Influencing Factors and Neural Mechanisms of the Sound-Induced Flash Illusion

Authors: Wang Aijun, Huang Jie, Lu Feifei, He Jiaying, Tang Xiaoyu, Zhang Ming, Tang Xiaoyu, Zhang Ming

Date: 2020-05-25T00:00:00+00:00

Abstract

The sound-induced flash illusion is a typical phenomenon of audiovisual integration illusion, referring to the condition where, when visual flash stimuli and auditory sound stimuli are presented in unequal numbers within an interval of 100 ms, participants perceive the number of visual flashes as equal to the number of auditory sounds. The influencing factors of the sound-induced flash illusion include both bottom-up and top-down factors, as well as inter-subject difference factors such as audiovisual stimulus dependency, development of audiovisual integration, and perceptual sensitivity to audiovisual stimuli. The generation of this effect is mainly manifested in the early processing stages temporally, and primarily involves multiple cortical and related subcortical brain regions spatially. Future research should further investigate the influence of cognitive processing such as attention, reward, and audiovisual integration modes on the sound-induced flash illusion, while also examining its effect on memory and learning, and further exploring its cognitive neural mechanisms through the integration of computational models and neuroscience approaches.

Full Text

Preamble

The Influential Factors and Neural Mechanisms of Sound-Induced Flash Illusion

WANG Aijun[1], HUANG Jie[1], LU Feifei[1], HE Jiaying[1], TANG Xiaoyu[2]*, ZHANG Ming[1]*

[1](Department of Psychology, Soochow University, Research Center for Psychology and Behavioral Sciences, Suzhou 215123, China)
[2](School of Psychology, Liaoning Normal University, Liaoning Collaborative

Innovation Center for Assessment and Cultivation of Healthy Personality in Children and Adolescents, Dalian 116029, China)

Abstract: Sound-induced flash illusion (SiFI) represents a classic auditory-dominated multisensory integration phenomenon wherein participants perceive the number of visual flashes as equal to the number of auditory beeps when these stimuli are presented with incongruent counts within a 100 ms window. The influencing factors of SiFI encompass bottom-up stimulus properties, top-down cognitive processes, and between-subject differences such as dependence on audiovisual information, developmental aspects of multisensory integration, and perceptual sensitivity to audiovisual stimuli. The temporal dynamics of this illusion primarily manifest during early processing stages, engaging multiple cortical and subcortical brain regions. Future research should further investigate how cognitive processes such as attention and reward modulate SiFI, examine its impact on memory and learning, and integrate computational modeling with neuroscience approaches to elucidate its cognitive and neural mechanisms.

Keywords: sound-induced flash illusion; fusion illusion; fission illusion; multisensory integration

Our perception of the external world derives from multiple sensory systems, with stimuli across different modalities interacting to enable more effective responses through integration and competition among sensory channels. Multisensory integration refers to the process by which individuals combine information from different sensory channels (e.g., visual, auditory, tactile) into a unified, coherent, and stable meaningful percept [**?**, **?**]. One manifestation of multisensory integration is multisensory illusion, with classic examples including the McGurk effect (visual dominance) and the sound-induced flash illusion (auditory dominance). Both phenomena represent crossmodal bistable perception, where identical stimuli elicit two distinct, competing percepts [**?**, **?**]. Although vision typically dominates in most bimodal interactions, audition often prevails in the temporal dimension, as demonstrated by SiFI [**?**, **?**] and temporal ventriloquism [**?**].

First described by Ladan Shams, SiFI is a typical auditory-dominated audiovisual integration phenomenon occurring when visual flashes paired with an incongruent number of auditory beeps presented within 100 ms are illusorily perceived as numerically equivalent [**?**, **?**]. This illusion is considered both a classic multisensory integration phenomenon and a crossmodal bistable percept. Shams et al. (2000, 2002) pioneered SiFI research, demonstrating that auditory information can dominate visual processing during audiovisual integration. Their paradigm presented two auditory beeps separated by 57 ms accompanied by a single flash occurring 23 ms after the first beep, requiring participants to report the number of perceived flashes. Most participants reported seeing two flashes, a phenomenon termed fission illusion [**?**, **?**]. Subsequently, Andersen et al. (2004) adapted this paradigm, presenting multiple flashes with a single beep and asking participants to report flash count. When two flashes accompanied one beep, participants frequently misperceived them as a single flash, termed fusion illusion [?, ?].

Two primary explanations account for SiFI generation. First, subjective judgments of external stimuli alter perceived flash counts [?, ?]. Second, reduced visual perceptual sensitivity and altered decision criteria occur [?]. A crucial factor underlying these changes is causal inference—the perception of causality that auditory beeps exert on visual flash judgments [?]. Research on causal inference in multisensory integration shows that synchronously presented, non-informative auditory stimuli congruent with visual motion direction can facilitate visual processing [?]. In SiFI, Shams et al. (2005) demonstrated that beep count influences flash perception, with this process conforming to Bayesian causal inference models [?].

SiFI has remained a focal topic in audiovisual integration research since its discovery by Shams et al. (2000). However, comprehensive reviews of SiFI are currently lacking. This paper attempts a thorough review, first examining influencing factors from multiple perspectives, then detailing cognitive and neural mechanisms using various neuroscientific techniques, and finally proposing future research directions.

2.1 Factors Influencing Within-Subject Variability in Sound-Induced Flash Illusion

The magnitude of SiFI varies under certain conditions, meaning different experimental conditions produce different illusion effects within the same participants. Previous research identifies two main categories of within-subject factors: bottom-up physical stimulus properties and top-down cognitive factors.

