Influences of attention on auditory aftereffects following purely visual adaptation

JI HONG and THOMAS V. PAPATHOMAS*

Laboratory of Vision Research, RuCCS, and Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854, USA

Received 18 February 2006; accepted 13 May 2006

Abstract—Recently, Kitagawa and Ichihara (2002) demonstrated that visual adaptation to an expanding or contracting disk produces a cross-modal visually-induced auditory loudness aftereffect (VALAE), which they attributed to cross-correlations of motion in three-dimensional space. Our experiments extend their results by providing evidence that attending selectively to one of two competing visual stimuli of the same saliency produces a cross-modal VALAE that favors the attended stimulus. These cross-modal attentional effects suggest the existence of integrative spatial mechanisms between vision and audition that are affected by attention.

Keywords: Cross-modal; selective attention; competing stimuli.

INTRODUCTION

There has been a renewal of interest, among psychophysicists and physiologists, in crossmodal influences in perception over the last decade or so (for reviews, see Eimer, 2004; Shimojo and Shams, 2001). This renewal has been fueled by, among other factors, advances in brain imaging techniques that have enabled researchers to investigate neural correlates of crossmodal perception (see Macaluso and Driver, 2003; Spence and Driver 2004, for reviews). These recent studies have confirmed the existence of significant crossmodal interactions, especially between auditory and visual processes in spatial perception. Thus, the classical ventriloquist illusion (Howard and Templeton, 1966), in which the location of a sound source is shifted drastically toward a simultaneously presented congruent visual stimulus, has been studied extensively under a wide variety of conditions (e.g. Alais and Burr, 2004; Bertelson, 1999; Bertelson et al., 2000; Vroomen and De Gelder, 2004).

*To whom correspondence should be addressed. E-mail: papathom@rci.rutgers.edu
Most of the early psychophysical research on crossmodal influences in the perception of visual and auditory stimuli dealt with localization of stationary crossmodal stimuli (e.g. Bertelson and Radeau, 1981; Lovelace and Anderson, 1993), but there have been increasingly more studies with moving crossmodal stimuli (e.g. Ehrenstein and Reinhardt-Rutland, 1996; Lakatos, 1995). These psychophysical experiments were accompanied by corresponding neurophysiological (e.g. Harris \textit{et al.}, 1980; Stein and Wallace, 1996), brain imaging (e.g. Calvert \textit{et al.}, 2001) and event-related potential (e.g. Fort \textit{et al.}, 2002) studies. One of the main findings of these neurophysiological studies was that the superior colliculus (SC) plays an important role in integrating such bimodal signals. In addition, there is evidence that two cortical sites, the anterior ectosylvian and rostral lateral suprasylvian sulci, send signals to the SC that enable it to integrate stimuli from the auditory and visual modalities (for a review, see Stein, 2005).

There have been numerous reports in which the perception of dynamic auditory events is significantly influenced by dynamic visual stimuli. Thus, the McGurk effect (McGurk and MacDonald, 1976) demonstrates that the speech sounds we perceive are influenced by the visual input from the speaker’s mouth; there are several examples where auditory motion is affected by simultaneous visual motion (Kitajima and Yamashita, 1999; Mateeff \textit{et al.}, 1985; Vroomen and de Gelder, 2003) or is modulated as a result of adaptation to visual motion (Ehrenstein and Reinhardt-Rutland, 1996; Kitagawa and Ichihara, 2002). There have also been reports of effects in the reverse direction, in which auditory stimuli influence the processing of visual stimuli, such as the phenomenon of audio-visual motion capture (Meyer and Wuerger, 2001). Regan and Spekreijse (1977) observed that the rate of a clicking sound affected the perceived visual flicker rate. Shams \textit{et al.} (2000, 2002) took this to the extreme: a single light flash is seen as multiple flashes when it is presented simultaneously with multiple sound beeps. Sekuler \textit{et al.} (1997) designed a compelling effect with a visual stimulus where two disks approach each other, then stream through, and recede away from each other; the presence of an auditory click biases the percept in favor of the two disks bouncing off, rather than streaming through, one another.

