Sound-induced illusory flash perception: role of gamma band responses

Joydeep Bhattacharya,[1,2,CA] Ladan Shams[1] and Shinsuke Shimojo[1,3]

[1]Division of Biology, Mail Code 139-74, California Institute of Technology, Pasadena, CA 91125, USA; [2]Commission for Scientific Visualization, Austrian Academy of Sciences, Vienna, A-1220, Austria; [3]NTT Communication Science Laboratories, Human and Information Science Laboratory, Atsugi, Kanagawa, 243-0198, Japan

[CA]Corresponding Author: joydeep@its.caltech.edu

Received 21 May 2002; accepted 9 July 2002

In the recently discovered sound-induced illusory flash phenomenon, a single flash accompanied with two auditory beeps is perceived as two flashes in a majority of trials. Here we asked what the neural substrates distinguishing illusion and no-illusion (i.e. perception of single flash) percepts are under identical stimulus configuration. Wavelet based method was used to analyze γ band (> 30 Hz) responses in the event-related potential (ERP) signals recorded over visual cortical regions. We found: (i) significantly higher oscillatory and induced γ band responses in illusion than in no-illusion trials, and (ii) significant supra-additive audio-visual interactions only in illusion trials. These results provide a clear neurophysiological correlate to the perception of illusion. Furthermore, the results suggest that auditory stimuli modulate cortical processing of visual stimuli, and the flash illusion (qualitative alteration of visual percept) only takes place when this modulation exceeds some critical threshold for the registration of conscious awareness. *NeuroReport* 13:1727–1730 © 2002 Lippincott Williams & Wilkins.

Key words: Cross-modal interaction; Gamma band; Illusion; Oscillation; Wavelet

INTRODUCTION

Information carried by different sensory channels is effortlessly integrated within the human brain [1–3], although the underlying mechanisms are yet to be fully understood. In addition to the enhancement of saliency of one modality by the concurrent activity of another modality [4–6], the subjective experience within one modality can also be dramatically altered by another modality, thus leading to an illusion. Such a radical cross-modal alteration of the phenomenological quality of the percept has recently been reported in visual perception [7], even though vision has been traditionally considered to be the dominant modality [8–10]. Shams *et al.* [7] showed that when a single visual flash is accompanied by multiple auditory beeps, the single flash is perceived as multiple flashes. This phenomenon has been termed 'sound-induced illusory flash effect'. To study the neural substrates of this cross-modal influence on visual perception, ERP signals were recorded in a previous study [11] from visual cortical regions of human subjects using the sound-induced illusory flash paradigm. Through traditional averaging method, it was found that the ERP profile for the illusory flash is highly similar to that of a physical flash, suggesting a common underlying mechanism. This previous study also compared the auditory–visual interactions between the condition where the visual stimulus was in the periphery (when the illusion occurs strongly) and the condition where it was in the fovea (when the illusion occurs only occasionally). No significant auditory–visual interaction was found in the foveal no-illusion trials, whereas significant interactions were found in the peripheral illusion trials. However, there were 25% of trials in the periphery (on average across subjects) in which the observers did not perceive the illusion. No direct comparison between illusion and no-illusion trials in identical stimulus configuration (in the periphery) was made in the earlier study.

The primary aim of the present study was to investigate possible neural correlation between high frequency gamma (γ) band responses in visual cortical regions and the perceptual phenomenology (illusion or no-illusion) under identical stimulus configuration (one flash in the periphery accompanied by two beeps). To this end, we applied wavelet-based time-frequency analyses to the ERP data of the earlier study [11]. Time frequency analysis is suitable for the characterization of evoked (phase-locked to stimulus) as well as induced (non-phase locked) high-frequency γ band responses [12,13]. γ band was chosen due to widespread evidence of the importance of neuronal oscillations in γ band in feature integration or object recognition including perception of illusory contours [13–15], selective attention [16], associative learning [17], lexical processing [18], and other perceptual functions [19].