2.1.1 Bottom-Up Physical Stimulus Factors

Spatial characteristics of stimuli influence audiovisual processing and thus SiFI magnitude. Lewald and Guski (2003) found that flash count judgments depend on both temporal and spatial disparities, with proximity in time and space being crucial for experiencing causality [?]. Innes-Brown and Crewther (2009) added spatial congruency conditions (ipsilateral vs. contralateral presentation) to Shams et al.' s (2002) classic paradigm, finding that spatial incongruency did not affect SiFI occurrence but enhanced perceptual illusion sensitivity under spatially congruent conditions [?]. Máté et al. (2015) further examined spatial congruency using dynamic continuous flash suppression (DCFS), presenting different visual stimuli to each eye (dynamic Mondrians to the suppressed eye; three gray disks to the target eye). When a flash occurred at one disk location, the auditory stimulus was presented on the corresponding screen side. Results showed that simultaneous sounds enhanced unconscious visual signals in a spatially congruent manner, with sounds affecting reported flash locations based on flash visibility—strongest spatial bias occurred for flashes judged invisible [?]. Abadi and Murphy (2014) added motion direction manipulations (congruent vs. incongruent audiovisual motion) to Shams et al.' s (2002) paradigm, finding larger fission than fusion effects and demonstrating that audiovisual stimulus combinations (spatial congruency) influenced both illusions [?]. DeLoss and Andersen (2015) increased spatial disparity from 20° to 50°, finding no effect on SiFI [?].

Stimulus intensity differences affect responses, suggesting that intermodal signal strength differences may influence SiFI magnitude. Kawabe (2009) manipulated flash luminance to examine whether fusion illusions could be generated by varying audiovisual temporal intervals. Results showed that when audiovisual intervals were close, sounds more easily captured flashes, producing fusion illusions with enhanced perceived flash brightness. Enhanced brightness was also perceived when some sounds had high frequencies. However, when intervals exceeded the capture range, stimulus attribute matching disappeared. These findings indicate that changes in sensory channel intensity can affect fusion illusion strength [?]. Andersen et al. (2004) examined whether sound properties influenced SiFI by adding two intensity levels: normal (80 dB, matching Shams' parameters) and low (10 dB, near auditory threshold). Fission illusions showed no significant differences across intensity levels, proving resistant to sound intensity changes. However, fusion illusions persisted until auditory levels approached threshold, disappearing at low intensities. Thus, fusion illusions vary with sound intensity, demonstrating that illusion phenomena change with subjective stimulus judgments when external inputs are altered [?]. Fission illusions appear stable and unaffected by sound intensity, whereas fusion illusions are less stable and vulnerable to stimulus condition changes. Additionally, Setti and Chan (2011) investigated whether complex visual stimuli affect SiFI susceptibility, finding that familiar visual stimuli significantly reduced SiFI sensitivity (fewer illusions), suggesting that visual stimulus intensity can modulate SiFI [?].

Temporal intervals constitute a crucial factor in multisensory integration [?], potentially influencing audiovisual integration processing and thus within-subject variability. Shams et al. (2002) presented one flash with multiple beeps, requiring flash count judgments. They demonstrated that SiFI reflects subjective visual perception rather than task difficulty or perceptual bias. Strong fission illusions required beep intervals below 70 ms, weakening when intervals exceeded 150 ms, showing that the temporal window between beeps affects fission perception [?, ?]. Zhang and Chen (2006) used fMRI to examine crossmodal interactions underlying fission illusions, finding that at a 25 ms audiovisual stimulus onset asynchrony (SOA), fMRI signals in visual cortex were significantly higher for illusory flashes than for real single flashes. This enhancement disappeared at 300 ms delays, indicating that fission illusions dynamically affect visual cortical activity in an SOA-sensitive manner [?]. Recently, Kostaki and Vatakis (2016) used varying SOAs to examine competition between unequal sensory inputs, finding that fission illusions produced different illusion strengths across SOAs [?].

Overall, SiFI is susceptible to physical stimulus factors, yet remains robust—particularly fission illusions, which persist despite changes in audiovisual stimulus properties. Fission effects are maximal when beep intervals are short or when visual stimuli appear in peripheral vision. Fusion illusions, however, are sometimes absent in studies, possibly due to experimental conditions or equipment, and are vulnerable to physical stimulus attributes (e.g., sound intensity) that can weaken or eliminate them. Fusion illusions occur more readily when audiovisual intervals are brief. Thus, relative to fission illusions, fusion effects are weaker and less stable.

2.1.2 Top-Down Cognitive Factors

While physical stimulus properties influence SiFI, cognitive state differences when facing identical stimuli can also produce varying illusion magnitudes—a hallmark of bistable perception where illusory and veridical percepts alternate despite constant bottom-up input. Moreover, top-down cognitive factors are more common in everyday life than physical stimulus changes. Existing research has primarily investigated attention, feedback patterns, and cognitive expectations.

