Postprint: The Effect of Channel-Based Endogenous Attention on the Sound-Induced Flash Illusion

Authors: Zhang Ming, Tang Xiaoyu, Yu Wei, Ningbo, Wang Zhinan, Wang Aijun

Date: 2018-09-07T00:00:00+00:00

Abstract

The sound-induced flash illusion refers to the phenomenon where, when visual flashes accompanied by an unequal number of auditory sounds are presented sequentially or simultaneously within 100 ms, individuals illusorily perceive the number of visual flashes to be equal to the number of auditory sounds. Using the classic sound-induced flash illusion paradigm, this study endogenously directed attention to either the visual or auditory modality to investigate the effect of modality-based endogenous attention on the sound-induced flash illusion. The results showed that when attention was endogenously directed to the visual modality, the magnitude of the fission illusion was significantly reduced relative to the baseline condition; when attention was endogenously directed to the auditory modality, there was a trend for the magnitude of the fission illusion to increase relative to the baseline condition. This indicates that modality-based endogenous attention can influence the fission illusion in the sound-induced flash illusion, but does not affect the fusion illusion.

Full Text

The Effects of Modal-Based Endogenous Attention on Sound-Induced Flash Illusion

ZHANG Ming[1], TANG Xiaoyu[2], YU Wei[3], NING Bo[1], WANG Zhinan[1], WANG Aijun[1]

[1] Department of Psychology, Research Center for Psychology and Behavioral Sciences, Soochow University, Suzhou 215123, China
[2] School of Psychology, Liaoning Collaborative Innovation Center of Children and Adolescents Healthy Personality Assessment and Cultivation, Liaoning Normal University, Dalian 116029, China

[3] Admission and Employment Office, Changchun University of Chinese Medicine, Changchun 130117, China

Abstract

The sound-induced flash illusion refers to the phenomenon where, when a visual flash is accompanied by an unequal number of auditory beeps presented within 100 ms (either sequentially or simultaneously), individuals illusorily perceive the number of visual flashes as equal to the number of auditory beeps. Using the classic sound-induced flash illusion paradigm, the present study investigated how modal-based endogenous attention—directed endogenously toward either the visual or auditory modality—affects this illusion. The results revealed that when attention was endogenously directed to the visual modality, the magnitude of the fission illusion decreased significantly compared to the baseline condition. When attention was endogenously directed to the auditory modality, the magnitude of the fission illusion showed a trend toward increase relative to the baseline condition. These findings demonstrate that modal-based endogenous attention can modulate the fission illusion in sound-induced flash illusion, but does not affect the fusion illusion.

Keywords: endogenous attention; sound-induced flash illusion; auditory dominance; visual and auditory modalities
Classification: B842

Introduction

The sound-induced flash illusion (SIFI) is an auditory-dominant phenomenon in multisensory integration where individuals illusorily perceive the number of visual flashes as equal to the number of auditory beeps when these stimuli are presented within 100 ms of each other. This phenomenon demonstrates that auditory information is processed preferentially over visual information—that is, auditory input can dominate visual perception (Abadi & Murphy, 2014; Cecere, Rees, & Romei, 2015; van Erp, Philippi, & Werkhoven, 2013; Kamke, Vieth, Cottrell, & Mattingley, 2012; Kumpik, Roberts, King, & Bizley, 2014; Shams, Kamitani, & Shimojo, 2000, 2002; Whittingham, Mcdonald, & Clifford, 2014). The SIFI manifests in two forms: the fission illusion, where a single visual flash accompanied by two auditory beeps is misperceived as two flashes (Shams et al., 2000, 2002), and the fusion illusion, where two visual flashes accompanied by a single beep are misperceived as one flash (Andersen, Tiippana, & Sams, 2004). Research has consistently shown that the fission illusion produces a stronger effect than the fusion illusion (Shams et al., 2000; Wozny, Beierholm, & Shams, 2008).

