VISION RESEARCH

An International Journal for Functional Aspects of Vision

Pre-photoactivation

Post-photoactivation

A   B

C   D

Biochemistry & Cell Biology • Molecular Biology & Genetics
Anatomy, Physiology, Pathology & Pharmacology • Optics, Accommodation & Refractive Error
Circuitry & Pathways • Psychophysics • Perception • Attention & Cognition
Computational Vision • Eye Movements & Visuomotor Control

ISSN 0042-6989 | Volume 93 | 18 December 2013

CrossMark

Audition influences color processing in the sound-induced visual flash illusion

Jyoti Mishra [a],*, Antigona Martinez [b], Steven A. Hillyard [b]

[a] Department of Neurology, Physiology and Psychiatry, University of California, San Francisco, San Francisco, CA 94158, United States
[b] Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, United States

ARTICLE INFO

ABSTRACT

Multisensory interactions can lead to illusory percepts, as exemplified by the sound-induced extra flash illusion (SIFI: Shams, Kamitani, & Shimojo, 2000, 2002). In this illusion, an audio–visual stimulus sequence consisting of two pulsed sounds and a light flash presented within a 100 ms time window generates the visual percept of two flashes. Here, we used colored visual stimuli to investigate whether concurrent auditory stimuli can affect the perceived features of the illusory flash. Zero, one or two pulsed sounds were presented concurrently with either a red or green flash or with two flashes of different colors (red followed by green) in rapid sequence. By querying both the number and color of the participants' visual percepts, we found that the double flash illusion is stimulus specific: i.e., two sounds paired with one red or one green flash generated the percept of two red or two green flashes, respectively. This implies that the illusory second flash is induced at a level of visual processing after perceived color has been encoded. In addition, we found that the presence of two sounds influenced the integration of color information from two successive flashes. In the absence of any sounds, a red and a green flash presented in rapid succession fused to form a single orange percept, but when accompanied by two sounds, this integrated orange percept was perceived to flash twice on a significant proportion of trials. In addition, the number of concurrent auditory stimuli modified the degree to which the successive flashes were integrated to an orange percept vs. maintained as separate red–green percepts. Overall, these findings show that concurrent auditory input can affect both the temporal and featural properties of visual percepts.

1. Introduction

When a single flash of light is presented interposed between two brief auditory stimuli separated by 60–100 ms, individuals typically report perception of two flashes (Shams, Kamitani, & Shimojo, 2000, 2002). The neural basis of this multisensory sound-induced flash illusion (SIFI) has been investigated in several electrophysiological and neuroimaging studies (Arden, Wolf, & Messiter, 2003; Bhattacharya, Shams, & Shimojo, 2002; Mishra et al., 2007; Shams et al., 2001, 2005; Watkins et al., 2006, 2007). In a detailed investigation of the phenomenon using event related potential recordings (ERPs), Mishra et al. (2007) found that the illusion is based on a rapid interplay between auditory, visual and polysensory cortical areas. Notably, however, the neural activity pattern underlying the illusory flash was found to be very different from the activity elicited by a real flash. These neurophysiological differences raise questions regarding the properties of the illusory

flash, in particular whether it can possess distinctive visual features like those of a real flash such as color, shape, contrast and size. In the present study we extend the Shams paradigm by probing additional information about the final visual percept (its color specificity) as modified by concurrent sounds. By doing this, the current experiment provides a more detailed understanding of the featural attributes of multisensory percepts that may occur in real-life situations; for example, the integration of sounds and lights at a music concert or on a busy highway.

Researchers who first described the SIFI have shown that the phenomenon can be elicited under a wide range of stimulus parameters of shape, size, texture and duration (Shams, Kamitani, & Shimojo, 2000; Shams et al., 2005; Watkins et al., 2006), and this was recently expanded to flashed visual objects such as faces and buildings (Setti & Chan, 2011). McCormick and Mamassian (2008) further showed that the illusory flash can have a measurable contrast. In this case, the presence of two sounds lowered the threshold contrast of the second flash in a sequence of two flashes. An unresolved question, however, is whether the SIFI is stimulus specific; i.e., does the illusory flash have the same or similar features as the inducing real flash.

* Corresponding author. Address: University of California, San Francisco-Mission Bay, Sandler Neurosciences Center, Room 502, MC 0444 675 Nelson Rising Lane, San Francisco, CA 94158, United States. Fax: +1 4155027538.
E-mail address: jyoti@gazzaleylab.ucsf.edu (J. Mishra).

The present study investigated the stimulus-specificity of the SIFI in terms of its color features. A variable number of sounds (0, 1 or 2) were paired with either a red or a green colored flash, and participants were asked to report the number as well as the color of their final visual percept. If the color of the illusory flash matched the color of the real flash, this would be taken as evidence for feature-specificity of the illusion. In a recent study Setti and Chan (2011) demonstrated the SIFI with face and scene stimuli, but they did not query the feature content of the illusory flash. Hence, their study did not directly test the feature-specificity of the SIFI, which we aim to investigate here by asking individuals to report the featural content of their visual percepts as well as the number of visual stimuli perceived. In addition, we investigated whether concurrent sounds would influence color integration by presenting a variable number of sounds (0, 1 or 2) paired with a rapid visual sequence of a red followed by a green flash. A rapid red–green sequence by itself usually results in a fused orange percept. Again, by asking individuals to report the number and color of their visual percepts, we examined whether auditory input could influence this red–green to orange color integration.