Mishra et al. (2010) proposed that sound-induced visual illusions involve not only crossmodal integration but also attentional allocation. Using ERPs with Shams et al.' s (2002) paradigm, they manipulated stimulus location (upper vs. lower visual field) and attention (attended vs. unattended), requiring flash count judgments in designated fields. Results showed that for lower-field stimuli, the PD120/PD110 component localized to ventral occipital and extrastriate visual cortex, while subsequent PD180 and ND250/ND240 components localized to the superior temporal gyrus—regions implicated in multisensory interactions. These ERP components closely resembled those found in Mishra et al.' s (2007) SiFI study and were enhanced by attentional allocation. These findings demonstrate that attentional allocation influences crossmodal interactions and SiFI, indicating that fission illusions result not solely from automatic integration but also from attentional involvement in audiovisual signal combination [**?**]. Beyond attentional allocation, endogenous attention also modulates SiFI. Georgios and Julian (2018) examined endogenous attention effects, finding that under high working memory load, participants perceived more flash illusions independent of response bias. This provides evidence that audiovisual integration can be modulated by cognitive resource availability, enhancing non-speech audiovisual integration under reduced attentional resources and supporting the importance of top-down attentional control in multisensory integration [**?**]. Similarly, Yu et al. (2017) used the classic SiFI paradigm to manipulate attentional resource allocation, examining how active attention to auditory stimuli affects SiFI. Results showed that fission illusions were influenced by attentional allocation degree, whereas fusion illusions were not [**?**]. Recently, Zhang et al. (2018) directed endogenous attention modality-specifically to visual or auditory channels, finding that attention to the visual channel significantly reduced fission illusion magnitude, while attention to the auditory channel increased it. This demonstrates that modality-based endogenous attention can modulate fission but not fusion illusions in SiFI [?].

Rosenthal et al. (2009) used Shams et al.’ s (2002) paradigm to investigate whether feedback could influence SiFI by providing participants with accuracy feedback during the task. Results showed that neither correct nor incorrect feedback significantly reduced or eliminated the illusion, indicating SiFI’ s stability. However, when monetary rewards were provided based on accuracy, the illusory dominance effect showed a decreasing trend [?]. Van Erp et al. (2013) argued that since neuroimaging reveals differential activation in visual cortex for illusory versus real flashes, feedback might reduce perceived illusions despite non-significant group differences. They criticized previous studies for overlooking potential differences in contrast, intensity, and duration between real and illusory flashes, erroneously assuming similar, indistinguishable percepts. They proposed adding a third response option: “something different from one or two flashes.” Results showed that 44% of type-3 responses corresponded to illusory flashes versus 4% for real flashes, suggesting participants could distinguish them based on intensity, contrast, or duration characteristics. Additionally, conflict detection may influence judgments, as participants might notice “conflict” between visual and auditory pulses rather than qualitative differences between real and illusory flashes. In congruent trials, flash and beep timing and numerosity matched, whereas in illusory trials, they differed (23 ms audiovisual interval; beep count flash count). Thus, despite instructions to base responses solely on visual perception, participants’ conflict detection manifested in their responses, meaning judgments may rely on intersensory conflict rather than qualitative flash characteristics [?]. These studies indicate that direct accuracy feedback does not significantly affect SiFI, though increasing response options can alter perceived illusions. It remains uncertain what specific criteria participants use for discrimination and whether conflict detection truly influences responses, warranting further investigation.

Beyond attention and feedback, research has examined top-down cognitive expectations. Studies manipulated expectations through instructions and trial probability ratios. Experiments 1 and 2 used three instruction conditions (“report flash count,” “80% two-flash trials, report flash count,” “80% one-flash trials, report flash count”) with different trial ratios; Experiment 3 used three identical instructions (“report flash count,”“all conditions equally probable,”“all conditions equally probable”) with trial ratios matching Experiment 1. Results showed that when instruction probabilities matched actual trial ratios, judgment accuracy improved and response times decreased. When they mismatched, instructions still influenced accuracy and RTs for fission illusions. This demonstrates that top-down cognitive expectations can significantly reduce fission illusions and accelerate judgments, with less pronounced effects on fusion illusions, possibly due to their relative instability [?].

In summary, within-subject SiFI variability research has examined both bottom-up physical stimulus factors and top-down cognitive factors. Physical factors primarily reflect reduced visual perceptual sensitivity under different stimulus conditions [?]. For instance, Kumpik et al. (2014) reduced visual spatial reliability to examine how visual sensitivity affects SiFI, finding more frequent illusions in peripheral vision with greater fission than fusion magnitudes, suggesting SiFI depends on perceptual sensitivity to flashes in the visual field [?]. Top-down cognitive factors primarily reflect how subjective judgments of external inputs alter perceived flash counts. Shams et al. (2002) argued that SiFI reflects subjective visual perception rather than perceptual biases from task difficulty or physical stimulus differences [?].

2.2 Factors Influencing Between-Subject Differences in Sound-Induced Flash Illusion

SiFI magnitude varies across individuals and groups. Mishra et al. (2007) found substantial individual differences, with illusion occurrence rates ranging from 3% to 86% [?]. Three main factors contribute to these differences: dependence on audiovisual information, development of audiovisual integration, and perceptual sensitivity to audiovisual stimuli.

2.2.1 Dependence on Audiovisual Stimuli

Individual differences in SiFI may stem from differential reliance on visual versus auditory information. Visually dependent individuals should be less susceptible to auditory influence (smaller illusions), while auditorily dependent individuals should be more susceptible (larger illusions). One study compared musicians and non-musicians, finding that musically trained individuals processed audiovisual cues faster and more accurately, showed enhanced SiFI susceptibility, and had narrower temporal binding windows—likely due to enhanced auditory processing from long-term training [?]. Another study found that individuals with early monocular vision loss showed reduced SiFI magnitude compared to normal binocular observers [?], demonstrating adaptive compensation in multisensory integration. When unisensory information is unreliable, the brain can enhance perception by integrating multisensory inputs. For monocular individuals, when visual input is reduced, auditory information can provide additional environmental information [?].

