to an auditory influence on visual processing but can also operate in the reverse direction with vision disambiguating sound.

Experiment 7: generalization to touch—a tactile pattern aids in disambiguating vision

To this point, our experiments have involved vision and audition. Here, we introduce a tactile stimulus to test the prediction that trimodal congruency aids in attentional control over the visual looming pattern. To make a temporally congruent tactile signal, we attached a vibrating loudspeaker membrane to the skin on the back of the hand (supplemental Fig. 6*a*, available at www.jneurosci.org as supplemental material) and played the same

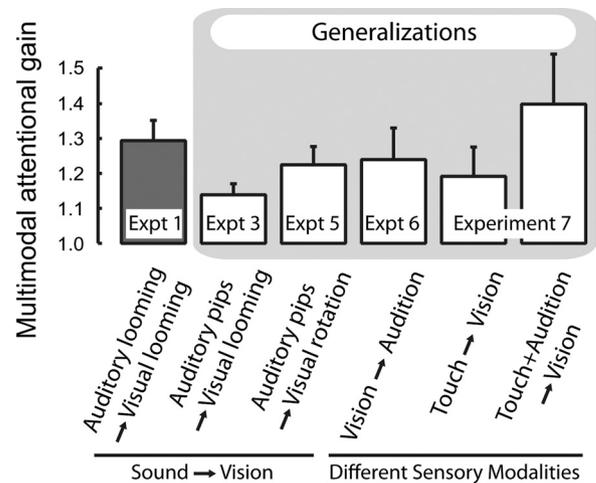


Figure 4. The first bar from the left shows the basic finding (Fig. 3*a*, experiment 1). The second bar depicts a generalization across attended sound patterns: tone pips provided significant multimodal attentional gain in holding the visual looming pattern dominant (Fig. 3*c*, experiment 3). Next, the third bar shows that congruent tone pips aided the visual radial pattern as well, generalizing our findings to visual stimuli other than looming patterns (Fig. 3*e*, experiment 5). The three right most bars show generalizations to other sensory domains. The fourth bar shows the role of vision on the dominance of competing sounds (supplemental Fig. 5, available at www.jneurosci.org as supplemental material, experiment 6). The fifth bar shows the influence of touch on active visual ambiguity resolution, and the sixth bar depicts the combined effect of touch and audition on the ability to actively control visual ambiguity resolution (supplemental Fig. 6, available at www.jneurosci.org as supplemental material, experiment 7). Error bars denote ± 1 SE.

competing sounds as we used in experiment 1, being the looming and the tone triad. The looming sound was felt as a pulsing pattern that was temporally matched to the looming visual pattern. The tone triad was felt as tactile noise. Again, we explicitly asked observers to pay attention to looming feeling as pilot experiments made it readily obvious that without actively attending to the tactile looming there was no effect of the presence of the tactile stimulus.

Five subjects (3 male, 2 female) participated, all of whom had participated in all other experiments. First, we tested the bimodal visuo-tactile condition. With only the congruent tactile stimulus accompanying the ambiguous visual stimuli, there was a significant multimodal gain in the ability to hold the visual looming pattern dominant ($19.1 \pm 8.4\%$, $p < 0.05$, one-tailed t test) (supplemental Fig. 6*c*, available at www.jneurosci.org as supplemental material; Fig. 4, fifth bar). In the trimodal condition, when congruent stimuli were present in both the auditory and the tactile domain, an even stronger effect was observed, with multimodal gain in holding the visual looming pattern dominant increasing to $39.7 \pm 14.3\%$ ($p < 0.05$) (supplemental Fig. 6*c*, available at www.jneurosci.org as supplemental material; Fig. 4, sixth bar), which was significantly higher than the tactile condition ($p < 0.05$, one-sided paired t test). Importantly, and as found above, the ability to hold the radial pattern was not facilitated by tactile stimuli ($18.6 \pm 12.4\%$; $p > 0.2$), nor by combined audio-tactile stimuli ($-3.5 \pm 4.0\%$, $p > 0.4$). Individual subject data are provided in supplemental Fig. 6, available at www.jneurosci.org as supplemental material.

