Sound alters visual evoked potentials in humans

Ladan Shams,[CA] Yukiyasu Kamitani, Samuel Thompson and Shinsuke Shimojo

California Institute of Technology, Division of Biology BI 139-74, Pasadena, CA 91125, USA

[CA]Corresponding Author

Received 5 September 2001; accepted 26 September 2001

When a single flash is accompanied by two auditory beeps, the single flash is perceived as two flashes. We investigated whether this crossmodal influence on visual perception occurs at the level of the modality-specific visual pathway or later. We compared the visual evoked potentials (VEPs) in the presence and absence of sound. Activity was modulated extensively and with short latency in trials in which an illusory flash was perceived. In addition, the brain potentials for the illusory flash were qualitatively very similar to those for a physical flash, suggesting that the same mechanism underlies the percept of both illusory and physical flashes. These results suggest that the activity in the visual cortex can be modulated by sound. This implication challenges the general belief that the visual cortical processing is independent of other modalities. *NeuroReport* 12:3849–3852 © 2001 Lippincott Williams & Wilkins.

Key words: Auditory–visual; Brain potentials; Crossmodal illusion; Crossmodal interactions; Event-related potential; Illusory flash effect; Multisensory integration; Visual evoked potential; Visual illusion

INTRODUCTION

It has recently been shown that when a single flash is accompanied by two auditory beeps, the single flash is perceived as two flashes [1]. This illusion illustrates that sound can radically alter visual perception. The illusory flash effect is much stronger in the periphery than in the fovea. The stimulus configuration and task involved in the illusion are both very simple. Furthermore, the illusion is highly robust with respect to a number of parameters. We have observed that manipulation of many parameters does not eliminate the illusion, at most makes it degrade gracefully [2]. This degree of simplicity and robustness suggests that the illusory flash phenomenon reflects a mainstream mechanism in the brain, as opposed to an accidental or marginal neuronal activity. Several other psychophysical studies have also provided evidence for modulation of visual perception by sound (for review see [3]). However, knowledge of the effects of auditory stimulation on visual perception still remains largely within the realm of phenomenology, and the underlying mechanisms are not extensively studied nor understood. It is not clear at what level of perceptual processing these crossmodal effects take place. These interactions may occur at early/late visual areas, or at polysensory associative cortical areas.

Event-related potential (ERP) recording provides an appropriate methodology for tackling this question due to its high temporal resolution. Schröger and Widdman [4] used ERP to explore the site of audio-visual interactions. They employed an odd-ball paradigm, and found no early interactions between the auditory and visual processes. They interpreted their results as suggesting that the audio-visual integration occurs somewhere beyond the modality-specific areas but before the decision-making stage. They pointed out, however, that the reason for the lack of evidence for early modulation in their study may be due to the fact that their task relies on memory mechanisms and thus may not be appropriate for uncovering early sensory interactions. Giard and Perronet [5] used ERP for tackling the same question employing a pattern recognition task. They reported very early crossmodal effects in the occipital area, and interpreted these results as modulation of activity in the modality-specific visual cortical areas by sound. In their study, however, they used two visual deformation patterns which unfolded over a course of 230 ms, and the subjects were trained in advance to associate each of the two visual patterns with a specific tone. It is not clear whether their results generalize to situations in which subjects are not trained to associate specific visual stimuli with specific auditory stimuli, or where the visual stimulus is a static image as opposed a deforming pattern. We recorded ERPs in a framework based on the illusory flash effect in order to examine the locus of alterations of visual perception by sound. Unlike the two studies mentioned above, the task used in our study was a very simple perceptual task not involving memory. More importantly, the subjects were not instructed *a priori* to associate a certain visual stimulus with a certain auditory stimulus. The stimuli were extremely simple – brief tones and flashes.

METHODS AND MATERIALS

We employed a flash VEP paradigm and introduced a sound stimulus to examine whether sound would modulate the VEPs. Illusory flash effect was used as the basic framework, i.e. a single flash was paired with two brief beeps leading to the percept of two flashes (or illusory double-flash). Our psychophysical observation showed that the illusion is significantly stronger in the periphery than in the fovea. In order to search for any physiological correlation with this perceptual effect, we recorded VEPs for flashes presented in the fovea and the periphery separately.