2. Materials and methods

Seventeen right-handed healthy adults (8 males and 9 females, age mean ± standard deviation 23.4 ± 3.3 years) participated in the study after giving informed consent as approved by the University of California San Diego Human Research Protections Program. Each participant had normal or corrected-to-normal vision and normal hearing. Participants in the study were pre-selected as individuals who perceived the SIFI on at least 40% of the trials in a short 2-min screen prior to the main experiment. The screening block consisted of identical visual and audio–visual stimuli as previously used in Mishra et al. (2007) to study the illusion. Approximately 34% of individuals screened met this criterion for participation in the study.

The experiment was conducted in a sound-attenuated chamber having a background sound level of 32 dB SPL and a background luminance of 2 cd/m^2. Visual stimuli were presented on the horizontal meridian at 8° of visual angle (va) in the left visual field on an LCD monitor, as in our prior investigations of the SIFI (Mishra, Martinez, & Hillyard, 2008, 2010; Mishra et al., 2007). Visual stimuli were annuli (3.7° va outer diameter and 0.8° va thickness) flashed for 32 ms at a luminance level (measured by photometer) of 75 cd/m^2. Participants maintained fixation on a cross positioned at the center of the mid-level gray screen at a viewing distance of 83 cm. Auditory stimuli were presented in free field simultaneously from speakers attached to the right and left sides of the monitor display, thereby resulting in centrally localized sounds. Auditory stimuli were 76 dB SPL noise bursts with 10 ms durations.

During the experimental runs, participants were presented with the following 17 different visual (V) and audio–visual (AV) stimulus combinations (see Fig. 1): 8 of the stimuli contained either red (r) or green (g) colored stimuli (V$_{1r}$, V$_{1g}$, V$_{2r}$, V$_{2g}$, A$_1$V$_{1r}$, A$_1$V$_{1g}$, A$_2$V$_{1r}$, A$_2$V$_{1g}$), while 9 of the stimuli contained a first red and a second green visual stimulus (V$_{2rg}$, A$_1$V$_{2rg}$, A$_2$V$_{2rg}$) that were presented at three different red–green SOAs of 50 ms, 84 ms and 100 ms. Suffixes 1 or 2 denote presence of one or two auditory or visual components within each stimulus combination. For audiovisual stimuli, the first sound was always temporally aligned with the first visual stimulus at onset. For audiovisual stimuli containing two sounds, the SOA between the two auditory stimuli was set at a constant 67 ms as this SOA reliably produced the SIFI in our previous studies (Mishra et al., 2007). The 17 experimental stimuli were presented equiprobably and in random order on each experimental run. Each run included 12 trials each of the 17 stimulus
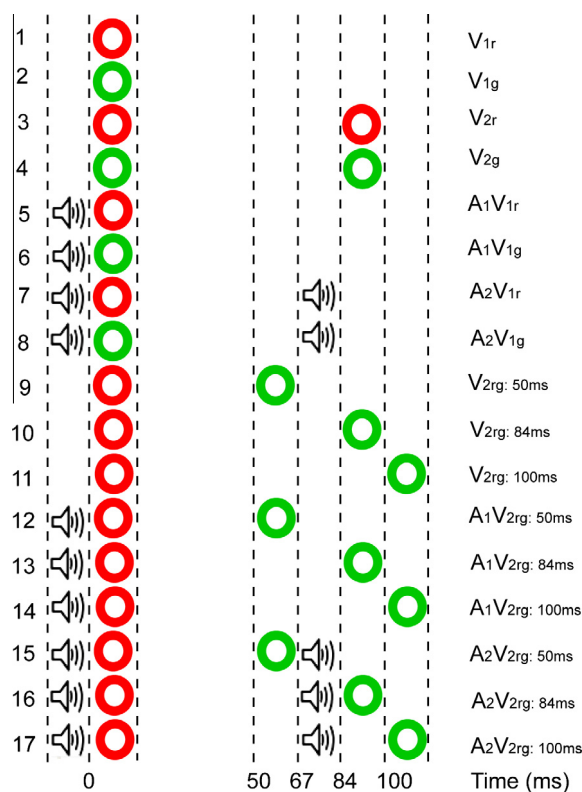
Fig. 1. Overview of the experimental design. The stimulus presentation timeline is shown at the bottom, each labeled time point below the dotted lines corresponds to stimulus onset.

types. Inter-trial intervals were set at 800 ms with a ±300 ms jitter. Each experimental run of 204 stimuli lasted roughly 3 min. A total of four runs were conducted in the experiment.