Since Shams et al. (2000) first discovered the SIFI phenomenon, numerous researchers have investigated factors influencing it, with particular interest in how top-down attention modulates the illusion. Previous studies have found that the second auditory beep produces early modulations of visual cortical activity amplitude approximately 30-60 ms after presentation (Mishra, Martínez, Sejnowski, & Hillyard, 2007). Mishra, Mishra, Martínez, and Hillyard (2010) further examined how stimulus location (upper vs. lower visual field) and attention (attended vs. unattended) interactively affect the SIFI. Their results showed that the PD120/110 component (in the lower visual field) originated in ventral occipital and extrastriate visual cortex, while subsequent PD180 and ND250/240 components (in the lower visual field) were localized to the superior temporal gyrus—regions considered critical for multisensory interaction. These findings demonstrate that the SIFI is not solely the product of automatic integrative processing but also involves top-down attention in the erroneous integration of visual and auditory signals (Mishra et al., 2010). Additionally, Kamke et al. (2012) used transcranial magnetic stimulation (TMS) to investigate attention-related cortical networks in the SIFI, employing different audiovisual stimulus combinations (one visual with two auditory stimuli, one visual alone, two auditory alone) and different temporal intervals (70 ms and 160 ms). They found that when TMS was applied to the right angular gyrus (AG), participants' reports of flash illusions decreased significantly. The authors concluded that disrupting the angular gyrus enhanced veridical perception in multisensory events, effectively reducing the influence of irrelevant auditory stimuli on visual processing and decreasing the likelihood of erroneous audiovisual integration. However, TMS did not alter perception when single visual stimuli were presented or when audiovisual stimuli had longer temporal intervals. Thus, the angular gyrus appears to be involved in modulating the integration of visual and auditory stimuli in the SIFI, though this study was limited to the fission illusion (Kamke et al., 2012).

Recent research distinguishing different types of attention has revealed that endogenous attention can influence multisensory integration through modal attention (Talsma, 2015), where attention to a particular modality enhances processing within that modality while attenuating processing in unattended modalities (Spence, 2011). Studies have also found that dividing attention between two modalities incurs greater costs than attending to a single modality, costs associated with stronger activation in frontoparietal or superior temporal cortices when attention is distributed across modalities (Santangelo, Fagioli, & Macaluso, 2010). A recent study further demonstrated that dividing attention across modalities, compared to focusing on a single modality, can affect the SIFI, showing that modality-divided attention influences the illusion but only for the fission component (Yu, Wang, & Zhang, 2017). One possible explanation for how modal attention modulates sensory dominance effects is the principle of prior entry: when two stimuli are presented simultaneously, observers tend to perceive the stimulus they attended to first as having occurred earlier. This prior entry effect may arise either because observers endogenously attend to a visual or auditory pathway, or because the visual or auditory stimulus itself exogenously captures attention (Posner, Nissen, & Klein, 1976; Turatto, Benso, Galfano, & Umiltá, 2002). Therefore, when observers endogenously direct attention to the visual modality, they tend to perceive visual stimuli as preceding auditory stimuli in bimodal trials; when attention is endogenously directed to the auditory modality, they tend to perceive auditory stimuli as preceding visual stimuli. Many studies have used different methods to endogenously direct attention to one modality, with the most common approach being to manipulate the proportion of trials in different modalities. For example, Sinnett, Spence, and Soto-Faraco (2007) investigated how endogenous modal attention affects the Colavita visual dominance effect by increasing the proportion of unimodal visual targets to 60% while decreasing unimodal auditory targets to 20% (with the remaining 20% being bimodal trials). They found that the Colavita visual dominance effect increased; conversely, when unimodal visual targets comprised 20% and unimodal auditory targets comprised 60%, the effect decreased. This demonstrates that such manipulations can successfully shift endogenous attention to the more frequently occurring modality.