MATERIALS AND METHODS

The ERP experiment has been reported previously [11]; however, here we describe the relevant aspects of the experiment for the sake of completeness. Sixteen subjects

0959-4965 © Lippincott Williams & Wilkins

(six females, age range 17–45 years; 13 subjects previously studied in [11] and three additional subjects), with normal or corrected-to-normal vision and with normal hearing ability, participated in the study. The experiment consisted of six stimulus conditions, however, here, we only considered the following three conditions which were relevant for our study: (i) two auditory beeps (A), (ii) a visual flash in the periphery (V_p), and (iii) a flash in the periphery accompanied by two beeps (AV_p). The auditory beep (frequency 3.5 kHz with 77 dB sound pressure level) was presented by two loudspeakers placed symmetrically. The duration of each auditory beep was 8 ms and the beeps were separated by 57 ms. The visual flash was a uniformly illuminated white disk presented on a computer monitor at a visual angle of 2° with 8° eccentricity and duration of 14 ms. In the AV_p condition, the visual stimulus was presented 14 ms after the onset of first auditory beep. There were 100 trials for each condition and the order of the trials was interleaved and randomized. The total time duration of each trial was 600 ms including the pre-stimulus interval of 100 ms. The task was to report the number of flashes perceived.

Signals were recorded from three occipital electrodes (O1, Oz and O2) placed according to the International 10–20 placement system with reference electrode placed at nose. Eye movements were monitored using two electrodes located above and below the right eye. The sampling frequency was 1 kHz. The signals were further digitally high pass filtered with cut-off at 100 Hz followed by a notch filter of 60 Hz. Out of 1600 (16 subjects × 100 trials for each subject) trials, 75% of the AV_p trials produced illusion.

The illusion trials were those trials in AV_p configuration when the subjects reported the perception of two flashes despite the presence of a single physical visual flash. Remaining trials in AV_p were termed as no-illusion trials and both of these trials (illusion and no-illusion) were subject to comparison (with their respective baselines or with summed unimodal responses) followed by the time-frequency analysis.

For the estimation of time-frequency energy variations, the signal $s(t)$ was convolved with complex Morlet wavelets $w(t,f_c)$ which have a Gaussian shape in both time (s.d. σ_t) and in frequency domain (s.d. σ_f) around f_c:

$$w(t,f_c) = \frac{1}{\sqrt{\sigma_t\sqrt{\pi}}} exp\left(-\frac{t^2}{2\sigma_t^2}\right) exp(i2\pi f_c t) \tag{1}$$

The wavelet family is characterized by a constant ratio $f_c/\sigma_f = 7$, with $\sigma_f = 1/(2\pi\sigma_t)$, as has previously been used [12]. The center frequency varies from 20 Hz to 60 Hz in 0.5 Hz steps. The time-varying energy of the signal, $E(t,f_c)$, is defined as follows:

$$E(t,f_c) = \left|w(t,f_c) \otimes s(t)\right|^2 \tag{2}$$

where $\otimes$ represents the convolution operator. The stimulus locked responses can be evaluated by averaging the following quantity across trials:

$$N_k(t,f_c) = \frac{w(t,f_c) \otimes s_k(t)}{\left|w(t,f_c) \otimes s_k(t)\right|} \tag{3}$$

where k denotes the trial number.

The modulus of this quantity, termed event-related inter-trial coherence [12], varies from 0 (non-locked to stimulus or event) to 1 (strictly locked to stimulus). Using eqn (2) and (3), phase-locked as well as non-phase locked responses can be analyzed. The time varying energy of each configuration was always corrected by their individual baselines or pre-stimulus interval (100 ms).

RESULTS

Figure 1a,b shows the baseline-subtracted grand-averaged time frequency representations (TFR) for illusion and no-illusion trials, respectively, in identical stimulus configuration, AV_p (a flash in the periphery accompanied by two beeps). There are higher overall γ-band (> 30 Hz) responses (compared with their individual pre-stimulus interval or baseline) in illusion than in no-illusion trials. In addition, the following observations can be made: (i) The strong gamma (centered at frequency 39 Hz, spanning 28–50 Hz) responses started as early as 30 ms and lasted until 140 ms during illusion trials, whereas in no-illusion trials strong responses were found primarily in beta range (< 25 Hz) and also at center frequency of 30 Hz with latency 90 ms. (ii) Late but strong gamma responses (center frequency 42 Hz) were found in illusory trials which lasted from 420 ms to 480 ms. (iii) γ-band responses seem to be distinct from the lower frequency components (< 25 Hz) because TF energy decreases at ~ 24 Hz and increases for > 30 Hz in illusion trials.

Since the responses as represented by TFR can be due to phase locked (with respect to stimulus) and/or non-phase locked component, event-related inter-trial coherence (eqn (3)) was computed for both aforementioned conditions and the results are shown in Fig. 1c,d, respectively. It is clear that the early responses (through 100 ms) are phase-locked to the stimulus (thus similar between the illusion and no-illusion trials) but the late responses (> 400 ms) of illusion trials do not appear in the profiles of event-related coherence; they are, thus, non-phase locked to the stimulus onset and standard averaging procedure is inappropriate to uncover such responses in which there are variations in latency across trials [20].