However, in most of these studies, it is the presence of a strong, unambiguous, stimulus in one modality that affects the perceptual processing of a stimulus in another modality, even in cases where the two stimuli are not present simultaneously. In particular, Kitagawa and Ichihara (2002) showed that adapting to an unambiguous purely visual stimulus, i.e. an expanding or contracting disk, produces not only a visually induced visual motion aftereffect (VVMAE), but also a cross-modal visually induced auditory loudness aftereffect (VALAE); namely, a sound of fixed intensity appears to decrease or increase over time, respectively. The most parsimonious explanation of this phenomenon, which is also supported by evidence from neuroscience (Andersen, 1997; Driver and Spence, 2000; Poirier \textit{et al.}, 2005), is that the neural mechanisms involved in visual motion processing interact with those involved in the processing of auditory motion. The experiments in this study
sought to test whether selective attention to one of two competing balanced visual stimuli during adaptation, which is known to modulate the strength of visual adaptation (Lankheet and Verstraten, 1995), affects the magnitude of cross-modal auditory aftereffects. Thus, this study attempts to add to our understanding of the role of attention in experiments with cross-modal stimuli (Busse et al., 2005; Luo and Wei, 1999; Spence and Driver, 1996, 1997).

**METHODS**

**General**

In all conditions, there were two phases in each trial: an adaptation phase, followed by a test phase, in which auditory aftereffects were measured; subjects reported whether a test tone increased or decreased in intensity over time. Nine subjects participated in the main experiment. They were instructed to fixate their eyes on the fixation mark throughout each trial. All had normal or corrected-to-normal visual acuity, and normal hearing.

The main experiment comprised four conditions to test the strength of the ensuing auditory aftereffect. The two primary experimental conditions that were designed to examine the influence of visual attention on auditory processing are described first: In the first condition, referred to as ‘single-peripheral visual-to-auditory’ and schematized in Fig. 1a, we measured the VALAE after adapting to a single disk in the periphery: a disk of fixed color (red or green) expanded (or contracted) over an interval of 2 s; its position alternated in time between the left and right of the fixation mark, as shown in Fig. 1a, producing a stimulus with a period of 4 s. As expected from Kitagawa and Ichihara’s (2002) study, such adaptation produced a VALAE. In the second condition, ‘double-peripheral visual-to-auditory’, the VALAE was assessed after the following adaptation pattern: there were always two competing disks on both sides of the fixation mark. The red disk on the left expanded, while the green disk on the right contracted for 2 s. In the next cycle, the positions of the two disks were exchanged, with the red disk expanding and the green one contracting, as before. This position alternation was repeated throughout the adaptation phase, producing a stimulus with a period of 4 s. To ensure attentional engagement, in both of the above conditions, the observer was instructed to perform an attentive task during adaptation on the disk that always expanded (or contracted); see the subsection below on ‘Details on stimuli and procedure’.

The key question is whether attention to one of the two competing visual stimuli (expanding or contracting) of Fig. 1b during adaptation would ‘break the balance’ of the visual stimuli and produce a VALAE. Here, we assume that there is a near-zero net VALAE without selective attention to one of the two disks; this assumption was verified in a separate control experiment, described below. We assessed the effect of attention by comparing the magnitude of the VALAE in the two conditions of Fig. 1a and 1b. We next describe the two calibrating conditions. In the third condition,
referred to as ‘central visual-to-auditory’, we obtained a reference value, namely the strength of a VALAE induced by purely visual adaptation in the center of the visual field, as reported by Kitagawa and Ichihara (2002). In the fourth condition, ‘auditory-to-auditory’, we simply measured the aftereffect following adaptation to a purely auditory signal of ascending or descending intensity, to establish a reference value for the intramodal aftereffect. We ought to point out that the main, and only clearly interpretable, comparison is between conditions 1 (spVA) and 2 (dpVA). Indeed, we have no prediction about the role of eccentricity in visual stimulation, nor do we know what particular parameters to use for the ‘auditory-to-auditory’ condition that would allow a comparison of results across these conditions.

Details on stimuli and procedure

All visual stimuli were displayed against a dark background. Red and green settings were set at equiluminance, using a flicker minimization adjustment procedure. All auditory stimuli were presented binaurally through headphones. The order of experimental conditions was randomized for each participant.