Building on previous research, the present study investigated whether endogenous modal attention affects auditory dominance in the SIFI by manipulating the proportion of trials in different modalities to endogenously direct attention toward the more frequently presented modality. Using the classic SIFI paradigm (Shams et al., 2002), we conducted three experiments: Experiment 1 served as a baseline with equal proportions (1:1:1) of visual, auditory, and audiovisual trials; Experiment 2 directed endogenous attention to the visual modality with proportions of 3:1:1; and Experiment 3 directed endogenous attention to the auditory modality with proportions of 1:3:1. We hypothesized that endogenously attending to visual or auditory modalities would affect the direction of sensory dominance, thereby altering the auditory-dominant SIFI effect. Specifically, when attention was endogenously directed to the visual modality, enhanced neural responses to visual information would improve behavioral accuracy in judging the number of flashes, resulting in reduced flash illusion effects. When attention was endogenously directed to the auditory modality, enhanced neural responses to auditory information would increase interference with visual judgments, thereby increasing the illusion effect.

Experiment 1: Baseline Condition

Participants

Twenty-three university students (8 male, 15 female) aged 18–26 years participated in the experiment. All had normal or corrected-to-normal vision, none had previously participated in similar experiments, and all received monetary compensation after completing the study.

Apparatus and Materials

All stimuli were presented on an iiyama MA203DT Vision Master Pro 513 monitor with a resolution of 1024×768 pixels and a refresh rate of 85 Hz. The experimental program was developed using Presentation software (Neurobehavioral Systems Inc.). Visual flash stimuli were presented on a black background as white disks (2° visual angle) appearing 5° below the central fixation point, with a duration of 17 ms. The visual stimuli were positioned in the peripheral visual field because previous research has shown that the illusion effect is maximal at this location when accompanied by auditory stimuli (Shams et al., 2002). Auditory stimuli were presented through Audio-Technica ATH-WS99 headphones at 75 dB intensity, 3.5 kHz frequency, and 7 ms duration.

Design and Procedure

In each trial, participants were presented with either a single visual flash or two consecutive visual flashes. These visual stimuli could be presented alone, accompanied by one auditory beep, or accompanied by two consecutive auditory beeps. Additionally, two consecutive visual flashes could be presented alone, with one beep, or with two beeps. Furthermore, a single auditory beep or two consecutive beeps could be presented alone. This resulted in eight experimental conditions: F1, F1B1, F1B2, F2, F2B1, F2B2, B1, and B2. For clarity, these trial types are denoted using a consistent notation: F1B2 refers to a trial with one visual flash accompanied by two auditory beeps, while F2 refers to a trial with two visual flashes and no auditory stimuli. F1 and F2 represent "unimodal visual stimuli," B1 and B2 represent "unimodal auditory stimuli," and the remaining conditions represent "bimodal audiovisual stimuli."

The experimental procedure is illustrated in Figure 1 [Figure 1: see original paper]. The first auditory beep preceded the first visual flash by 23 ms. The interval between two visual flashes was 50 ms, and the interval between two auditory beeps was 57 ms. The conditions capable of producing flash illusions were F1B2 and F2B1, where participants could either experience the illusion or respond veridically to identical physical stimuli. Based on participants' responses, these two conditions were classified post-hoc into F1B2_W (illusion present, also called fission illusion) versus F1B2_R (illusion absent, i.e., correct response) and F2B1_W (illusion present, also called fusion illusion) versus F2B1_R (illusion absent, i.e., correct response). Participants were instructed to maintain fixation on the central point throughout the experiment and judge whether they perceived one or two flashes; no response was required for unimodal auditory trials. Each participant completed 512 trials, with 64 trials per condition (F1, F1B1, F1B2, F2, F2B1, F2B2, B1, and B2) presented in random order. The inter-trial interval varied randomly between 400–700 ms in 100 ms steps.

Results and Analysis

As shown in Table 1 , accuracy was high in the F1, F2, F1B1, and F2B2 conditions, indicating that participants could judge the number of flashes accurately when visual stimuli were presented alone or when the number of visual flashes matched the number of auditory beeps.