Participants used a joystick to report the color and number of visual percepts on each trial. Perceived color was reported as one of four choices (i) red, (ii) green, (iii) orange or (iv) both red and green annuli. Choice (iii) was reported on trials on which the sequential red and green annuli fused to form an orange percept. Color choice (iv) was reported when both red and green colors were perceived on any given trial. The number of perceived flashes (two vs. one) was reported with a separate button. For color choice (iv), subjects were instructed to make the two-flash numerical response if either one or both of the red and green colors was perceived to flash twice. At the end of the experiment participants were asked if they perceived any other color (e.g., white) not provided as one of the red/green/orange color response choices; all participants consistently replied "no".

Percentages of the different types of responses across conditions were analyzed by repeated measures ANOVAs. Post hoc analyses consisted of two-tailed dependent sample t-tests. For each comparison, effect sizes were reported as the Cohen's d measure (Cohen, 1988).

3. Results

The percentages of one-flash and two-flash responses to the visual and audio–visual stimuli that contained a single color component, either red (V$_{1r}$, V$_{2r}$, A$_1$V$_{1r}$, A$_2$V$_{1r}$) or green (V$_{1g}$, V$_{2g}$, A$_1$V$_{1g}$, A$_2$V$_{1g}$), are shown in Fig. 2A and Table 1. A 2 × 4 repeated measures ANOVA on the percentage of two-flash responses with stimulus color (red/green) and stimulus type (V$_1$, V$_2$, A$_1$V$_1$, A$_2$V$_1$) as factors showed a main effect of color with more two-flash reports
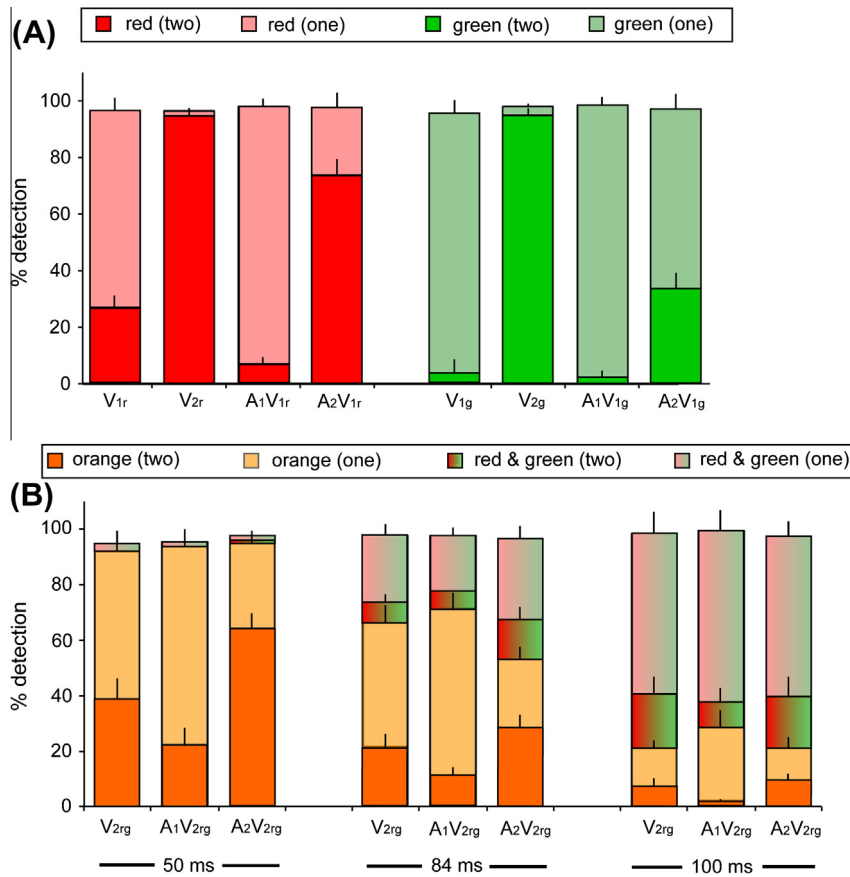
Fig. 2. The color and number (in parentheses) of perceptual reports for each stimulus type containing either a single color component (A) or two color components (B). In (A), dark and light bars (for red stimuli on the left and green stimuli on the right) depict the mean percentages of two and one colored flash reports, respectively. Note that the increased two-flash reports (depicted in darker color) on single-flash, two-sound trials (A_2V_{1r}, A_2V_{1g}) reflect the SIFI. In (B) dark and light orange bars represent percentage reports of two and one orange flash percepts, respectively. The red–green-gradient bars indicate percentage reports of separate red and green percepts: dark bars indicate that more than one red or green flash was perceived and light bars indicate that only one red and one green flash was perceived. Note that in (B) a color report of orange was due to perceptual fusion of the red and green flashes; orange colored stimuli were never actually presented in the experiment. In each bar the summation of the different types of obtained responses add up to 100% or near 100% (in cases where a few responses were missed by some individuals on some trials). There were no exclusively red or exclusively green responses obtained to the stimuli represented in (B). Error bars are standard errors of the mean (s.e.m.). Corresponding numerical data are given in Tables 1, 2 and 4.

Table 1
Percent two flash reports (mean ± standard error) for stimuli in the screening block and for red and green color stimuli in the experiment.