Studies of special populations (synesthetes, migraine patients, multiple sclerosis patients, amblyopes, autism spectrum disorder patients, mild cognitive impairment patients, schizophrenia patients) support this view. Whittingham et al. (2014) found that synesthetes showed reduced SiFI magnitude with age [?]. Brighina et al. (2015) found that migraine patients, especially during attacks, showed reduced or absent fission illusions when one flash paired with two beeps [?]. Yalachkov et al. (2019) found that multiple sclerosis patients perceived more flash illusions than healthy controls [?]. Narinesingh et al. (2017) found that for fission illusions with auditory-leading asynchrony, normal controls showed reduced illusion strength with increasing asynchrony, whereas amblyopes maintained constant illusion strength (wider temporal binding window).

For fusion illusions with visual-leading asynchrony, amblyopes showed reduced illusion strength [?]. Bao et al. (2017) found that autism spectrum disorder patients and typically developing individuals both showed fission illusion sensitivity, but autism patients were more prone to fusion illusions [?]. Chan et al. (2015) found that mild cognitive impairment patients perceived more illusions than healthy controls at longer auditory intervals [?]. Vanes et al. (2016) found that schizophrenia patients showed significantly reduced fusion illusion magnitude [?]. Similarly, Haß et al. (2017) found differences in audiovisual integration between schizophrenia patients and healthy controls using varying intervals to induce fission illusions [?]. Recently, Gieseler et al. (2018) examined differences between hearing aid users and non-users with mild hearing loss, finding that both mild hearing loss and hearing aid use affected illusion magnitude, possibly because hearing aid use reverses the effects of hearing loss on audiovisual integration [?].

Overall, group comparisons demonstrate that differential dependence on visual and auditory information during cognitive processing influences individual SiFI differences. Specifically, visually dependent individuals show reduced auditory-driven SiFI magnitude, while auditorily dependent individuals show enhanced auditory dominance effects. However, beyond dependence differences, other factors may contribute to illusion magnitude across groups. Future research should examine additional populations to provide further evidence for the dependence hypothesis and, more importantly, quantify and investigate how dependence on audiovisual information influences SiFI in these populations.

2.2.2 Development of Audiovisual Integration

The brain employs a temporal binding window to effectively integrate crossmodal information. Research shows that this window widens with age to maximize use of multisensory input as sensory transduction declines. DeLoss et al. (2013) used SiFI to examine age differences, finding that older adults showed reduced ability to suppress irrelevant crossmodal information and enhanced integration ability—manifesting as more frequent illusions. This suggests a dissociation between multisensory integration and aging, with older adults being more susceptible to multisensory integration effects [?]. McGovern et al. (2014) found that older adults experienced fission illusions even at large stimulus onset asynchronies (SOAs), but showed no significant age differences in fusion illusion magnitude [?]. Similarly, Chan et al. (2017) found that older adults were more susceptible to SiFI than younger adults at longer SOAs [?]. Chan et al. (2018) subsequently examined young and older adults with long (±70, ±110, ±150, ±190, ±230 ms) and short (±70, ±150, ±230 ms) SOA ranges, finding both groups showed higher susceptibility to illusions with small asynchronies but only within longer temporal windows. Changing SOA magnitude also affected sensitivity to perceived illusions [?].

Recently, Hernández et al. (2019) provided the first large-scale cross-sectional evidence on SiFI, sampling 3,955 adults over 50 to examine age, cognitive status, and gender effects on SiFI susceptibility. Consistent with previous research, aging and lower Montreal Cognitive Assessment (MoCA) scores correlated with higher SiFI sensitivity, supporting the role of aging in multisensory processing and expanded temporal binding windows [?]. O' Brien et al. (2017) tested older adults' audiovisual integration using SiFI before and after exercise (60-80 minutes), finding that healthy, regularly exercising older adults showed more efficient multisensory processing [?]. Merriman et al. (2015) investigated whether balance function improvement correlates with enhanced multisensory processing in fall-prone older adults. After balance training 76 healthy fall-prone older adults, they found correlations between balance improvement and enhanced multisensory integration, suggesting that balance training interventions can improve postural control and produce more effective multisensory processing in fall-prone older adults [?].

These studies indicate that age-related differences in audiovisual integration temporal windows affect SiFI magnitude. But what causes these window differences between groups? Cecere et al. (2015) noted that SiFI's integration window is approximately 100 ms, matching the alpha oscillation cycle. They proposed that alpha oscillations modulate the temporal window for audiovisual event integration, finding positive correlations between individual alpha frequency (IAF) peak and illusion window size. The illusion window varied with alpha oscillatory activity, suggesting that flash illusions result from abrupt changes in visual cortex excitability induced by successive sounds within a critical window for rapid visual processing [?]. Keil and Senkowski (2017) extended these findings, showing negative correlations between IAF and illusion rates (lower IAF = stronger illusions) and implicating the calcarine sulcus in visual cortex representation [?].

In summary, stronger multisensory integration ability correlates with larger SiFI magnitude. Older adults show greater fission and fusion illusion effects than younger adults [?], likely due to enhanced multisensory integration. Age-related enhancement may reflect compensatory mechanisms where stronger multisensory processing compensates for declining unisensory processing [?]. When unisensory information becomes unreliable, the brain enhances perception through multisensory integration. For older adults, when visual channel information declines, auditory information can provide compensatory environmental information [?].