The experiment consisted of 6 conditions: V_p: a flash in the periphery, AV_p: a flash in the periphery accompanied with two beeps, V_f: a flash in the fovea, AV_f: a flash in the fovea accompanied with two beeps, A: two beeps (and no flashes), and V_{p^2}: two flashes in the periphery.

The flashing stimulus was a uniform white disk subtending a visual angle of 2°, displayed in the fovea (0° eccentricity) or in the periphery at 8° eccentricity for 14 ms. The auditory stimulus consisted of two brief beeps each lasting for 8 ms and separated by 57 ms (stimulus onset asynchrony, SOA). The sound stimulus (3.5 kHz frequency at 77 dB SPL) was presented from two speakers symmetrically placed adjacent to the two sides of the computer screen on which the visual stimulus was presented. The height of the speakers was between the foveal and the peripheral positions of the visual stimulus (i.e. 4° below the fixation point). In the bimodal condition, the flash onset was between those of the two beeps, i.e. 14 ms after the onset of the first beep. Because the control condition V_{p^2} (the physical double flash) was meant to be contrasted against the illusory double flash condition AV_p, we calibrated the timing between the two flashes such that it perceptually matched the percept of the illusory double flash. The SOA of the two flashes was thus set to 67 ms. There were 100 trials for each condition and the order of trials was random. Data was collected from 13 participants (ages 17–45 years, normal or corrected to normal vision, five females). The participant's task was to judge the number of flashes they saw on the screen at the end of each trial in a three-alternative forced-choice paradigm – zero, one or two flashes. They responded by pressing keys.

VEPs were recorded using three Ag–AgCl electrodes placed in the occipital area at O_z, O_1 and O_2 (according to the international 10-20 system) referred to the nose. Eye blinks and vertical eye movements were monitored using two electrodes above and below the right eye. ERP and eye movement signals were recorded at epochs of −100 to 360 ms relative to the onset of the (first) flash, filtered with a bandpass of 0.5–100 Hz, and digitized at 1 kHz. Epochs containing eye movement artifacts were rejected. The epoched data were averaged and filtered digitally with a low pass cut-off frequency at 50 Hz. To assess possible auditory–visual interactions, the difference wave [AV−(A + V)] obtained by subtracting the sum of responses to the unimodal stimuli from the response to the bimodal stimuli was calculated. This is a common method of measuring auditoryvisual interactions using ERP [4–6]. The logic behind this analysis is that if there are no interactions between the audio and visual processes, the activity in the bimodal condition should be equal to the sum of the activities in the uni-modal conditions, and therefore the amplitude of the difference wave should be equal to zero.

RESULTS

On average, observers reported seeing two flashes in 81% of the AV_p trials, and in only 21% of the AV_f trials. These results confirm our previous observation that the illusory flash effect is stronger in the periphery than in the fovea. Observers correctly reported seeing two flashes in the V_{p^2} condition in 92% of the trials. We analyzed the ERP data from selected trials based on the behavioral response in order to explore correlation between the ERPs and the perceptual phenomenology. In the AV_p condition, we only retained the illusion trials where two flashes were seen (the majority of trials); in the AV_f condition we only retained the non-illusion trials where one flash was seen (again the majority of trials, 79%), and in the V_{p^2} condition we only retained the correctly perceived trials (two flashes perceived, again the majority of trials).

Grand average ERPs were calculated over all trials and all participants for each condition and each electrodes separately. The grand average waveforms of all the conditions in which there was a visual stimulus (i.e., V_p, AV_p, V_f, AV_f, and V_{p^2}) had the classic flash VEP morphology. In condition A, in which there was no visual stimulus, the grand average response was nearly zero at all latencies, confirming that the responses measured from the three occipital electrodes are primarily due to activity in the visual areas (and hence referred to as VEPs).