Stimulus	Screening	Red	Green
V_1	4.90 ± 1.71	26.47 ± 4.71	3.92 ± 2.51
V_2	98.53 ± 1.71	94.61 ± 2.49	95.10 ± 1.97
A_1V_1	1.47 ± 0.77	6.86 ± 2.52	1.96 ± 1.30
A_2V_1	48.03 ± 6.37	73.53 ± 5.93	33.33 ± 7.96

occurring on red than on green trials ($F(1,16) = 39.35$, $p < 0.0001$), and a main effect of stimulus type ($F(3,48) = 138.60$, $p < 0.0001$) with the largest percentage of two-flash reports occurring for V_2, the least for A_1V_1 and V_1, and an intermediate percentage for A_2V_1. Note that the main effect of color (i.e., greater two-flash reports on red than green trials) may have occurred because we did not precisely equate perceptual luminance for these colors (e.g., by flicker photometry) prior to the experiment, which is a limitation of the present study. Post hoc paired t-tests comparing responses to these different stimulus combinations were conducted separately for red and for green stimuli. For red stimuli, illusory two-flash reports were significantly greater when single red flashes were accompanied by two sounds compared to when these were accompanied by one sound (A_2V_{1r} vs. A_1V_{1r},

$p < 0.0001$, effect size: $d = 3.83$) or by no sound (A_2V_{1r} vs. V_{1r}, $p < 0.0001$, $d = 2.14$). Two-flash reports for the single red flash accompanied by two sounds were significantly fewer than for two red flashes accompanied by no sounds (A_2V_{1r} vs. V_{2r}, $p = 0.004$, $d = -1.21$). The least number of two-flash reports (in error) were reported for single red flashes accompanied by one sound, significantly fewer than for single red flashes with no sound (A_1V_{1r} vs. V_{1r}, $p = 0.0002$, $d = -1.31$). Similar results were obtained for two-flash responses to the green stimuli: $A_2V_{1g} > A_1V_{1g}$ ($p = 0.001$, $d = 1.64$), $A_2V_{1g} > V_{1g}$ ($p = 0.0005$, $d = 1.36$), $A_2V_{1g} < V_{2g}$ ($p < 0.0001$, $d = -3.02$), but the number of erroneous two-flash responses to the single green flash accompanied by one sound vs. no sound did not differ significantly (A_1V_{1g} vs. V_{1g}, $p = 0.5$, $d = -0.25$).

We further compared the two-flash reports on the above described single color component stimuli with the results obtained in the screening block prior to the experiment (Table 1). For this analysis, two flash reports on red vs. green stimuli in the main experiment were averaged together and compared to the two flash reports in the screening block. This was done in a 2×4 repeated measures ANOVA with experiment (main/screening) and stimulus type (V_1, V_2, A_1V_1, A_2V_1) as factors. A main effect of experiment did not reach significance ($p = 0.09$) suggesting overall equivalent performance in the screening block and the main experiment. A main effect of stimulus type was observed ($F(3,48) = 203.24$, $p < 0.0001$)

with the largest percentage of two-flash reports occurring for V_2, the least for A_1V_1 and V_1, and an intermediate percentage for A_2V_1. A significant experiment × stimulus type interaction ($F(3,48) = 3.60$, $p = 0.02$) was further parsed in post hoc tests. Post hoc t-tests between screening and main experiment stimuli showed no difference in two flash reports for A_2V_1 ($p = 0.34$, $d = -0.19$) or for A_1V_1 ($p = 0.13$, $d = -0.53$) or for V_2 ($p = 0.16$, $d = 0.45$), but significantly reduced two flash responses were reported for V_1 ($p = 0.0005$, $d = -0.94$) during screening.

This pattern of results was generally consistent with prior studies that demonstrated a high incidence of two-flash reports to uncolored single flash stimuli accompanied by two sounds (Mishra et al., 2007; Shams, Kamitani, & Shimojo, 2000, 2002). In the present study where simultaneous color and number reports were required, it was further demonstrated that participants could reliably categorize the color of the double flash percepts as either red or green. Most importantly, this was true even for the illusion inducing one-flash–two-sound stimulus. Even though four color choices were available to participants (see Section 2), two red flashes were consistently reported for the red-flash–two-sound combination while two green flashes were reported for the green-flash–two-sound stimulus. Significantly, there were virtually no two-color (red and green) responses made to either of these illusion-inducing

stimulus combinations. These results demonstrated color stimulus specificity for the illusion.