Discussion

The primary finding is that the presence of an attended sound matching the temporal rate of one of a pair of competing ambiguous visual stimuli allows subjects much more control over vol-

untarily holding that stimulus dominant. Attentional control over the other, temporally mismatched, visual pattern was also influenced by the sound but in the opposite manner. The size of this effect is remarkably large, given that attentional control over binocular rivalry is usually found to be quite weak (Meng and Tong, 2004; Chong et al., 2005; van Ee et al., 2005; Paffen et al., 2006). Importantly, we also showed that active attention to both the sound and the visual stimulus promoted enhanced voluntary control. Below, we argue that this may help to explain why other researchers in psychophysics have failed to find such intimate links between auditory and visual attentional control. We also demonstrated a facilitatory relationship in the opposite direction in that attentional control over audio ambiguity is markedly aided by a matching visual stimulus. Extending this generalization, we demonstrated that a matching tactile stimulus enhanced attentional control in perceptually selecting competing visual stimuli and that this control was further strengthened in a trimodal condition that combined congruent audio-tactile stimuli with the bistable visual stimulus. Figure 4 summarizes the generalization of results across different visual patterns, sound patterns, and sensory modalities.

When the sound was temporally delayed, subjects still sensed that vision and sound were linked because of their constant phase relationship (Fig. 2*b*). In addition, although we have only provided formal evidence for a mandatory involvement of directed attention in the sound-on-vision experiments (Fig. 3*d*), our pilot work (supplemental Fig. 5*d*, available at www.jneurosci.org as supplemental material) and the available literature suggest that attention must be engaged to promote cross-modal interactions (Calvert et al., 1997; Gutfreund et al., 2002; Degerman et al., 2007; Mozolic et al., 2008; for review, see Shinn-Cunningham, 2008). Nevertheless, although a systematic investigation of temporal offset and automation for the cross-modal effects goes beyond the scope of the present paper, it is interesting to note that the underlying rhythm mechanism for our rhythm-based effect may be different from the mechanism underlying automatically occurring coincidence-based auditory-visual interactions (such as in the reported enhanced perception of visual change by a coincident auditory tone pip) (van der Burg et al., 2008).

Our study is unique in that it uses competing bistable visual and bistable auditory stimuli, providing the opportunity to study how competing sensory processing in two modalities (related to percepts rather than physical stimuli) are influenced by signals from other modalities. How do our findings shed light on the mechanisms underlying the resolution of perceptual ambiguity? We suggest that the enhanced capacity for attentional selection of the congruent stimulus results from a boost of its perceptual gain, which is attributable to top-down feedback from multisensory attentional processes that select the congruent feature of the input signal. In support of this, for vision, it has been shown previously that the effect of top-down attention on extending dominance durations for perceptually competing stimuli is equivalent to a boost in stimulus contrast (Chong et al., 2005; Chong and Blake, 2006; Paffen et al., 2006). This is in line with recent studies on visual spatial and feature attention in psychophysics (Blaser et al., 1999; Carrasco et al., 2004; Boynton, 2005) and neurophysiology (Reynolds and Chelazzi, 2004) which demonstrate that the neural mechanism underlying attentional selection involves boosting the gain of the relevant neural population. This is observed in the early cortical stages of both visual (Treue and Maunsell, 1996; Treue and Martínez Trujillo, 1999; Lamme and Roelfsema, 2000; Womelsdorf et al., 2006; Wannig et al., 2007) and auditory processing (Bidet-Caulet et al., 2007). From

the present results, we can conclude that the scope of this feedback process can be extended to incorporate relevant multimodal signals. Thus, it appears that voluntary control over ambiguity resolution can be modeled as an increase in effective contrast (perceptual gain) of stimulus elements involving feature attention, as opposed to spatial attention. Dovetailing with this, voluntary control in perceptual bistability depends multiplicatively on stimulus features (Suzuki and Peterson, 2000), and an equivalence between stimulus parameter effects and attentional control is evident even at the level of fit parameters to distributions of perceptual duration data (Brouwer and van Ee, 2006; van Ee et al., 2006). It can also be demonstrated quantitatively, as in a recently developed theoretical neural model (Noest et al., 2007), that attentional gain modulation at early cortical stages is sufficient to explain all reported data on attentional control of bistable visual stimuli (Klink et al., 2008). Thus, there is converging evidence that an early gain mechanism is involved in attentional control of perceptual resolution of ambiguous stimuli, although it is too early to entirely rule out high-level modification.