Figure 1a,b shows the difference wave [AV−(A + V)] averaged across subjects, calculated for the fovea and periphery conditions, respectively. Each column corresponds to one of the three electrodes. To determine significant effects, the amplitude of the difference wave was compared with zero using *t*-test for each time sample at each electrode across subjects. The time intervals which were significantly different from zero ($p < 0.05$) for $\geqslant 15$ consecutive time samples (15 ms) were considered as stable interaction time intervals [7]. The black blocks in Fig. 1 represent these time intervals. As can be seen in Fig. 1a, there are no interactions in the fovea. (Even though the mean amplitude is fairly large at some latencies, because of the large variance across subjects these amplitudes are not significantly different from zero, $p > 0.05$.) In contrast, the difference waves displayed in Fig. 1b show evidence for extensive and early interactions in the periphery. There is a large contiguous interval between 174 and 200 ms post-stimulus, and another large interval between 262 ms and 360 ms post-stimulus of significant effects. It should be noted that the data shown in Fig. 1a,b correspond to trials in which the observers did not experience the illusion and did experience the illusion, respectively. These results, therefore, provide a neurophysiological correlate to the percept of the illusion. We were not able to do a reliable analysis of the AV_p trials in which illusion was not perceived, or AV_f trials in which illusion was perceived, because these trials constituted a small minority of trials and thus, the signal-to-noise ratio was not sufficiently high. However, when we included the non-illusion AV_p trials in the analysis of AV_p condition the significant effects shown in Fig. 1b deteriorated significantly. This indicates that the interaction effects in the ERP data are indeed correlated with the perception of the illusion.

The difference waves shown in Fig. 1b can also be interpreted as activity corresponding to the percept of an