To investigate the effect of auditory input on color integration, participants were also presented with visual stimuli having two color components, a red flash followed by a green flash at three SOAs of 50, 84 and 100 ms (Fig. 2B and Table 2). Color integration (color fusion) occurred when the rapid sequence of red followed by green was fused into an orange percept. The influence of concurrent sound stimuli on the percentage of double orange flash reports was investigated in a repeated measures ANOVA with stimulus type (0, 1 or 2 sounds accompanying the rapid red–green flash sequence) and SOA as factors. This ANOVA showed a main effect of SOA with the highest percentage of two-orange flash reports at the shortest SOA (50 ms > 84 ms > 100 ms: $F(2,32) = 27.02$, $p < 0.0001$). A main effect of stimulus type ($F(2,32) = 25.49$, $p < 0.0001$) revealed the highest percentage of two-orange flash reports for the red–green flash sequence accompanied by two sounds (A_2V_{2rg}) and lowest percentage of two-orange flash reports for the red–green flash sequence accompanied by one sound (A_1V_{2rg}). The stimulus type × SOA interaction was also highly significant ($F(4,64) = 13.66$, $p < 0.0001$), and specific comparisons between the different stimuli types at the different SOAs that underlie this interaction are given in Table 3. Similar t-test comparisons for all other response choices made by individuals are also elaborated in Table 3. Overall, these results showed that the number of pulsed sounds (0, 1 or 2) can modify the perceived number of fused orange color percepts. Thus, we show that the pulsed sounds not only influence the number of percepts generated to a physically presented color, but also the number of percepts of a color (orange) that itself is a product of sensory color (red and green) integration.

To further analyze the influence of sounds on visual color integration, we tested whether 0 or 1 or 2 sounds associated with the red–green flash sequence influenced the total percentage of integrated color (i.e., percent orange) reports. The numerical orange flash reports (two vs. one) were collapsed for this analysis (see Table 4) and a 3 × 3 ANOVA was conducted with factors of stimulus type (0, 1 or 2 sounds accompanying the red–green flash sequence) and SOA. A highly significant main effect of SOA ($F(2,32) = 120.25$, $p < 0.0001$) showed the most color integration (i.e., orange reports) occurred at 50 ms > 84 ms > 100 ms. The main effect of stimulus type was significant ($F(2,32) = 4.69$, $p = 0.02$) because there were fewer orange reports for the red–green sequence paired with two sounds (A_2V_{2rg}) in comparison with one sound (A_1V_{2rg}) or no sounds (V_{2rg}). This suggested that two sounds individuated the red and green colors relative to one or no sounds. A stimulus type × SOA interaction was also significant ($F(4,64) = 4.95$, $p = 0.002$), suggesting that the influence of sounds on orange

Table 2
Percept reports (mean ± standard error) for red–green flash sequence stimuli as either two orange flashes (orange (two)), one orange flash (orange (one)), separate red and green flashes with either the red or green color flashing more than once (red and green (two)), and separate red and green flashes with one red and one green flash perceived (red and green (one)). The summation of these four different types of obtained responses adds up to 100% or near 100% (in cases where a few responses were missed by some individuals on some trials). Note that there were no exclusively red or exclusively green responses obtained to these red–green flash sequence stimuli. Also note that a color report of orange was due to perceptual fusion of the red and green flashes; orange colored stimuli were never actually presented in the experiment.

SOA	Stimulus	Orange (two)	Orange (one)	Red and green (two)	Red and green (one)
50 ms	V_{2rg}	38.73 ± 7.78	53.43 ± 7.56	0	2.94 ± 1.69
	A_1V_{2rg}	22.06 ± 6.35	71.57 ± 6.54	0	1.96 ± 1.10
	A_2V_{2rg}	64.22 ± 5.66	30.39 ± 4.70	1.47 ± 1.43	1.96 ± 0.86
84 ms	V_{2rg}	21.57 ± 4.86	44.61 ± 6.16	7.35 ± 2.76	24.51 ± 3.89
	A_1V_{2rg}	10.78 ± 3.52	60.78 ± 5.40	5.88 ± 3.00	20.10 ± 3.18
	A_2V_{2rg}	28.43 ± 4.50	24.51 ± 4.73	14.22 ± 4.99	29.41 ± 4.81
100 ms	V_{2rg}	7.35 ± 2.58	13.73 ± 3.10	19.61 ± 6.31	57.84 ± 7.58
	A_1V_{2rg}	1.96 ± 0.86	26.47 ± 6.32	9.31 ± 5.23	61.76 ± 7.27
	A_2V_{2rg}	9.31 ± 2.29	11.76 ± 3.67	18.63 ± 6.85	57.35 ± 6.08

Table 3
Paired t-tests and effect sizes (d) for comparisons between the percent perceived responses across the different auditory combinations (0, 1 or 2 sounds) accompanying the red–green flash sequence stimuli presented at the three different SOAs. Comparisons were made separately at each of the four response choices; non-significant results are in gray; negative Cohen's d occurred when the second of the two compared stimuli (in second column) had greater mean percent responses.