Based on participants' responses, the two illusion-inducing conditions (F1B2 and F2B1) were classified post-hoc into F1B2_W (illusion present, i.e., fission illusion) versus F1B2_R (illusion absent, i.e., correct response) and F2B1_W (illusion present, i.e., fusion illusion) versus F2B1_R (illusion absent, i.e., correct response). Table 1 shows that accuracy in the F1B2 condition was 34%, significantly lower than in the F1, F1B1, F2, and F2B2 conditions (all ps < 0.001). Accuracy in the F2B1 condition was 62%, also significantly lower than in the F1, F1B1, F2, and F2B2 conditions (all ps < 0.001). For conditions with mismatched visual and auditory stimuli (F1B2 and F2B1), the pattern of accuracy demonstrated auditory dominance, where the number of auditory beeps influenced judgments of visual flash count.

Experiment 2: Endogenous Attention to the Visual Modality

Participants

Twenty-five university students (9 male, 16 female) aged 18-26 years participated. All had normal or corrected-to-normal vision, none had previously participated in similar experiments, and all received monetary compensation.

Design and Procedure

The apparatus and procedure were identical to Experiment 1, except that the ratio of "unimodal visual," "unimodal auditory," and "bimodal audiovisual" trials was set to 3:1:1. Participants were instructed to maintain central fixation and judge whether they perceived one or two visual flashes; no response was required for B1 and B2 conditions. Each participant completed 768 trials: 192 trials each for F1 and F2 conditions, and 64 trials each for the remaining conditions (F1B1, F1B2, F2B1, F2B2, B1, and B2), all presented randomly. The inter-trial interval varied randomly between 400-700 ms in 100 ms steps.

Results and Analysis

As shown in Table 1 , accuracy remained high in the F1, F2, F1B1, and F2B2 conditions, confirming that participants could accurately judge flash number when visual stimuli were presented alone or when visual and auditory counts matched.

Post-hoc classification of the illusion-inducing conditions (F1B2 and F2B1) yielded the same categories as in Experiment 1. Table 1 shows that accuracy in the F1B2 condition was 50%, significantly lower than in the F1, F1B1, F2, and F2B2 conditions (all ps < 0.001). Accuracy in the F2B1 condition was 68%, also significantly lower than in the F1, F1B1, F2, and F2B2 conditions (all ps < 0.001). For mismatched visual-auditory conditions, the accuracy pattern again demonstrated auditory dominance, with auditory beep count influencing visual flash judgments.

Experiment 3: Endogenous Attention to the Auditory Modality

Participants

Twenty-four university students (9 male, 15 female) aged 18–26 years participated. All had normal or corrected-to-normal vision, none had previously participated in similar experiments, and all received monetary compensation.

Design and Procedure

The apparatus and procedure were identical to Experiment 1, except that the ratio of "unimodal visual," "unimodal auditory," and "bimodal audiovisual" trials was set to 1:3:1. Participants maintained central fixation and judged whether they perceived one or two visual flashes; no response was required for B1 and B2 conditions. Each participant completed 768 trials: 192 trials each for B1 and B2 conditions, and 64 trials each for the remaining conditions (F1, F1B1, F1B2, F2, F2B1, and F2B2), all presented randomly. The inter-trial interval varied randomly between 400–700 ms in 100 ms steps.

Results and Analysis

As shown in Table 1 , accuracy was high in the F1, F2, F1B1, and F2B2 conditions, indicating accurate flash number judgments when visual stimuli were presented alone or when visual and auditory counts matched.

Post-hoc classification of F1B2 and F2B1 conditions followed the same procedure as previous experiments. Table 1 shows that accuracy in the F1B2 condition was 29%, significantly lower than in the F1, F1B1, F2, and F2B2 conditions (all ps < 0.001). Accuracy in the F2B1 condition was 64%, also significantly lower than in the F1, F1B1, F2, and F2B2 conditions (all ps < 0.001). For mismatched conditions, the accuracy pattern demonstrated auditory dominance, with auditory beep count influencing visual flash judgments.