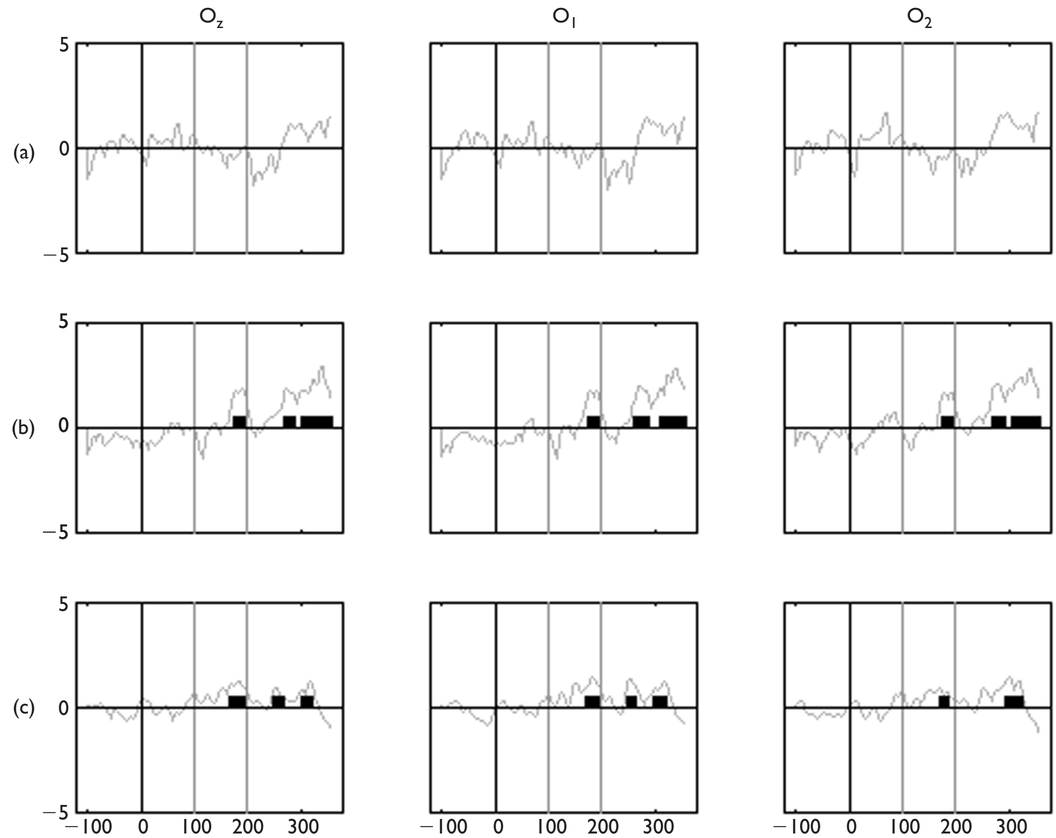

Fig. 1. Auditory–visual interactions reflected in the ERP difference waves. Each row in this figure corresponds to a type of difference waves, and each column corresponds to one electrode. The horizontal and vertical axes denote time in milliseconds with respect to the onset of the (first) flash, and brain potential in μV, respectively. The gray lines represent the mean amplitudes across participants, and the black blocks denote the time intervals in which the amplitudes are significantly different from zero ($p < 0.05$). The waveforms corresponding to $[AV_f-(A+V_f]$ (**a**), $[AV_p-(A+V_p)]$ (**b**), and $[V_{p^2}-V_p]$ (**c**) are plotted for each electrode separately.

illusory second flash. By taking away the activities corresponding to the percept of the two beep (condition A) and the percept of the single (physical) flash (condition V_p) from the activity corresponding to the percept of the illusory double flash and the double beeps (condition AV_p), what remains is the activity corresponding to the percept of the illusory second flash. On the other hand, difference wave $[V_{p^2}-V_p]$ can be interpreted as the activity corresponding to the percept of the physical second flash. Figure 1c illustrates this waveform. The results are qualitatively very similar to those of $[AV_p-(A+V_p)]$ (Fig. 1b). Here we also obtain a contiguous interval of significant activity (amplitude significantly different from zero ($p < 0.05$) for $\geqslant 15$ consecutive time samples) between 170 and 201 ms and another interval of significant activity between 249 and 325 ms post-stimulus. In addition to the timing and duration of these intervals, the morphology of the waves is also qualitatively similar. In both difference waves, the intervals 170–200 ms and 250–350 ms contain a positive peak and dual positive peaks, respectively.

DISCUSSION

The data presented here indicate extensive and early modulation of VEP by sound in the illusion trials, in contrast to lack of modulation of VEP by sound in the non-illusion trials. These results provide a neurophysiological correlate for the perception of the illusory flash.

Modulations of VEP by sound occurred as early as 170 ms post-stimulus. Considering that ERPs prior to 200 ms post-stimulus are believed to be due to the activity in the modality-specific pathways [5], these modulations appear to occur in the visual pathway. Most interesting, however, is the finding that similar modulations were induced by sound and by an additional physical flash. The comparison of the difference waves displayed in Fig. 1b,c revealed a striking similarity between the activity associated with an illusory second flash and that of a physical second flash. This similarity suggests that similar brain mechanisms underlie the processing of these two percepts. Because evoked response to a physical flash involves activity in visual cortex, this implies that the representation of the illusory flash also involves the activity in the visual cortex. It may appear that the onset of activity associated with a physical second flash (Fig. 1c) starting at 170 ms is somewhat late. It should be pointed out that this waveform is plotted with respect to the onset of the first physical flash which precedes the second flash by 67 ms. Therefore, the latency of the onset activity associated with the second flash with respect to the onset of the second flash would be 103 ms post-stimulus. The same reasoning also applies to the onset timing of the illusory second flash; the illusory flash does occur (perceptually) with a delay with respect to the onset of the first flash.

CONCLUSION

These results taken altogether suggest that the activity along the visual cortex can be modulated by the auditory stimulation. Therefore, the multisensory integration already seems to be at work at the level of modality-specific areas. These findings counter the traditional view that the sensory modalities operate independently of each other and the convergence of information does not occur until very late stages of perceptual processing.

REFERENCES

1. Shams L, Kamitani Y and Shimojo S. *Nature* **408**, 788 (2000).
2. Shams L, Kamitani Y and Shimojo S. *Cogn Brain Res*, in press.
3. Shimojo S and Shams L. *Curr Opin Neurobiol* **11**, 505–509 (2000).
4. Schröger E and Widmann A. *Psychophysiology* **35**, 755–759 (1998).
5. Giard MH and Peronnet F. *J Cogn Neurosci* **11**, 473–490 (1999).
6. Barth DS, Goldberg N, Brett B *et al. Brain Res* **678**, 177–190 (1995).
7. Rugg MD, Doyle MC, and Wells T. *J Cogn Neurosci* **7**, 209–227 (1995).