Next, we performed non-parametric Wilcoxon signed-rank test on the TFR of the energy averaged across single trials at electrode O1 for each point in joint time-frequency space. Significant responses ($p < 0.05$), larger than baseline, are plotted as probability maps in Fig. 2a,b for illusion and no-illusion trials, respectively. The other two (occipital) electrodes have qualitatively similar profiles. As expected from Fig. 1a, two clusters of significantly enhanced γ-band responses are formed in illusory AV_p trials as compared to baseline. The effect is found for frequencies > 30 Hz. In contrast, little enhancement of γ-band responses is found for non-illusory AV_p trials compared with their baseline. No effect ($p > 0.7$) was found between the baseline levels corresponding to illusion and no-illusion trials.

In order to study the cross-modal (i.e. audio-visual) interaction, the standard approach is to investigate the difference between the bimodal responses and the absolute sum of unimodal responses (i.e. $AV_p-(A+V_p)$) and interaction is reported at any latency at which the difference waveform is significantly different from zero. Although this approach has been routinely adopted [11,21,22], a more accurate method [23] is to analyze the difference

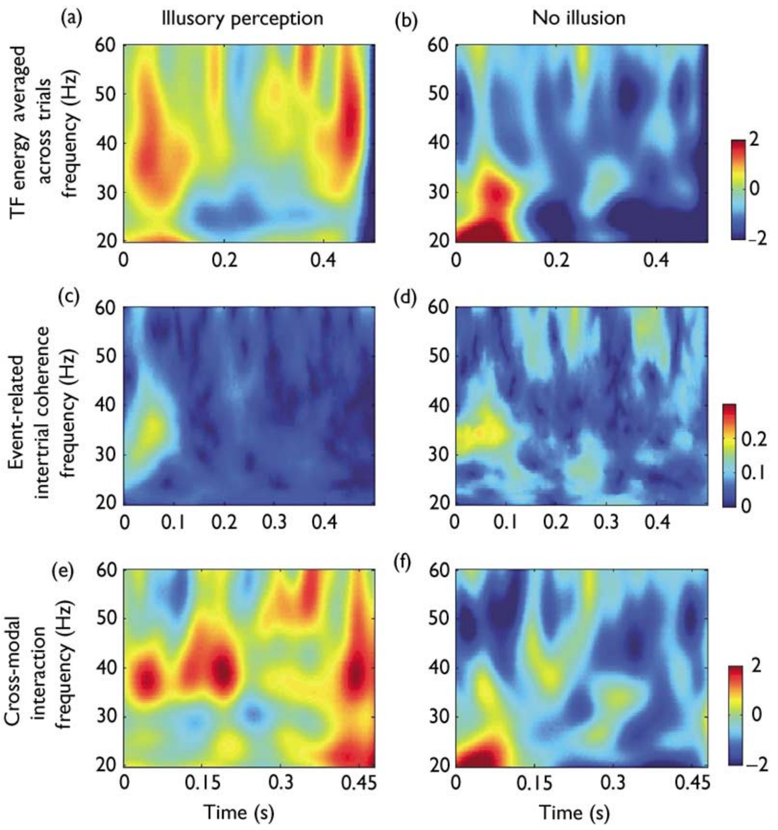

Fig. 1. (**a**,**b**) Time-frequency (TF) representations of the energy during illusory perception, and non-illusory perception, respectively. Results, corrected with respect to 100 ms baseline or pre-stimulus interval, were averaged across trials, all subjects ($n = 16$), and over three occipital electrodes. The onset of visual stimulus is at time zero. (**c**,**d**) Event-related inter-trial coherence or phase-locking factor (averaged across electrodes and subjects). (**e**–**f**) Cross-modal interaction as computed by $[AV_p - \text{baseline}] - ([A - \text{baseline}] + [V_p - \text{baseline}])$ for both illusion and non-illusion trials, respectively.