General Discussion

Accuracy Analysis

A 3 (Experiment 1 vs. Experiment 2 vs. Experiment 3) × 2 (F1B2 vs. F2B1) repeated-measures ANOVA was conducted on accuracy rates for mismatched visual-auditory trials. The results revealed a significant main effect of experiment, $F(1, 22) = 4.01$, $p < 0.05$, $^2 = 0.16$; a significant main effect of illusion condition, $F(1, 22) = 60.73$, $p < 0.001$, $^2 = 0.73$; and a non-significant interaction, $F(1, 22) = 1.16$, $p > 0.05$. To further examine potential differences across experiments for F1B2 and F2B1 conditions, separate one-way ANOVAs were conducted (Figure 2 [Figure 2: see original paper]).

For the F1B2 condition, differences across experiments were significant, $F(1, 2) = 5.11$, $p < 0.01$, $^2 = 0.23$. Post-hoc tests showed significant differences between Experiment 1 and Experiment 2 ($p < 0.05$) and between Experiment 2 and Experiment 3 ($p < 0.005$), but not between Experiment 1 and Experiment 3 ($p > 0.05$). For the F2B1 condition, no significant differences were found across experiments (all ps > 0.05). These results indicate that modal-based endogenous attention affects the fission illusion (F1B2) in the SIFI: endogenous attention to the visual modality significantly increased accuracy (reduced illusion), while attention to the auditory modality decreased accuracy (increased illusion), though not significantly. The difference in F1B2 accuracy between visual and auditory endogenous attention conditions was significant. In contrast, the fusion illusion (F2B1) was not modulated by endogenous modal attention.

Although fission and fusion illusions are stable and universal phenomena, individual differences exist in illusion magnitude. To account for this variability, Andersen et al. (2004) introduced an odds ratio metric to quantify fission and fusion illusion effects. To calculate the odds ratio, the total number of bimodal trials capable of producing fission illusions is first determined, then the number of trials where fission actually occurred (i.e., perceived flash count exceeded actual flash count) is identified. The difference between these values represents trials without fission illusions. The ratio of fission-present to fission-absent trials is then compared to the corresponding ratio from unimodal conditions (where flash count matches that in bimodal conditions). This odds ratio provides a clearer index of illusion magnitude: values greater than 1 indicate stronger fission illusions in bimodal versus unimodal conditions, with larger values reflecting greater illusion magnitude. The same calculation applies to fusion illusions (Andersen et al., 2004). Using this algorithm, we calculated illusion magnitudes to more clearly compare effects across experimental conditions (Table 2 and Table 3). The results show that fission illusions were stronger than fusion illusions across all experiments. Additionally, the odds ratios reflect significant differences between Experiment 1 and Experiment 2 and between Experiment 2 and Experiment 3 for the F1B2 condition, but no significant differences between Experiment 1 and Experiment 3. For the F2B1 condition, no significant differences were observed across the three experiments.

Reaction Time Analysis

A 3 (Experiment: Experiment 1 vs. Experiment 2 vs. Experiment 3) × 4 (Condition: F1B2_R vs. F1B2_W vs. F2B1_R vs. F2B1_W) mixed repeated-measures ANOVA was conducted on reaction times (Table 4). The results showed significant main effects of experiment, $F(1, 2) = 23.91$, $p < 0.001$, $^2 = 0.41$, and condition, $F(3, 69) = 124.81$, $p < 0.001$, $^2 = 0.64$, as well as a significant interaction, $F(6, 69) = 2.47$, $p < 0.05$, $^2 = 0.07$. To further explore the interaction between experiment and illusion type, we compared F1B2_R versus F1B2_W and F2B1_R versus F2B1_W within each experiment.

For the fission illusion (F1B2), reaction times in the no-illusion condition (F1B2_R) were significantly longer than in the illusion condition (F1B2_W) across all experiments: Experiment 1 (889 ms vs. 839 ms), $t(22) = 4.07$, p =

0.001, d = 1.22; Experiment 2 (798 ms vs. 710 ms), $t(24) = 5.44$, $p < 0.001$, d = 1.56; and Experiment 3 (732 ms vs. 707 ms), $t(23) = 2.15$, $p < 0.05$, d = 0.63. In contrast, for the fusion illusion (F2B1), no significant differences were found between no-illusion (F2B1_R) and illusion (F2B1_W) conditions in any experiment: Experiment 1 (757 ms vs. 762 ms), $t < 1$; Experiment 2 (640 ms vs. 639 ms), $t < 1$; Experiment 3 (581 ms vs. 613 ms), $t(23) = 1.99$, $p > 0.05$.