Although there is support for the idea that auditory and visual attention are processed separately (Shiffrin and Grantham, 1974; Bonnel and Hafter, 1998; Soto-Faraco et al., 2005; Alais et al., 2006; Pressnitzer and Hupé, 2006; Hupé et al., 2008), our findings support the neurophysiological literature (Calvert et al., 1997; Gutfreund et al., 2002; Shomstein and Yantis, 2004; Amedi et al., 2005; Brosch et al., 2005; Budinger et al., 2006; Degerman et al., 2007; Lakatos et al., 2007, 2008; Shinn-Cunningham, 2008) that the mechanisms mediating multisensory attentional control are intimately linked. To understand these seemingly disparate results, note first that psychophysical studies finding separate processing, focused on spatial attention, as opposed to our study. Our findings concern feature attention and agree with recent findings that feature attention can more profoundly influence processing of stimuli than spatial attention (Melcher et al., 2005; Kanai et al., 2006). Note further that we presented the matched audio and visual stimuli simultaneously. The only other study on attentional control of ambiguous auditory and visual stimuli (Pressnitzer and Hupé, 2006) presented the stimuli from the two modalities separately in time, finding that results from the two modalities were unrelated. Although there are studies reporting that audiovisual stimulus combination is mandatory (Driver and Spence, 1998; Guttman et al., 2005), this is not a general view (Shiffrin and Grantham, 1974; Bonnel and Hafter, 1998; Soto-Faraco et al., 2005; Alais et al., 2006; Hupé et al., 2008). Our experiments address this by using perceptually ambiguous competing auditory and visual stimuli, thereby dissociating attention and stimulation to reveal that active attention to both modalities promotes audiovisual combination, in line with other recent studies (Calvert et al., 1997; Gutfreund et al., 2002; Degerman et al., 2007; Mozolic et al., 2008).

Our data suggest a functional role for neurons recently found in human posterior parietal, superior prefrontal, and superior temporal cortices that combine voluntarily initiated attentional functions across sensory modalities (Gutfreund et al., 2002; Shomstein and Yantis, 2004; Degerman et al., 2007). We suggest that when the brain can detect a rhythm in a task, attention feeds back to unisensory cortex to enforce coherent and amplified output of the matching perceptual interpretation. Recently, neurophysiologists were able to demonstrate that an attended rhythm in a task enforced the entrainment of low-level neuronal excitability oscillations across different sensory modalities (Lakatos et al., 2008). The fact that oscillations in V1 entrain to attended auditory stimuli just as well as to attended visual stimuli rein-

forces the view that the primary cortices are not the exclusive domain of a single modality input (Foxe and Schroeder, 2005; Macaluso and Driver, 2005; Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007; Lakatos et al., 2007) and confirms the role of attention in coordinating heteromodal stimuli in the primary cortices (Brosch et al., 2005; Budinger et al., 2006; Lakatos et al., 2007, 2008; Shinn-Cunningham, 2008). We suggest that the same populations of neurons may control multimodal sensory integration and attentional control, suggesting that the neural network that creates multimodal sensory integration may also provide the interface for top-down perceptual selection. However, our understanding of multisensory neural architecture is still developing (Driver and Noesselt, 2008; Senkowski et al., 2008) and a competing view, rather than focusing on feedback from multisensory to unisensory areas, proposes that multisensory interactions can occur because of direct feedforward convergence at very early cortical areas previously thought to be exclusively unisensory (Foxe and Schroeder, 2005; Ghazanfar and Schroeder, 2006). Testing competing views will require further studies, possibly using neuroimaging techniques with high temporal resolution or neurodisruption techniques to temporarily lesion the putative higher-level area.

Conclusion

In sum, our novel paradigm involving ambiguous stimuli (either visual or auditory) enabled us to demonstrate that active attention to both the auditory and the visual pattern was necessary for enhanced voluntary control in perceptual selection. The audiovisual coupling that served awareness was therefore not fully automatic, not even when they had the same rate and phase. This suggests a functional role for neurons that combine voluntarily initiated attentional functions across different sensory modalities (Calvert et al., 1997; Gutfreund et al., 2002; Shomstein and Yantis, 2004; Amedi et al., 2005; Brosch et al., 2005; Budinger et al., 2006; Degerman et al., 2007; Lakatos et al., 2007, 2008), because in most of these studies congruency effects were not seen unless attention was actively used. This squares with psychophysics and neurophysiology showing intimate links between active attention and cross-modal integration (Spence et al., 2001; Kanai et al., 2007; Lakatos et al., 2007; Mozolic et al., 2008; Shinn-Cunningham, 2008). Thus, these attention-dependent multisensory mechanisms provide structure for attentional control of perceptual selection in two ways. First, in responding to intermodal congruency, they may boost the baseline response of the congruent alternative (as there is more “proof” for a perceptual interpretation when it is supported by two converging modality sources). Second, they may increase attentional control over perceptual selection because a multiplicative gain will be more significant when acting on a higher baseline, therefore allowing more attentional control.