Response choice	Stimulus comparison	50 ms SOA	84 ms SOA	100 ms SOA
Orange (two)	$A_2V_{2rg} > A_1V_{2rg}$	$p < 0.0001$, $d = 1.70$	$p = 0.0006$, $d = 1.07$	$p = 0.005$, $d = 1.13$
	$A_2V_{2rg} > V_{2rg}$	$p = 0.0008$, $d = 0.92$	$p = 0.13$, $d = 0.36$	$p = 0.51$, $d = 0.19$
	$V_{2rg} > A_1V_{2rg}$	$p = 0.002$, $d = 0.57$	$p = 0.007$, $d = 0.62$	$p = 0.02$, $d = 0.76$
Orange (one)	$A_2V_{2rg} < A_1V_{2rg}$	$p < 0.0001$, $d = 1.78$	$p < 0.0001$, $d = 1.74$	$p = 0.008$, $d = 0.71$
	$A_2V_{2rg} < V_{2rg}$	$p = 0.002$, $d = 0.91$	$p = 0.002$, $d = 0.90$	$p = 0.55$, $d = 0.14$
	$V_{2rg} < A_1V_{2rg}$	$p = 0.002$, $d = 0.62$	$p = 0.003$, $d = 0.68$	$p = 0.03$, $d = 0.66$
Red and green (two)	$A_2V_{2rg} > A_1V_{2rg}$	$p = 0.33$, $d = 0.5$	$p = 0.005$, $d = 0.51$	$p = 0.02$, $d = 0.37$
	$A_2V_{2rg} > V_{2rg}$	$p = 0.33$, $d = 0.5$	$p = 0.04$, $d = 0.43$	$p = 0.83$, $d = -0.04$
	$V_{2rg} > A_1V_{2rg}$	-	$p = 0.48$, $d = 0.12$	$p = 0.03$, $d = 0.43$
Red and green (one)	$A_2V_{2rg} > A_1V_{2rg}$	$p = 1.00$, $d = 0$	$p = 0.05$, $d = 0.57$	$p = 0.56$, $d = -0.16$
	$A_2V_{2rg} > V_{2rg}$	$p = 0.50$, $d = -0.19$	$p = 0.17$, $d = 0.27$	$p = 0.93$, $d = -0.02$
	$V_{2rg} > A_1V_{2rg}$	$p = 0.33$, $d = 0.17$	$p = 0.23$, $d = 0.30$	$p = 0.61$, $d = -0.13$

Table 4
Percent total orange percept reports and percent total red and green percept reports (one and two number reports combined) in response to red–green stimulus sequences (mean ± standard error). These two response types summed up to 100% or near 100% (in cases where a few responses were missed by some individuals on some trials). There were no exclusively red or exclusively green responses obtained to the red–green sequence stimuli. The color report of orange was due to perceptual fusion of the red and green flashes; orange colored stimuli were never actually presented in the experiment.

SOA	Stimulus	Total orange reports	Total red and green reports
50 ms	V_{2rg}	92.16 ± 3.36	2.94 ± 1.69
	A_1V_{2rg}	93.63 ± 2.14	1.96 ± 1.10
	A_2V_{2rg}	94.61 ± 2.49	3.43 ± 1.97
84 ms	V_{2rg}	66.18 ± 4.01	31.86 ± 3.87
	A_1V_{2rg}	71.57 ± 3.40	25.98 ± 3.85
	A_2V_{2rg}	52.94 ± 4.44	43.63 ± 4.73
100 ms	V_{2rg}	21.08 ± 4.28	77.45 ± 5.31
	A_1V_{2rg}	28.43 ± 6.62	71.08 ± 7.07
	A_2V_{2rg}	21.08 ± 4.71	75.98 ± 4.38

color integration was different at the three SOAs; this was further parsed in separate ANOVAs at each SOA with the factor of stimulus type. The effect of stimulus type on orange color integration was not significant at either 50 ms ($p = 0.65$) or 100 ms ($p = 0.21$) SOAs but was highly significant at the 84 ms SOA ($F(2, 32) = 10.85$, $p = 0.0003$). Post hoc t-tests were used to compare the 0, 1 and 2 sound stimuli at 84 ms. At 84 ms, orange color integration was significantly less for the two sound vs. one sound stimulus ($A_2V_{2rg} < A_1V_{2rg}$, $p = 0.0003$) and for the two sound vs. no sound stimulus ($A_2V_{2rg} < V_{2rg}$, $p = 0.005$), whereas color integration did not differ between the one vs. no sound stimulus ($A_1V_{2rg} < V_{2rg}$, $p = 0.21$). These results showed that specifically for the audiovisual stimulus combination in which the second sound preceded the second flash by a few milliseconds (17 ms) (A_2V_{2rg}, 67 ms auditory SOA, 84 ms visual SOA), the two sounds facilitated individuation of the red and green colors; this did not occur when the second sound followed the second flash by 17 ms (A_2V_{2rg}, 67 ms auditory SOA, 50 ms visual SOA) or when the second sound preceded the second flash at a larger audiovisual separation of 33 ms (A_2V_{2rg}, 67 ms auditory SOA, 100 ms visual SOA). These results show that for specific audiovisual temporal combinations, two sounds not only induce the SIFI but also modulate the extent of color integration.

4. Discussion

The present results provide insight into how auditory stimulation affects visual perception in the sound-induced flash illusion (SIFI). By probing both the number and color of the final visual percept, it was found that the illusory flash possesses the same color as the real flash; that is, a single red or green flashed stimulus paired with two sounds was perceived as either two red or two green flashes, respectively. This result implies that the illusory flash is induced at a level of visual processing after perceived color has been encoded. We further investigated the pairing of sounds with a rapid sequence of a red then green flash that fuse to form an orange percept. In this case, it was found that the concurrent sounds affected the number of perceived orange flashes. To our knowledge, this is the first report of auditory stimulation influencing an illusory visual color percept as well as the sensory integration of color information.