A one-way ANOVA on the difference scores between F1B2_R and F1B2_W across experiments revealed significant differences, $F(2, 68) = 4.61$, $p < 0.05$, $^2 = 0.22$. Post-hoc tests showed no significant difference between Experiment 1 and Experiment 2 ($p > 0.05$) or between Experiment 1 and Experiment 3 ($p > 0.05$), but a significant difference between Experiment 2 and Experiment 3 ($p < 0.005$). These results indicate that the auditory-dominant fission illusion remains stable regardless of attentional manipulations, though the cost of resisting the illusion varies. For the fusion illusion, the absence of reaction time differences across conditions in all three experiments suggests that this auditory-dominant effect is not manifested at the reaction time level.

To compare changes in fission and fusion effects across different endogenous attention conditions, separate one-way ANOVAs were conducted on reaction times for F1B2_W and F2B1_W across experiments. For F1B2_W, significant differences emerged across experiments, $F(2, 69) = 17.49$, $p < 0.001$, $^2 = 0.37$, with post-hoc tests showing significant differences between Experiment 1 and Experiment 2 ($p < 0.001$) and between Experiment 1 and Experiment 3 ($p < 0.001$), but not between Experiment 2 and Experiment 3 ($p > 0.05$). For F2B1_W, significant differences were also found across experiments, $F(2, 69) = 16.94$, $p < 0.001$, $^2 = 0.35$, with post-hoc tests showing significant differences between Experiment 1 and Experiment 2 ($p < 0.001$) and between Experiment 1 and Experiment 3 ($p < 0.001$), but not between Experiment 2 and Experiment 3 ($p > 0.05$).

Overall Discussion

Using the classic SIFI paradigm (Shams et al., 2000, 2002), the present study investigated whether endogenously attending to visual or auditory modalities could modulate the auditory-dominant SIFI effect by manipulating trial proportions to direct attention toward the more frequently presented modality. The results replicated the classic SIFI phenomenon, with fission illusions being stronger than fusion illusions. Critically, endogenous attention to different sensory modalities affected the magnitude of fission illusions but not fusion illusions. Specifically, when attention was endogenously directed to the visual modality, the fission illusion decreased significantly compared to baseline (Figure 2). When attention was endogenously directed to the auditory modality, the fission illusion increased relative to baseline, though this difference was not statistically significant (Figure 2).

The decision to manipulate proportions as a between-subjects variable was based on several considerations. First, the current experimental duration was already substantial, and the low-luminance testing environment with high-contrast white flashes on a black background could induce fatigue if all three experiments were combined into one session. Second, although individual differences exist in SIFI magnitude, the phenomenon is remarkably stable, with virtually all participants showing the effect (though to varying degrees). Moreover, we have replicated Experiment 1 multiple times with different participant samples, yielding consistent results (see also Yu et al., 2017). Therefore, using between-subject proportion manipulations would not compromise the results due to individual differences. Third, in studies of the Colavita visual dominance effect, Sinnett et al. (2007) similarly used different experiments (with different participants) to set different modality proportions, demonstrating that proportion-based manipulations of endogenous modal attention can influence the magnitude of visual dominance effects.