2.2.3 Perceptual Sensitivity to Audiovisual Stimuli

As a multisensory integration phenomenon, SiFI may be affected by differences in early perceptual processing of visual and auditory stimuli. Building on Shams et al. (2002), research examined how perceptual sensitivity affects SiFI, finding that lower perceptual sensitivity correlates with larger illusion magnitude [?]. Kumpik et al. (2014) investigated spatial factors in auditory influence on visual perception, manipulating spatial congruency, sound localization cues, visual eccentricity, and size to determine how visual sensitivity determines flash illusion occurrence. Signal detection analyses revealed more frequent SiFI in peripheral vision with greater fission than fusion magnitudes [?].

Given this relationship, De Haas et al. (2012) proposed that individual differences in brain structure underlie SiFI susceptibility. They found that smaller gray matter volume in early visual cortex (BA17&18) correlated with greater illusion proneness, regardless of whether visual stimuli appeared in upper or lower visual fields [?]. Additionally, gamma-band oscillations (>30 Hz) play important roles in multisensory processing, with GABA neurotransmission in the superior temporal sulcus (STS) contributing to gamma-band oscillations (GBO) via metabotropic glutamate receptor activation [?]. Balz et al. (2016) therefore proposed that GABA and glutamate system differences might contribute to individual differences in multisensory processing. Using EEG and MRI, they examined relationships among GABA/glutamate concentrations in STS, local GBO, and SiFI magnitude. Results showed that GABA levels shaped individual differences in audiovisual perception by modulating GBO, with strong positive correlations between STS GABA concentration and GBO power, and between GABA concentration and individual susceptibility to flash illusions [?].

In summary, perceptual sensitivity in visual and auditory channels influences SiFI, with reduced visual sensitivity producing larger illusions. This sensitivity relates to brain structural differences. Given these findings, Sun et al. (2020) examined how repeated auditory stimulus presentation before the classic SiFI paradigm affects auditory sensitivity and consequently illusion magnitude. They found that auditory repetition reduced auditory sensitivity, thereby decreasing illusion magnitude [?]. Thus, changes in visual and auditory perceptual sensitivity are considered primary causes of SiFI [?, ?]. Current research has examined sensitivity changes in single sensory channels, but future studies should investigate how relative sensitivity changes across both channels affect SiFI magnitude.

3.1 Temporal Mechanisms of Sound-Induced Flash Illusion

After receiving audiovisual input, the brain integrates this information. How does the brain process these stimuli at different stages to produce different illusions? Researchers have addressed this question using high-temporal-resolution techniques (EEG, ERP, MEG), with results consistently showing that audiovisual integration in SiFI occurs during early processing stages.

Shams et al. (2001) used ERPs and visual evoked potentials (VEPs) to examine whether the crossmodal integration underlying fission illusions occurs in modality-specific visual pathways. Comparing VEPs with and without accompanying sounds, they found crossmodal effects in occipital cortex, with fission illusion VEPs resembling those from real flashes. This suggests similar neural mechanisms in visual cortex for both percepts [?]. Shams et al. (2002) further investigated SiFI neural correlates using gamma-band (>30 Hz) oscillatory responses, finding significantly higher gamma-band activity on illusion trials versus non-illusion trials, with audiovisual interactions occurring only on illusion trials. These results indicate that sounds can modulate visual cortex processing, with modulation strength affecting perceptual output [?]. Shams et al. (2005) used MEG to examine sound-to-visual modulation mechanisms, finding occipitoparietal cortex modulation 35-65 ms after visual stimulus onset, with occipital, parietal, and anterior regions also modulated shortly after stimulation (~150 ms) [?]. Mishra et al. (2007) used ERPs to examine SiFI neural correlates, combining different numbers of visual (0, 1, 2) and auditory (0, 1, 2) stimuli, including two configurations for one flash with two beeps ("beep-flash-beep" vs. "beep-beep-flash"). They found early modulation of visual cortex activity amplitude 30-60 ms after the second beep during fission illusions [?]. Mishra et al. (2008) subsequently examined fusion illusion processing characteristics, finding early modulation of visual cortex activity amplitude 80-112 ms after the second flash [?]. Balz et al. (2016) recorded EEG from schizophrenia (SCZ) and healthy control (HC) groups, finding that multiple auditory stimuli paired with single visual stimuli could induce illusory multiple flashes. ERPs showed reduced amplitude in SCZ ~135 ms post-stimulus, indicating altered multisensory processing. Additionally, neural oscillation analyses revealed altered 25-35 Hz activity in SCZ occipital regions 100-150 ms post-stimulus [?]. Recently, Kaiser et al. (2019) used logistic regression to examine relationships between pre-stimulus oscillatory brain activity and crossmodal influences in SiFI. They found that increased power in occipital electrodes at 25-41 Hz 0.17-0.05 s before stimulus onset predicted subsequent illusory perception, with higher power on illusion trials. This suggests that trial-by-trial oscillatory activity predicts multisensory signal integration, providing further evidence that pre-stimulus oscillations govern multisensory processing [?].

3.2 Brain Regions in Sound-Induced Flash Illusion

Another critical question concerns which brain regions process audiovisual information in SiFI and how they operate. Researchers have used high-spatial-resolution techniques (fMRI, tDCS, TMS, MEG) to investigate this, revealing that audiovisual integration in SiFI engages extensive regions including occipital visual cortex, superior temporal sulcus, superior colliculus, prefrontal cortex, and cerebellum.