Visual integration of color information is hypothesized to be a cortical process (Chatterjee & Callaway, 2003; Terao et al., 2010; Nijhawan, 1997; Nishida et al., 2007; Watanabe & Nishida, 2007). The color content of a visual stimulus is initially encoded by 'red/green' and 'blue/yellow' opponent cells in the retina and the lateral geniculate nucleus (LGN) and relayed as segregated signals to

visual cortex (Dacey, 2000; Dacey et al., 2003; Lennie & D'Zmura, 1988). Within visual cortical circuits, the information encoded in the color-opponent input signals is integrated to form the full range of perceived colors. Here, we found that the simultaneous presentation of sounds with a sequence of red and green colored stimuli robustly influenced the number of integrated color percepts, such that the perceived (but not physically presented) color orange was observed to flash twice in the presence of two sounds. This result suggests that the auditory signals may influence visual perception by creating the percept of an extra flash immediately after the stage of cortical color processing at which inputs from color-opponent cells are integrated. These results are in agreement with our prior electrophysiological findings, which showed that the early visual processing associated with perception of the sound-induced illusory flash was localized to ventral extrastriate visual cortex – a cortical region specializing in visual form and color processing (Mishra et al., 2007).

Watkins et al. (2006, 2007) also investigated the neural basis of the SIFI, using instead functional magnetic resonance imaging (fMRI). They described generally enhanced BOLD activations for audiovisual relative to visual stimuli throughout retinotopic visual cortex (areas V1, V2 and V3) as well as non-retinotopic cortical regions such as the superior temporal sulcus and gyrus. Only the V1 locus was analyzed to compare illusion vs. non-illusion trials of the SIFI stimulus, and differential activations were indeed found, but these did not correlate with the percentage of illusory reports across individuals. The extra-striate and superior temporal activations were not specifically analyzed for associations with the illusory percept; hence a striate vs. extrastriate visual cortex locus for the neural basis of the illusion cannot be ascertained from these studies.

In the present analysis of the integrated orange color percept elicited by the red–green stimulus sequence, we found that for certain audiovisual stimulus combinations the presence of two sounds alters the relative percentage of integrated (orange) vs. separated (red and green) perceptual reports. Specifically, for the red–green flash sequence separated by 84 ms paired with two sounds, orange percepts (collapsed over one vs. two orange reports) were significantly reduced relative to trials with one or no sounds; in other words, at the 84 ms SOA the presence of two sounds individuated the red and green color tokens to a greater extent. This is akin to the pip and pop effect in which a sound 'pip' makes a temporally aligned visual stimulus 'pop' or appear more salient (Van der Burg et al., 2008). Interestingly, the present individuation effect was observed even though the second sound and second flash were not exactly temporally aligned as in the pip and pop effect; specifically, the individuation was present for a 17 ms sound–flash separation (sound preceding flash), but not for a 17 ms flash–sound separation (flash preceding sound) or a longer 33 ms sound–flash separation. That a sound immediately preceding the second flash makes the second flash more salient and reduces color integration with the first flash, is suggestive of rapid priming of vision by audition (Logeswaran & Bhattacharya, 2009). This observation aligns with our prior electrophysiological evidence that the second sound can very rapidly modify visual processing (Mishra et al., 2007) possibly via direct connections between auditory and visual sensory cortices (Clavagnier, Falchier, & Kennedy, 2004; Clemo et al., 2008; Falchier et al., 2002; Hall & Lomber, 2008; Rockland & Ojima, 2003). The finding that a temporally preceding sound can modify visual color integration while a sound that follows the visual stimulus onset by 17 ms fails to do so may also be a consequence of the rapid bottom-up nature of visual color integration. In the case where the sound follows the visual stimulus, color integration may be nearly complete before the sound can individuate the two integrated colors. This remains to be confirmed in future experiments, especially with neural recordings accompanying the behavioral assays.

That audition can interact with a qualitative visual property such as color is in line with the prior studies that have shown influences of audition on other visual attributes. For example, Lippert, Logothetis, and Kayser (2007) showed that when a sound provided information about a visual stimulus such as its timing, the contrast of the visual stimulus was enhanced. Similarly, McCormick and Mamassian (2008) showed that the presence of two tones can lower the threshold contrast of a second flash in a sequence of two flashes. In the pip and pop effect, a concurrent auditory stimulus reduces visual search times, and notably this was shown not to be due to a general alerting effect (Van der Burg et al., 2008). The presentation of an auditory attention-directing cue has also been shown to boost the apparent contrast of a subsequent visual target at the cued location (Störmer, McDonald, & Hillyard, 2009) and to alter visual detectability and time-order perception (McDonald, Teder-Sälejärvi, & Hillyard, 2000; McDonald et al., 2005).