References

- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262.
- Alais D, Morrone C, Burr D (2006) Separate attentional resources for vision and audition. *Proc Biol Sci* 273:1339–1345.
- Amedi A, Malach R, Pascual-Leone A (2005) Negative BOLD differentiates visual imagery and perception. *Neuron* 48:859–872.
- Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O (2007) Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J Neurosci* 27:9252–9261.
- Blake R, Logothetis NK (2002) Visual competition. *Nat Rev Neurosci* 3:13–21.
- Blaser E, Sperling G, Lu ZL (1999) Measuring the amplification of attention. *Proc Natl Acad Sci U S A* 96:11681–11686.
- Bonnell AM, Hafter ER (1998) Divided attention between simultaneous auditory and visual signals. *Percept Psychophys* 60:179–190.
- Boynton GM (2005) Attention and visual perception. *Curr Opin Neurobiol* 15:465–469.
- Brosch M, Selezneva E, Scheich H (2005) Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci* 25:6797–6806.
- Brouwer GJ, van Ee R (2006) Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics. *Vision Res* 46:3393–3402.
- Budinger E, Heil P, Hess A, Scheich H (2006) Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. *Neuroscience* 143:1065–1083.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lip reading. *Science* 276:593–596.
- Carrasco M, Ling S, Read S (2004) Attention alters appearance. *Nat Neurosci* 7:308–313.
- Chong SC, Blake R (2006) Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. *Vision Res* 46:1794–1803.
- Chong SC, Tadin D, Blake R (2005) Endogenous attention prolongs dominance durations in binocular rivalry. *J Vis* 5:1004–1012.
- Degerman A, Rinne T, Pekkola J, Autti T, Jääskeläinen IP, Sams M, Alho K (2007) Human brain activity associated with audiovisual perception and attention. *Neuroimage* 34:1683–1691.
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57:11–23.
- Driver J, Spence C (1998) Attention and crossmodal construction of space. *Trends Cogn Sci* 2:254–262.
- Erkelenz CJ, Regan D (1986) Human ocular vergence movements induced by changing size and disparity. *J Physiol* 379:145–169.
- Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16:419–423.
- Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7:773–778.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285.
- Gilbert CD, Sigman M (2007) Brain states: top-down influences in sensory processing. *Neuron* 54:677–696.
- Gutfreund Y, Zheng W, Knudsen EI (2002) Gated visual input to the central auditory system. *Science* 297:1556–1559.
- Guttman SE, Gilroy LA, Blake R (2005) Hearing what the eyes see. *Psychol Sci* 16:228–235.
- Hancock S, Andrews TJ (2007) The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. *Perception* 36:288–298.
- Helmholtz H (1866) *Handbuch der Physiologischen Optik*. Hamburg: Voss.
- Hol K, Koene A, van Ee R (2003) Attention-biased multi-stable surface perception in three-dimensional structure-from-motion. *J Vis* 3:486–498.
- Hupé J-M, Joffo LM, Pressnitzer D (2008) Bistability for audiovisual stimuli: perceptual decision is modality specific. *J Vis* 8:1.1–15.
- Ichikawa M, Masakura Y (2006) Auditory stimulation affects apparent motion. *Jpn Psychol Res* 48:91–101.
- Kanai R, Tsuchiya N, Verstraten FA (2006) The scope and limits of top-down attention in unconscious visual processing. *Curr Biol* 16:2332–2336.
- Kanai R, Sheth BR, Verstraten FA, Shimojo S (2007) Dynamic perceptual changes in audiovisual simultaneity. *PLoS One* 2:e1253.
- Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain structure and function* 212:121–132.
- Klink PC, van Ee R, Nijs MM, Brouwer GJ, Noest AJ, van Wezel RJA (2008) Early interactions between neuronal adaptation and voluntary control determine perceptual choices in bistable vision. *J Vis* 8:16.1–18.
- Lack LC (1978) *Selective attention and the control of binocular rivalry*. The Hague, The Netherlands: Mouton.
- Lakatos P, Chen CM, O’Connell MN, Mills A, Schroeder CE (2007) Neuro-