Watkins et al. (2006) used retinotopic mapping to examine whether fission illusions could occur in early visual cortex areas. They found that sounds altered perception to reflect subjective rather than physical stimuli, with fission illusions showing higher activation in primary visual cortex (V1) [?]. Conversely, their fusion illusion study found lower V1 activation for illusory flashes [?]. Whole-brain analyses additionally revealed activation in superior temporal sulcus and superior colliculus. Zhang and Chen (2006) used dynamic fMRI to confirm that fission illusions occurred in early visual cortex, with effect magnitude dynamically related to audiovisual intervals, suggesting involvement of higher-level processing beyond retinotopic visual perception [?]. Jiang and Han (2007) used fMRI to examine SiFI neural mechanisms based on perceptual gain and loss.

When auditory count exceeded flash count, participants experienced illusory flashes (perceptual gain); when auditory count was fewer, perceived flashes decreased (perceptual loss). Results showed that perceptual gain in illusory flashes activated the supramarginal gyrus, prefrontal cortex, and cerebellum—networks associated with working memory integration—while perceptual loss activated medial occipital cortex and thalamus, regions related to early visual processing [**?**]. In structural MRI, De Haas (2012) examined whether individual SiFI differences relate to gray matter volume differences, finding that smaller gray matter volume in early visual cortex (BA17&18) correlated with greater illusion proneness [**?**].

Bolognini et al. (2011) used tDCS to alter audiovisual interaction likelihood, finding that increasing or decreasing cortical excitability could enhance or reduce illusion magnitude, with temporal and occipital cortical regions (except posterior parietal cortex) functionally related to SiFI [**?**]. Recently, Maccora et al. (2019) applied cathodal tDCS to visual cortex in migraine patients to restore physiological sensitivity to fission illusions, but found tDCS could not reliably modulate SiFI, possibly due to lack of effect on cortical excitability in migraine patients [**?**]. Kamke et al. (2012) used TMS to show that parietal networks including the angular gyrus integrate auditory and visual stimuli in fission illusions [**?**]. Shams et al. (2005) used MEG to study sound-to-visual modulation, finding occipitoparietal cortex modulation 35-65 ms after visual onset, with occipital, parietal, and anterior regions also modulated shortly after stimulation (~150 ms) [**?**]. Keil et al. (2013) used MEG to find enhanced beta-band phase synchrony between left middle temporal gyrus and auditory regions before illusion generation, with reduced synchrony in visual regions [**?**]. Chan et al. (2017) recorded MEG during SiFI tasks, finding that older adults showed enhanced prestimulus beta-band activity during fission illusions, indicating predictive processing of upcoming stimuli. Transfer entropy analysis and dynamic causal modeling of prestimulus MEG data revealed stronger illusion-related modulation of crossmodal connectivity between auditory and visual cortices in older versus younger adults [**?**].

Current SiFI neural mechanism research has focused primarily on unisensory cortices like primary visual and auditory cortex. However, accumulating evidence reveals that SiFI involves not only unisensory cortices but also interactions among multiple brain regions, further clarifying and developing our understanding of SiFI neural mechanisms (Table 1). However, Table 1 reveals inconsistencies in experimental designs and parameters across SiFI neural mechanism studies, leading to some divergent results. Future research should integrate findings to further examine SiFI neural mechanisms, potentially combining functional connectivity analyses with resting-state and structural data for predictive analyses to explore SiFI generation mechanisms.

4 Summary and Future Directions

The sound-induced flash illusion reflects auditory influence on visual perception during audiovisual integration. This paper systematically reviews SiFI research, covering: (1) factors influencing within-subject variability, including bottom-up physical stimulus factors (spatial characteristics, intensity differences, temporal intervals) and top-down cognitive factors (attentional allocation, endogenous attention, feedback patterns, cognitive expectations), plus between-subject factors including audiovisual dependence, integration development, and perceptual sensitivity; (2) cognitive and neural mechanisms, including early processing stages examined via EEG/ERP/MEG and widespread brain regions (occipital visual cortex, superior temporal sulcus, superior colliculus, prefrontal cortex, cerebellum) identified via MRI, tDCS, TMS, and MEG. Since SiFI' s discovery, extensive research has accumulated, and this review not only provides systematic understanding of SiFI and its mechanisms but also offers insights for related audiovisual integration studies.

Current SiFI research has limitations. Most studies focus on intrinsic mechanisms—influencing factors and neural mechanisms separately. What remains unresolved is comparison with other audiovisual integration phenomena to explore commonalities and SiFI-specific characteristics, which would provide more empirical support for multisensory integration research. Based on these limitations, we propose future research directions:

First, attentional effects on SiFI. Future studies could systematically manipulate stimulus competition levels, for example by manipulating perceptual load to examine top-down attention effects on multisensory integration [**?**]. Building on existing multisensory integration-attention interaction research [**?**], different attentional modes' effects on SiFI could be explored. This would provide empirical evidence for attention as a within-subject factor and support for sensory dominance effects under different modality weightings in audiovisual integration research.

Second, reward effects on SiFI. Rosenthal et al. (2009) found that monetary rewards could reduce perceived illusory flashes [**?**], but the underlying mechanism remains unclear. Why does reward alter perception? Is it related to reward circuitry? Do high versus low rewards differentially affect illusion perception? Future research could manipulate reward magnitude to examine its impact on SiFI, providing empirical evidence for reward as a within-subject factor and support for how reward influences audiovisual integration through auditory dominance effects.