$[AV_p - \text{baseline}(AV_p)] - ([A - \text{baseline}(A)] + [V_p - \text{baseline}(V_p)])$. This is considered better because statistical interaction effects between two factors point out those changes which would happen when two factors are simultaneously altered which cannot be predicted from the results of changing one factor separately. Figure 1e–f shows the cross-modal interactions for both illusion and no-illusion trials, respectively. Large supra-additive interaction is only found in illusion trials. Next, this difference TFR for 30–50 Hz frequency range was computed with 25 ms non-overlapping windows and subjected to statistical test (compared with zero by Wilcoxon test). Results are plotted in Fig. 2c,d for illusion and no-illusion trials, respectively. In illusion trials, two stable and relatively long intervals (150–225 ms and 425–500 ms) are found showing significant supra-additive cross-modal interactions. No interval of significant effects was found in no-illusion trials. It should be mentioned that the timing of significant audio-visual interactions remains the same even when equal number of illusion and no-illusion trials (as well as equal number of bimodal and unimodal trials) are considered; thus, the observed effect is not due to the difference in the number of illusion and no-illusion trials.

DISCUSSION

The method of wavelet-based averaging of time–frequency representation provides many advantages over classical averaging technique commonly used in ERP analysis, namely, (i) low-amplitude high frequency activity, which has low signal-to-noise ratio, can be cleverly separated from high-amplitude low-frequency (i.e., theta, alpha) activities which usually mask γ-band responses, (ii) variations in latency from trial to trial (i.e. induced response) can be accommodated whereas only evoked responses (precisely phase-locked to stimulus) can be detected by raw-data averaging, and (iii) the method is free from the requirement of stationarity. As a result, this method is particularly useful in uncovering the correlation between neural responses in γ-band and perceptual experience (illusion *vs* no-illusion).

Significant supra-additive interaction in γ-band responses has been reported previously in a bimodal condition (one beep and one flash, temporally synchronized) primarily in central electrodes [21]. Illusory flash effect may be a better testbed for investigating auditory–visual interactions, since it involves a radical change in the phenomenological quality of the visual percept by sound. The aim of our study was to investigate the γ-band activity of visual cortical regions in the illusory-flash condition, and thus, further trace the neural signature discriminating illusory and no-illusory percepts. The strong supra-additive audio-visual interaction starting at $\sim$150 ms and exclusive to the illusory trials is in close agreement with the reported time intervals of cross-modal interaction in an earlier study of illusory flash effect using standard ERP averaging technique [11]. It is also consistent

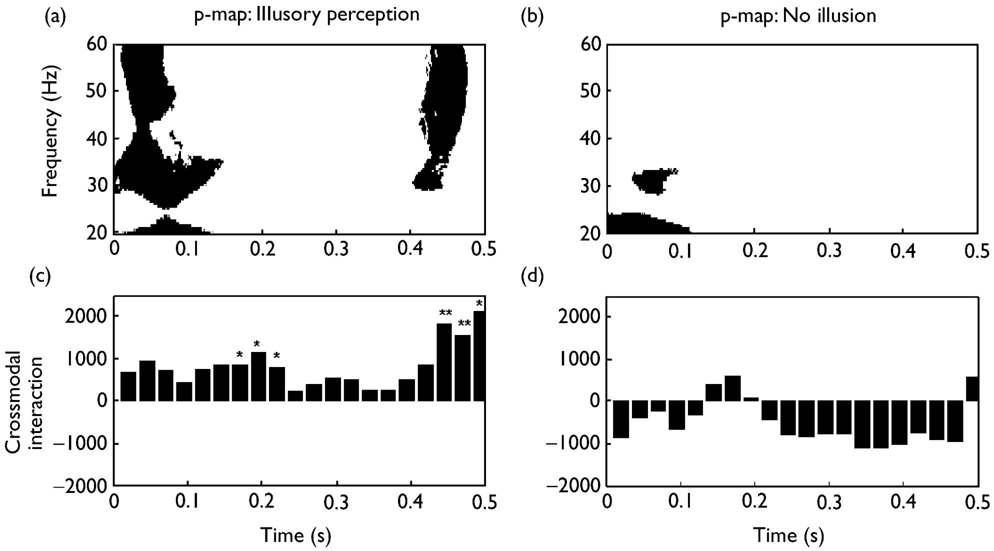

Fig. 2. (**a**,**b**) Statistical probability map (p-map) showing significant ($p < 0.05$) enhancement in TF responses during illusion and no-illusion trials as compared to their pre-stimulus interval. (**c**,**d**) Significant cross-modal supra-additive interaction in γ-band in different time windows for illusion and no-illusion trials, respectively. Significant levels ($*p < 0.05$, $**p < 0.01$) are shown on the top of the bar.

with time intervals reported in other studies using other tasks and paradigms [22]. Any interaction within 200 ms of stimulus presentation has been traditionally assumed to be due to modality-specific (here, visual) pathways [11,22], but a recent dynamic imaging study [24] has estimated the information propagation delay from primary visual areas to prefrontal areas to be ~30 ms. Thus, the possible contribution of feedback projection from multi-modal areas located in higher synaptic hierarchy cannot be ruled out. On the other hand, recent neuroanatomical findings [25,26] suggest a cortical pathway by which auditory signals can directly modulate activity in early visual cortical areas.