In summary, the present study provides new insights into the SIFI phenomenon. First, it was found that the illusory flash incorporates the color feature of the actual flash that is bracketed by two sounds. Second, concurrent auditory stimuli were found to modify the perceived number of colored flashes resulting from the integration of two separately presented colors. Overall, these findings add to our understanding of how auditory stimulation modifies the temporal and featural properties of a visual percept. The present results go beyond the widely accepted 'modality appropriateness hypothesis' proposed to explain the interaction between the auditory and visual sensory systems (Welch, DuttonHurt, & Warren, 1986; Welch & Warren, 1980). According to this hypothesis, audition dominates the temporal component of the multisensory percept because temporal coding is more accurate within audition, whereas vision dominates the qualitative form and spatial properties of the percept. Recent studies have further supported the idea that reliance on the auditory modality for temporal precision is due to its greater predictive certainty in the temporal domain, in accordance with a Bayesian ideal observer (Apthorp, Alais, & Boenke, 2013; Shams, Ma, & Beierholm, 2005). Here we show that auditory stimulation not only influences the temporal properties of visual percepts, but also, for very specific audiovisual combinations, audition can affect the degree of perceptual integration vs. separation when different colored flashes are presented in rapid succession. These results thus further our understanding of audio–visual interactions in perception.

Acknowledgments

References

Apthorp Alais & Boenke (2013). Flash illusions induced by visual, auditory, and audiovisual stimuli. *Journal of Vision, 13.* http://dx.doi.org/10.1167/13.5.3 (pii: 3).

Arden Wolf & Messiter (2003). Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. *Vision Research, 43,* 2469–2478.

Bhattacharya Shams & Shimojo (2002). Sound-induced illusory flash perception: Role of gamma band responses. *Neuroreport, 13,* 1727–1730.

Chatterjee & Callaway (2003). Parallel colour-opponent pathways to primary visual cortex. *Nature, 426,* 668–671.

Clavagnier Falchier & Kennedy (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective & Behavioral Neuroscience, 4,* 117–126.

Clemo Sharma Allman & Meredith (2008). Auditory projections to extrastriate visual cortex: Connectional basis for multisensory processing in "unimodal" visual neurons. *Experimental Brain Research, 191,* 37–47.

Cohen (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Routledge.

Dacey (2000). Parallel pathways for spectral coding in primate retina. *Annual Review of Neuroscience, 23,* 743–775.

Dacey Peterson Robinson & Gamlin (2003). Fireworks in the primate retina: In vitro photodynamics reveals diverse LGN-projecting ganglion cell types. *Neuron, 37,* 15–27.

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

Hall & Lomber (2008). Auditory cortex projections target the peripheral field representation of primary visual cortex. *Experimental Brain Research, 190,* 413–430.

Lennie & D'Zmura (1988). Mechanisms of color vision. *Critical Reviews in Neurobiology, 3,* 333–400.

Lippert Logothetis & Kayser (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173,* 102–109.

Logeswaran & Bhattacharya (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters, 455,* 129–133.

McCormick & Mamassian (2008). What does the illusory-flash look like? *Vision Research, 48,* 63–69.

McDonald Teder-Sälejärvi Di Russo & Hillyard (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience, 8,* 1197–1202.

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

Mishra Martinez & Hillyard (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research, 1242,* 102–115.

Mishra Martinez & Hillyard (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience, 22,* 1714–1729.

Mishra Martinez Sejnowski & Hillyard (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *The Journal of Neuroscience, 27,* 4120–4131.

Nijhawan (1997). Visual decomposition of colour through motion extrapolation. *Nature, 386,* 66–69.

Nishida Watanabe Kuriki & Tokimoto (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology: CB, 17,* 366–372.

Rockland & Ojima (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50,* 19–26.

Setti & Chan (2011). Familiarity of objects affects susceptibility to the sound-induced flash illusion. *Neuroscience Letters, 492,* 19–22.

Shams Iwaki Chawla & Bhattacharya (2005). Early modulation of visual cortex by sound: An MEG study. *Neuroscience Letters, 378,* 76–81.

Shams Kamitani & Shimojo (2000). Illusions. What you see is what you hear. *Nature, 408,* 788.

Shams Kamitani & Shimojo (2002). Visual illusion induced by sound. *Brain Research. Cognitive Brain Research, 14,* 147–152.

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

Shams Ma & Beierholm (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport, 16,* 1923–1927.

Störmer McDonald & Hillyard (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 22456–22461.

Terao Watanabe Yagi & Nishida (2010). Smooth pursuit eye movements improve temporal resolution for color perception. *PloS One, 5,* e11214.

Van der Burg Olivers Bronkhorst & Theeuwes (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1053–1065.

Watanabe & Nishida (2007). Veridical perception of moving colors by trajectory integration of input signals. *Journal of Vision, 7,* 1–16 (article no. 3).

Watkins Shams Josephs & Rees (2007). Activity in human V1 follows multisensory perception. *NeuroImage, 37,* 572–578.

Watkins Shams Tanaka Haynes & Rees (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage, 31,* 1247–1256.

Welch DuttonHurt & Warren (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics, 39,* 294–300.

Welch & Warren (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667.