Third, audiovisual integration mode effects on SiFI. Previous research assumes that illusions arise from differential weighting of visual and auditory channels during integration, leading to auditory dominance. However, it remains uncertain whether SiFI results from channel dominance during integration or from differential processing within each channel before integration. Future research should examine the roles of unimodal visual and auditory processing in SiFI, clarifying whether visual and auditory inputs are integrated before further cortical processing or whether unimodal processing occurs pre-integration. This would also examine commonalities and differences between SiFI and other audiovisual integration phenomena.

Fourth, SiFI effects on cognitive processes like memory and learning. Current SiFI research focuses on influencing factors and neural mechanisms, rarely examining impacts on subsequent cognitive processing. Quak et al. (2015) suggested that multisensory processing could illuminate how working memory stores and manipulates information [**?**], and related studies have confirmed that multisensory integration facilitates working memory encoding and retrieval [**?**]. Future research could examine how SiFI, as an auditory-dominant integration phenomenon, affects working memory and learning. This would investigate commonalities and differences between SiFI and other audiovisual integration phenomena while expanding SiFI research perspectives.

Fifth, applying computational models to SiFI neural mechanisms. Previous research has explained classic visual dominance effects (McGurk effect) using computational models like hierarchical predictive coding [**?**] and the noisy encoding of disparity model (NED) [**?**]. These approaches assume specific processing stages and use parameters to describe them, with parameters corresponding to specific processes [**?**]. However, few SiFI studies have combined neurophysiological techniques with computational models. One existing SiFI computational model used a simple two-layer visual-auditory neural network to examine audiovisual interactions [**?**]. Future SiFI modeling research could integrate hierarchical predictive coding and NED models, using neuroscientific measures from EEG/ERP/fMRI for parameter fitting, or use computational models to identify brain regions corresponding to model parameters, providing neural substrates for model parameters.

References

[1] Stein, B. E., & Stanford, T. R. Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 2008, 9: 255-266. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1038/nrn2331. DOI: 10.1038/nrn2331.

[2] 文小辉, 李国强, 刘强. 视听整合加工及其神经机制. 心理科学进展, 2011, 19(7), 976-982. (WEN Xiao-Hui, LI Guo-Qiang, LIU Qiang. Processing of Audiovisual Integration and Its Neural Mechanism. *Advances in Psychological Science*, 2011, 19(7): 976-982. Retrieved on May 22, 2020, from http://journal.psych.ac.cn/xlkxjz/CN/Y2011/V19/I7/976. DOI: 10.3724/SP.J.1042.2011.00976.)

[3] Sterzer, P., Kleinschmidt, A., & Rees, G. The neural bases of multistable perception. *Trends in Cognitive Sciences*, 2009, 13(7): 310-318. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.tics.2009.04.006. DOI: 0.1016/j.tics.2009.04.006.

[4] Wang, M., Arteaga, D., & He, B. J. Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(35): E3350-3359. Retrieved on May 22, 2020, from https://sci-hub.im/10.1073/pnas.1221945110. DOI: 10.1073/pnas.1221945110.

[5] Shams, L., Kamitani, Y., & Shimojo, S. What you see is what you hear. *Nature*, 2000, 408(6814): 788. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1038/35048669. DOI: 10.1038/35048669.

[6] Shams, L., Kamitani, Y., & Shimojo, S. Visual illusion induced by sound. *Brain Research Cognitive Brain Research*, 2002, 14(1): 147-152. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/S0926-6410(02)00069-1. DOI: 10.1016/S0926-6410(02)00069-1.

[7] Bertelson, P., & Aschersleben c, G. Temporal ventriloquism: Cross-modal interaction on the time dimension: Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, 2003, 50(1-2): 147-155. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/S0167-8760(03)00130-2. DOI:10.1016/S0167-8760(03)00130-2.

[8] Abadi, R. V., & Murphy, J. S. Phenomenology of the sound-induced flash illusion. *Experimental Brain Research*, 2014, 232(7): 2207-2220. Retrieved on May 22, 2020, from https://sci-hub.im/10.1007/s00221-014-3912-2. DOI: 10.1007/s00221-014-3912-2.

[9] Andersen, T. S., Tiippana, K., & Sams, M. Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 2004, 21(3): 301-308. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.cogbrainres.2004.06.004. DOI: 10.1016/j.cogbrainres.2004.06.004.

[10] McCormick, D. & Mamassian, P. "What does the illusory-flash look like?" *Vision Research* 2008, 48(1): 63-69. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.visres.2007.10.010. DOI: 10.1016/j.visres.2007.10.010.

[11] Shams, L., & Beierholm, U. Causal inference in perception. *Trends in Cognitive Sciences*, 2010, 14: 425-432. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.tics.2010.07.001. DOI: 10.1016/j.tics.2010.07.001.

[12] Kim, R., Peters, M. A. K., & Shams, L. 0 + 1 > 1: How adding noninformative sound improves performance on a visual task. *Psychological Science*, 2011, 23(1): 6-12. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1177/0956797611420662. DOI: 10.1177/0956797611420662.

[13] Shams, L., Ma, W. J., & Beierholm, U. Sound-induced flash illusion as an optimal percept. *Neuroreport*, 2005, 16(17): 1923-1927. Retrieved on May

22, 2020, from https://sci-hub.im/10.1097/01.wnr.0000187634.68504.bb. DOI: 10.1097/01.wnr.0000187634.68504.bb.

[14] Lewald, J., & Guski, R. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, 2003, 16(3): 468-478. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/S0926-6410(03)00074-0. DOI: 10.1016/S0926-6410(03)00074-0.