Two additional intriguing findings of the study are (i) the very late responses (centered at 450 ms after the onset of visual flash) in the illusion trials, and (ii) increase, though non-significant, in interaction level centered at 170 ms for non-illusion trials (illusion trials showed significant interaction for same time windows). Usually, late ($\geq$ 300 ms after stimulus onset) γ-band responses are considered to be associated with the generation of a coherent object perception, both in audio and visual modalities [12–15]. Here, these responses might reflect the propagation of the activity associated with the perception of illusory flash to higher-order areas (e.g. decision making). Observation (ii) suggests that the perception of illusory double flash is only possible when the modulated visual activity by sound exceeds a ‘perception threshold’ for being registered as a flash percept, otherwise the modulation is unable to trigger awareness. It is for future studies to test this hypothesis.

CONCLUSION

Our results demonstrate that γ band responses are correlated with percept of the illusory flash. The two main factors, discriminating the illusion from no-illusion are: (i) stronger γ band responses, and (ii) supra-additive audio-visual interactions. Taken together, these results suggest that processing of information in visual cortical regions can be modulated by sound and strength of modulation can influence the final perceptual outcome.

REFERENCES

1. Driver J and Spence C. *Curr Biol* **10**, R731–735 (2000).
2. Calvert GA, Brammer MJ and Iversen SD. *Trends Cogn Sci* **2**, 247–253 (1998).
3. Shimojo S and Shams L. *Curr Opin Neurobiol* **11**, 508–509 (2001).
4. McDonald JJ, Teder-Salejärvi WA and Hillyard SA. *Nature* **407**, 906–908 (2000).
5. Stein BE, London N, Wilkinson LK and Price DD. *J Cogn Sci* **8**, 497–506 (1996).
6. Macaluso E, Frith CD and Driver J. *Science* **289**, 1206–1208 (2000).
7. Shams L, Kamitani Y and Shimojo S. *Nature* **408**, 788 (2000).
8. McGurk H and MacDonald J. *Nature* **264**, 746–748 (1976).
9. Armony JL and Dolan RJ. *Neuroreport* **12**, 3407–3411 (2001).
10. Botvinick M and Cohen J. *Nature* **391**, 756 (1998).
11. Shams L, Kamitani Y, Thompson S and Shimojo S. *Neuroreport* **12**, 3849–3852 (2001).
12. Tallon-Baudry C, Bertrand O, Delpuech C and Pernier J. *J Neurosci* **16**, 4240–4249 (1996).
13. Tallon-Baudry C and Bertrand O. *Trends Cogn Sci* **3**, 151–162 (1999).
14. Singer W and Gray CM. *Annu Rev Neurosci* **18**, 555–586 (1996).
15. Csibra G, Davis G, Spratling MW and Johnson MH. *Science* **290**, 1582–1585 (2000).
16. Näätänen R *et al. Nature* **385**, 432–434 (1993).
17. Miltner WHR, Braun C, Arnold M *et al. Nature* **397**, 434–436 (1999).
18. Palva S *et al. J Neurosci* **22**, RC211 (2002).
19. Pulvermüller F, Birbaumer N, Lutzenberger W and Mohr B. *Prog Neurobiol* **52**, 427–445 (1997).
20. Makeig S, Westerfield M, Jung T-P *et al. Science* **295**, 690–694 (2002).
21. Sakowitz OW, Quiroga RQ, Schürmann M and Basar E. *Cogn Brain Res* **11**, 267–279 (2001).
22. Giard MH and Peronnet F. *J Cogn Neurosci* **11**, 473–490 (1999).
23. Calvert GA. *Cerebr Cortex* **11**, 1110–1123 (2001).
24. Foxe JJ and Simpson GV. *Exp Brain Res* **142**, 139–150 (2002).
25. Falchier A, Clavagnier S, Barone P and Kennedy H. *J Neurosci* **22**, 5749–5759 (2002).
26. Rockland KS and Ojima H. *Soc Neurosci Abstr* **511**, 20 (2001).

Acknowledgements: J.B. wishes to thank Sloan–Swartz Foundation for financial support.