- nal Oscillations and multisensory Interaction in primary auditory cortex. *Neuron* 53:279–292.
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110–113.
- Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23:571–579.
- Leopold DA, Logothetis NK (1999) Multistable phenomena: changing views in perception. *Trends Cogn Sci* 3:254–264.
- Levitt WJM (1965) On binocular rivalry. Assen, The Netherlands: Royal van Gorcum.
- Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci* 28:264–271.
- Maier JX, Ghazanfar AA (2007) Looming biases in monkey auditory cortex. *J Neurosci* 27:4093–4100.
- Melcher D, Paphomas TV, Vidnyánszky Z (2005) Implicit attentional selection of bound visual features. *Neuron* 46:723–729.
- Meng M, Tong F (2004) Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J Vis* 4:539–551.
- Mozolic JL, Hugenschmidt CE, Peiffer AM, Laurienti PJ (2008) Modality-specific selective attention attenuates multisensory integration. *Exp Brain Res* 184:39–52.
- Neuhoff JG (2001) An adaptive bias in the perception of looming auditory motion. *Ecol Psychol* 13:87–110.
- Noest AJ, van Ee R, Nijs MM, van Wezel RJA (2007) Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. *J Vis* 7:10.1–14.
- O’Leary A, Rhodes G (1984) Cross-modal effects on visual and auditory object perception. *Percept Psychophys* 35:565–569.
- Paffen CL, Alais D, Verstraten FA (2006) Attention speeds binocular rivalry. *Psychol Sci* 17:752–756.
- Parker AL, Alais D (2006) Auditory modulation of binocular rivalry. *J Vis* 6:855.
- Parker A, Alais D (2007) A bias for looming stimuli to predominate in binocular rivalry. *Vision Res* 47:2661–2674.
- Peterson MA, Hochberg J (1983) Opposed-set measurement procedure: a quantitative analysis of the role of local cues and intention in form perception. *J Exp Psychol: Hum Percept Perform* 9:183–193.
- Pressnitzer D, Hupé JM (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr Biol* 16:1351–1357.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
- Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. *Nature* 385:308.
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401–409.
- Shiffrin RM, Grantham DW (1974) Can attention be allocated to sensory modalities? *Percept Psychophys* 15:460–474.
- Shimojo S, Shams L (2001) Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol* 11:505–509.
- Shinn-Cunningham BG (2008) Object-based auditory and visual attention. *Trends Cogn Sci* 12:182–186.
- Shomstein S, Yantis S (2004) Control of attention shifts between vision and audition in human cortex. *J Neurosci* 24:10702–10706.
- Slotnick SD, Yantis S (2005) Common neural substrates for the control and effects of visual attention and perceptual bistability. *Brain Res Cogn Brain Res* 24:97–108.
- Soto-Faraco S, Morein-Zamir S, Kingstone A (2005) On audiovisual spatial synergy: the fragility of the phenomenon. *Percept Psychophys* 67:444–457.
- Spence C, Nicholls ME, Driver J (2001) The cost of expecting events in the wrong sensory modality. *Percept Psychophys* 63:330–336.
- Suzuki S, Peterson MA (2000) Multiplicative effects of intention on the perception of bistable apparent motion. *Psychol Sci* 11:202–209.
- Tong F (2003) Primary visual cortex and visual awareness. *Nat Rev Neurosci* 4:219–229.
- Toppino TC (2003) Reversible-figure perception: mechanisms of intentional control. *Percept Psychophys* 65:1285–1295.
- Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579.
- Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382:539–541.
- Van der Burg E, Olivers CN, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform* 34:1053–1065.
- van Ee R, van Dam LC, Brouwer GJ (2005) Voluntary control and the dynamics of perceptual bi-stability. *Vision Res* 45:41–55.
- van Ee R, Noest AJ, Brascamp JW, van den Berg AV (2006) Attentional control over either of the two competing percepts of ambiguous stimuli revealed by a two-parameter analysis: means don’t make the difference. *Vision Res* 46:3129–3141.
- van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. PhD thesis. Institute of Perception Research, Eindhoven, The Netherlands.
- Wade NJ, de Weert CM (1986) Binocular rivalry with rotary and radial motions. *Perception* 15:435–442.
- Wannig A, Rodríguez V, Freiwald WA (2007) Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron* 54:639–651.
- Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. *Neuron* 48:489–496.
- Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat Neurosci* 9:1156–1160.