The test phase involved a test tone to quantify the aftereffect. The tone’s intensity increased or decreased exponentially for 1.5 s (see Fig. 2); the exponential variation resulted in a linear slope in terms of dB/s. Subjects were asked to press one of two keys, depending on whether they perceived the tone as becoming louder or softer.
Figure 2. Schematic representation of the stimuli for the ‘auditory-to-auditory’ condition. The vertical axis represents the sound intensity, and the horizontal axis represents time. A solid black line denotes the adapting or the test sound. The experimental design was similar in all conditions of the experiment.

During the 1.5 s, using a single-interval binary-choice procedure. We quantified the aftereffect in the same manner as Kitagawa and Ichihara (2002), by assessing the ‘nulling slope’, i.e. the bias slope of the test tone that produced the percept of a flat tone, as follows: the slope of the test tone in the next trial was (algebraically) increased or decreased, depending on whether subjects reported the test tone in the current trial as becoming softer or louder, respectively. A dual staircase procedure controlled the decision to use increments/decrements of 1 dB/s. Each staircase was terminated after 10 reversals, and the nulling slope was estimated as the average of the last 6 reversals.

Each adaptation phase consisted of several repetitions of a basic adaptation pattern, each lasting for 2 s, as shown schematically in Fig. 2, which illustrates the adaptation stimulus for condition 3 (auditory-to-auditory). The adaptation phase for central visual stimuli, as well as for auditory stimuli, started with an initial adaptation of 120 s in the first trial, and continued with 12-s top-up adaptations in subsequent trials; for peripheral stimuli we increased the initial adaptation to 180 s and the top-up adaptations to 20 s, because the physical or attended stimuli were present for much less time at a particular peripheral location.

Details for each adaptation phase are described for each condition below: In conditions 1 and 2, the centers of the disks were 3.75 degrees to the right or to the left of the fixation mark. In conditions 1 to 3, disks expanded or contracted at a rate of 1°/s between extreme diameters of 1° and 3° of visual angle. In condition 4, the basic 2-s adaptation pattern was a 1-kHz pure sinusoid whose intensity ascended (or descended), with a slope of 20 dB/s, between 20 dB and 60 dB sound pressure.
level. We also tried a control condition, using eight additional subjects, with the same stimuli as in condition 2 (double-peripheral visual-to-auditory), but with no selective attention during adaptation; subjects were instructed to fixate without attending to either of the two disks. This control condition was used to ensure that there is no systematic auditory aftereffect when the visual stimuli are balanced during adaptation.

To ensure that observers attended to the expanding (or contracting) disk, we included brief ‘deformation events’ at random instances during adaptation, once (most often) or twice (rarely) every 2 seconds: each disk was transformed into an ellipse, and back into a circle over 200 ms. Subjects were asked to discriminate between transformations into a ‘vertical’ or ‘horizontal’ ellipse. The ellipse aspect ratio was selected so that subjects could perform the task correctly about 75% of the time; the responses of all the trials, whether correct or wrong, were included in ‘driving’ the staircase. This attentive task was used in conditions 1 and 2.

We also conducted an intra-modal experiment to quantify the effect of selective visual attention of the visual aftereffect (VAE) following visual adaptation. The adaptation stimuli and procedure was identical to that of condition 2 of the main experiment. Assessing the VAE involved a test disk that was presented for 0.5 s. Its initial diameter of 2 degrees was increased or decreased at a rate that was determined by a staircase procedure. Subjects were asked to report whether they perceived the disk as expanding or contracting, using a single-interval binary-choice procedure. We quantified the VVMAE by assessing the ‘nulling rate’, i.e. the bias rate of expansion (or contraction) that produced the percept of a fixed-sized disk during the test interval. A dual staircase procedure adjusted the rate of diameter change by a step size of 0.04 degrees/s. Each staircase was terminated after 10 reversals, and the nulling slope was estimated as the average of the last 6 reversals. The center of the test disk in conditions 2 and 3 was placed either to the left or to the right of the fixation mark, chosen randomly, at the same eccentricity as the adapting disks.