The accuracy results from all three experiments replicated the classic auditory-dominant SIFI effect, with fission illusions being stronger than fusion illusions, consistent with previous findings (Shams et al., 2000, 2002; Shams, Ma, & Beierholm, 2005; Watkins, Shams, Tanaka, Haynes, & Rees, 2006; Watkins, Shams, Josephs, & Rees, 2007; Wozny et al., 2008; Yu et al., 2017). Additionally, we found that fission illusions were modulated by endogenous modal attention, whereas fusion illusions were not (Figure 2). Several factors may explain this dissociation. First, previous research has shown that fusion illusions are less stable than fission illusions (Shams et al., 2000; Wozny et al., 2008). The fission illusion remains robust across variations in shape, contrast, size, visual stimulus duration, frequency, intensity, auditory stimulus duration, and the relative temporal and spatial positions of visual and auditory stimuli (Shams et al., 2000, 2002; Shams et al., 2005; Watkins et al., 2006; Watkins et al., 2007). Second, fission illusions are sensitive to attentional resources, whereas fusion illusions are not. Specifically, dividing attention between visual and auditory modalities reduces fission illusion magnitude compared to focusing attention solely on the visual modality, while fusion illusions remain unaffected (Yu et al., 2017). Research indicates that attention can be directed not only to spatial locations but also to sensory modalities (Talsma, 2015), enhancing processing in the attended modality while suppressing processing in unattended modalities (Gu & Lv, 2016; Spence, 2011; Sun et al., 2011). Therefore, when attention is directed to one sensory modality, neural responses to information in that modality are enhanced while responses to the ignored modality are suppressed. The SIFI represents an auditory-dominant phenomenon in multisensory integration involving both visual and auditory modalities. In the present study, endogenously directing attention to the visual modality enhanced neural responses to visual information, resulting in improved behavioral accuracy in judging flash number and consequently reduced fission illusion magnitude (Figure 2 and Table 2). Conversely, endogenously directing attention to the auditory modality enhanced neural responses to auditory information, increasing interference with visual flash judgments, reducing accuracy, and producing a trend toward increased fission illusion magnitude (Figure 2 and Table 2).

Our proportion manipulation successfully directed endogenous attention to visual and auditory modalities, revealing that, like the Colavita visual dominance effect, the sound-induced double-flash illusion (fission illusion) is modulated by endogenous modal attention, with the effect decreasing or increasing accordingly. These results align with Sinnett et al.' s (2007) findings on endogenous modal attention in the Colavita visual dominance effect, demonstrating that proportion-based manipulations of endogenous attention can modulate sensory dominance effects. In Experiment 2, endogenous attention was directed to the visual modality; in Experiment 3, it was directed to the auditory modality. We emphasize that different sensory modalities were attended across experiments. The task required judging visual flash number, so when attention was endogenously directed to the visual modality, auditory stimuli could be ignored, constituting focused attention on a single visual modality. However, when attention was endogenously directed to the auditory modality, participants needed to attend to both auditory and visual stimuli (due to task demands), constituting divided attention across modalities. Research has shown that dividing attention across modalities incurs greater costs than focusing on a single modality, costs associated with stronger activation in frontoparietal or superior temporal cortices (Degerman et al., 2007; Santangelo et al., 2010). Our recent study also found that dividing attention across modalities affects the SIFI, demonstrating that modality-divided attention can modulate the illusion (Yu et al., 2017). One possible explanation for how modal attention modulates sensory dominance is the prior entry principle: when two stimuli are presented simultaneously, the stimulus that is attended to first is perceived as occurring earlier. This prior entry effect may result from endogenous attention to visual or auditory pathways or from exogenous attentional capture by the stimuli themselves (Posner et al., 1976; Turatto et al., 2002). Therefore, when participants endogenously attend to the visual modality, they selectively focus attention on visual input, enhancing neural responses to visual information and reducing interference from auditory stimuli when judging flash number. When participants endogenously attend to the auditory modality, they must divide attention between visual and auditory modalities, resulting in enhanced neural responses to auditory information and increased interference from auditory stimuli when judging flash number.