RESULTS

Figure 3 shows the strength of the aftereffect in the various conditions, with the results averaged across all 9 subjects. All subjects participated in three sessions, each involving two staircases. Each bar in Fig. 3 was obtained by averaging the differences of two nulling slopes for the corresponding test condition: this involved subtracting the nulling slope following adapting (or attending) to the decreasing stimulus from the nulling slope following adaptation (or attending) to the increasing stimulus. As a consequence, in all four conditions, this operation doubles the net effect, i.e. the difference between the nulling slopes with and without adaptation.

As can be seen from Fig. 3, condition 4 (AA, auditory-to-auditory) produced the largest aftereffect, as expected, with a mean value of 3.95 dB/s. The aftereffect strength for the central disk of condition 3 (cVA) is significantly smaller ($t(8) = 3.32, p < 0.05$), as also reported by Kitagawa and Ichihara (2002), and it gets even
smaller for the peripheral disk of condition 1 (spVA). Nevertheless, the strength is significant ($t(8) = 3.12$, $p < 0.05$) for the attentive condition 2 (dpVA). Importantly, every one of the eight subjects who exhibited a larger aftereffect with an expanding, as compared to a contracting, unambiguous peripheral disk in condition 1 (spVA, single-peripheral visual-to-auditory), also exhibited a larger aftereffect when attending to the expanding, as compared to a contracting disk, in the ‘balanced’ visual stimulus of condition 2 (dpVA, double-peripheral visual-to-auditory). Only one subject exhibited an unexpected reverse effect in condition 1 (single peripheral disk), even though he exhibited an expected effect in condition 4 (auditory-to-auditory); however, this subject also produced the reverse effect in condition 3 (central disk), and no significant effect in condition 2 (attention); we have no explanation for the behavior of this unique subject.

There was no systematic pattern in the auditory aftereffect for the control condition, in which there were two competing disks during adaptation, one expanding and the other contracting, exactly as in condition 2, but subjects did not selectively attend to either. The average magnitude of the VALAE was $-0.043$ dB/s, with a standard error of 0.168. A $t$-test on the mean yielded $t(7) = -0.253$, $p = 0.81$, two-tailed, leading to the conclusion that the mean of the population is zero, as expected.

To quantify the effect in the intra-modal experiment, we computed the difference of two nulling slopes for the corresponding test condition: namely, we subtracted the nulling slope following attention to the contracting stimulus from the nulling slope following attention to the expanding stimulus during adaptation. The average of this difference for five subjects was 0.218 deg/s, with a standard error of 0.033 deg/s.
These results confirmed that selective attention to one of two competing disks produced a visual aftereffect in favor of the attended disk.

Observers were instructed to keep their eyes on the fixation mark throughout each trial. To exclude the possibility that subjects failed to maintain fixation during the experiments, we used the RK-726 PCI Pupil/Corneal Reflection Tracking System of ISCAN, Inc. to record real-time x-y eye coordinates every 16.7 ms. We repeated the adaptation part of conditions 1 to 3 with three observers. We verified that observers followed fixation instructions. Their fixation gaze was within 1 degree of the fixation mark for more than 94% of the time. Due to blinking, there were large excursions (mean value 2.2 degrees) away from fixation for about 2% of the time, but these were clustered systematically in the upper hemifield, and never approached any of the two disk positions for conditions 1 and 2. These considerations lead us to exclude eye movements as a source of the observed behavior.

CONCLUSIONS–DISCUSSION

One possible explanation for the influence of the visual adaptation on auditory processing can be based on the correlation of the two cross-modal signals for an approaching (or receding) object that is both visible and audible. For such an object, the retinal image size and the sound intensity at the ear increase congruently. Sub-cortical mechanisms, such as the SC, respond to the simultaneous presence of the visual and auditory signals (e.g. Stanford et al., 2005; see Stein, 1998, 2005 for reviews). One possibility is that such mechanisms may be activated by signals from the visual modality alone, which would account for such a transfer aftereffect. Another possibility is that these mechanisms may involve higher cortical areas that are also responsive to the simultaneous presence of auditory and visual stimuli (Andersen, 1997; Stein, 2005). The results of Kitagawa and Ichihara (2002) provide evidence that purely visual signals feed into cross-modal mechanisms, but they cannot specify the processing stage of such mechanisms.

A key question in this study is whether the modulation of the strength of visual adaptation by attention can affect the perception of subsequently presented auditory stimuli. This cross-modal influence was indeed observed in condition 2: attending to one of two simultaneously present competing stimuli in the visual domain, produced a VALAE in the auditory domain. The magnitude of this crossmodal VALAE was statistically significant; remarkably, it was of comparable magnitude to the VALAE produced by a single disk located at the same eccentricity. Such crossmodal influences may be expected in light of recent evidence for the modulation of activity in brain areas that had been hitherto regarded as unimodal areas (Bremmer et al., 2001; Calvert et al., 1997; Graziano, 2001; Kaas and Collins, 2004; Lewis et al., 2000) most probably because of feedback from multimodal areas (Calvert et al., 2000; Driver and Spence, 2000).

One may consider the possibility that these results may be due to some kind of multimodal feature-based attentional effect. In such effects, attending to a visual
feature (a particular color, motion direction, orientation, etc.) results in enhanced activity of neurons that are tuned to the attended feature throughout the visual field (Corbetta et al., 1991; McAdams and Maunsell, 2000; Motter, 1994; Saenz et al., 2002; Treue and Martinez-Trujillo, 1999). Recently, Melcher et al. (2005) have obtained evidence that this enhancement extends not only to the attended feature, but also to bound clusters of features that are associated with the attended feature, an effect that they termed ‘implicit cross-attribute attentional selection’. Thus, if one extends the implicit attentional selection hypothesis to cross-modal domains, one may reason that the observed effects can be partly explained as follows: attending to an increasing-size stimulus in the visual domain results in enhanced activity of neurons tuned to increasing-intensity stimuli in the auditory domain as well.

Lavie’s theory of attention (Lavie, 1995, 2005) posits that the extent of automatic processing of distractors — task-irrelevant stimuli — depends on the attentional load (Rees et al., 1997). Parenthetically, recent experiments indicate that high perceptual load reduces distractor processing, whereas cognitive control load, such as working memory load, enhances distractor processing (Lavie et al., 2004; Lavie, 2005). More relevant to the present study, Lavie and colleagues (Rees et al., 2001) found that attending to auditory stimuli during adaptation does not influence the magnitude of visual aftereffects, independently of the attentional load, leading them to conclude that attentional capacity is restricted within, but not between, sensory modalities. However, the absence of cross-modal effects in Rees et al. (2001) can be explained in light of Kitagawa and Ichihara’s (2002) finding that auditory stimuli did not affect visual motion aftereffects.

The issue of whether audition and vision share attentional resources or have entirely separate intra-modal attentional resources is still under debate. There is experimental evidence for separate attentional resources for vision and audition (Duncan et al., 1997; Rees et al., 1997, 2001). Of particular interest is a recent study by Alais et al. (2006), which offers evidence that there is no cost for two attentional tasks in separate modalities (audition and vision), whereas a price is paid for two tasks in the same modality. On the other hand, Eimer and Driver (2001) studied cross-modal links in spatial attention, and Beer and Roder’s (2004, 2005) experiments offered recent evidence for shared attentional resources. Based on Kitagawa and Ichihara’s (2002) and our own experiments, an ascending (descending) sound appears to be associated with an expanding (contracting) disk. Thus, it is reasonable to assume the existence of facilitatory connections originating from visual neurons tuned to expansion (or contraction) and terminating to auditory neurons tuned to an increasing-intensity (or decreasing-intensity) sound. Our experimental design, however, cannot distinguish whether attentional resources are shared between the visual and auditory modalities, or whether each modality has its own dedicated attentional resources.
Acknowledgements

The authors are indebted to Zoltán Vidnyánszky for advice on all aspects of the project, as well as with significant help with the manuscript. We thank Drs. R. Kashi, M. Sondhi, and J. Hall for valuable suggestions and technical support for the auditory stimuli, as well as Mr. Anshul Jain for his help in administering the control experiment. We also thank two anonymous reviewers, the US editor-in-chief and the action editor for valuable feedback, and the two reviewers for suggesting the control experiment. This research was supported by a grant from the NEI/NIH (EY 013758-01) to T. V. P.

REFERENCES