In the present study, endogenous attention to the visual modality can be conceptualized as reduced interference from auditory stimuli when judging visual flash number, whereas endogenous attention to the auditory modality can be conceptualized as increased interference from auditory stimuli. In multisensory integration/competition research, a critical factor is the extent to which distractor stimuli compete with target stimuli for attentional resources. According to Lavie' s (2005) perceptual load theory, under high-load conditions, participants must focus all attention on the current task. Therefore, if attention is divided across multiple sensory modalities (high-load condition), fewer attentional resources are allocated to any single modality, resulting in poorer behavioral performance (Tang, Wu, & Shen, 2016). In Experiment 2, participants attended only to the visual modality, completing the visual judgment task without auditory interference, resulting in reduced illusion magnitude. In Experiment 3, participants attended to both visual and auditory modalities, completing the visual judgment task while experiencing auditory interference, resulting in a trend toward increased illusion magnitude.

Modal-based endogenous attention can modulate the auditory-dominant effect in the sound-induced double-flash illusion (fission illusion), with visual attention reducing the illusion and auditory attention producing a trend toward increased illusion magnitude. The present study provides insight into top-down factors that can modulate sound-induced flash illusions.

References

Abadi, R. V., & Murphy, J. S. (2014). Phenomenology of the sound-induced flash illusion. *Experimental Brain Research*, 232(7), 2207–2220.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301–308.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology*, 25(2), 231–235.

Degerman, A., Rinne, T., Pekkola, J., Autti, T., Jaaskelainen, I.P., Sams, M., & Alho, K. (2007). Human brain activity associated with audiovisual perception and attention. *NeuroImage*, 34(4), 1683–1691.

Gu, J. Y., & Lv, Y. (2016). The different effects of selective attention and divided attention on multisensory integration. *Studies of Psychology and Behavior*, 14(2), 202–206. [顾吉有, 吕勇. (2016). 选择性注意和分配性注意对多感觉整合的不同影响. 心理与行为研究, 14(2), 202–206.]

Kamke, M. R., Vieth, H. E., Cottrell, D., & Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage*, 62(3), 1334–1341.

Kumpik, D. P., Roberts, H. E., King, A. J., & Bizley, J. K. (2014). Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion. *Journal of Vision*, 14(7), 1–15.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82.

Mishra, J., Martínez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience*, 22(8), 1714–1729.

Mishra, J., Martínez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, 27(15), 4120–4131.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, 83(2), 157–171.

Santangelo, V., Fagioli, S., & Macaluso, E. (2010). The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *Neurolmage*, 49(3), 2717–2727.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, 408(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147–152.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17), 1923–1927.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Attention, Perception, & Psychophysics*, 69(5), 673–686.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995.

Sun, Y-L., Hu, Z-H., Zhang, R-L., Xun, M-M., Liu, Q., & Zhang, Q-L. (2011). An investigation on the effect factors in the paradigm of multisensory integration. *Acta Psychologica Sinica*, 43(11), 1239–1246. [孙远路, 胡中华, 张瑞玲, 寻茫茫, 刘强, 张庆林. (2011). 多感觉整合测量范式中存在的影响因素探讨. 心理学报, 43(11), 1239–1246.]

Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, 9, 19.

Tang, X. Y., Wu, J. L., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, 61, 208–224.

Turatto, M., Benso, F., Galfano, G., & Umiltá, C. (2002). Nonspatial attentional shifts between audition and vision. *Journal of Experimental Psychology: Human Perception & Performance*, 28(3), 628–639.

Van Erp, J. B. F. V., Philippi, T. G., & Werkhoven, P. (2013). Observers can reliably identify illusory flashes in the illusory flash paradigm. *Experimental Brain Research*, 226(1), 73–79.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, 37(2), 572–578.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31(3), 1247–1256.

Whittingham, K. M., Mcdonald, J. S., & Clifford, C. W. G. (2014). Synesthetes show normal sound-induced flash fission and fusion illusions. *Vision Research*, 105, 1-9.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, 8(3), 1-11.

Yu, W., Wang, A. J., & Zhang, M. (2017). Effect of selective and divided attentions on auditory dominance in multisensory integration. *Acta Psychologica Sinica*, 49(2), 164-173. [于薇, 王爱君, 张明. (2017). 集中和分散注意对多感觉整合中听觉主导效应的影响. 心理学报, 49(2), 164-173.]

Note: Figure translations are in progress. See original paper for figures.