[15] Innes-Brown, H., & Crewther, D. The impact of spatial incongruence on an auditory-visual illusion. *PLoS ONE*, 2009, 4(7): e6450. Retrieved on May 22, 2020, from https://sci-hub.im/10.1371/journal.pone.0006450. DOI: 10.1371/journal.pone.0006450.

[16] Máté, A., Anette, G., Verena, C., Masataka, W., & Uta, N. A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience*, 2015, 9: 1-8. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3389/fnint.2015.00016. DOI: 10.3389/fnint.2015.00016.

[17] DeLoss, D. J., & Andersen, G.J. Aging, spatial disparity, and the sound-induced flash illusion. *PLoS ONE*, 2015, 10(11): e0143773. Retrieved on May 22, 2020, from https://sci-hub.im/10.1371/journal.pone.0143773. DOI: 10.1371/journal.pone.0143773.

[18] Kawabe, T. Audiovisual temporal capture underlies flash fusion. *Experimental Brain Research*, 2009, 198(2-3): 195-208. Retrieved on May 22, 2020, from https://sci-hub.im/10.1007/s00221-009-1877-3. DOI: 10.1007/s00221-009-1877-3.

[19] Setti, A., & Chan, J. S. Familiarity of objects affects susceptibility to the sound-induced flash illusion. *Neuroscience Letters*, 2011, 492(1): 19-22. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.neulet.2011.01.042. DOI: 10.1016/j.neulet.2011.01.042.

[20] Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, 2010, 14(9): 400-410. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.tics.2010.06.008. DOI: 10.1016/j.tics.2010.06.008.

[21] Zhang, N., & Chen, W. A dynamic fMRI study of illusory double-flash effect on human visual cortex. *Experimental Brain Research*, 2006, 172(1): 57-66. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1007/s00221-005-0304-7. DOI: 10.1007/s00221-005-0304-7.

[22] Kostaki, M., & Vatakis, A. Crossmodal binding rivalry: A "race" for integration between unequal sensory inputs. *Vision Research*, 2016, 127: 165-176. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.visres.2016.08.004.

DOI: 10.1016/j.visres.2016.08.004.

[23] Mishra, J., Martínez, A., & Hillyard, S. A. Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience*, 2010, 22(8): 1714-1729. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1162/jocn.2009.21295. DOI: 10.1162/jocn.2009.21295.

[24] Georgios, M., & Julian, K. High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports*, 2018, 8(1): 11530. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1038/s41598-018-30007-6. DOI: 10.1038/s41598-018-30007-6.

[25] 于薇, 王爱君, 张明. 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, 2017, 49(2): 164-173. (YU Wei, WANG Aijun, ZHANG Ming. Effect of selective and divided attentions on auditory dominance in multisensory integration. *Acta Psychologica Sinica*, 2017, 49(2): 164-173. Retrieved on May 22, 2020, from http://journal.psych.ac.cn/xlxb/CN/Y2017/V49/I2/164. DOI: 10.3724/SP.J.1041.2017.00164.)

[26] 张明, 唐晓雨, 于薇, 宁波, 王智楠, 王爱君. 基于通道的内源性注意对声音诱发闪光错觉的影响. 心理学报, 2018, 50(11): 1212-1221. (ZHANG Ming, TANG Xiaoyu, YU Wei, NING Bo, WANG Zhinan, WANG Aijun. The Effects of modal-based endogenous attention on sound-induced flash illusion. *Acta Psychologica Sinica*, 2018, 50(11): 1212-1221. Retrieved on May 22, 2020, from http://journal.psych.ac.cn/xlxb/CN/Y2018/V50/I11/1212. DOI: 10.3724/SP.J.1041.2018.01212.)

[27] Rosenthal, O., Shimojo, S., & Shams, L. Sound-induced flash illusion is resistant to feedback training. *Brain Topography*, 2009, 21(3-4): 185-192. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1007/s10548-009-0090-9. DOI: 10.1007/s10548-009-0090-9.

[28] Van Erp, J. B. F. V., Philippi, T. G., & Werkhoven, P. Observers can reliably identify illusory flashes in the illusory flash paradigm. *Experimental Brain Research*, 2013, 226(1): 73-79. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1007/s00221-013-3413-8. DOI: 10.1007/s00221-013-3413-8.

[29] Wang, A., Sang, H., He, J., Clara, S., Tang, X., & Zhang, M. Effects of cognitive expectation on sound-induced flash illusion. *Perception*, 2019, 48(12): 1214-1234. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1177/0301006619885796. DOI: 10.1177/0301006619885796.

[30] Kumpik, D. P., Roberts, H. E., King, A. J., & Bizley, J. K. Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion. *Journal of Vision*, 2014, 14(7): 12. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1167/14.7.12. DOI: 10.1167/14.7.12.

[31] Mishra, J., Martinez, A., Sejnowski, T. J. & Hillyard, S. A. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, 2007, 27(15): 4120–4131. Retrieved on May 22, 2020, from https://sci-hub.im/10.1523/JNEUROSCI.4912-06.2007. DOI: 10.1523/JNEUROSCI.4912-06.2007.

[32] Bidelman, G. M. Musicians have enhanced audiovisual multisensory binding: Experience-dependent effects in the double-flash illusion. *Experimental Brain Research*, 2016, 234(10): 3037–3047. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1007/s00221-016-4705-6. DOI: 10.1007/s00221-016-4705-6.

[33] Moro, S. S., & Steeves, J. K. E. Normal temporal binding window but no sound-induced flash illusion in people with one eye. *Experimental Brain Research*, 2018, 236(6): 1825–1834. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1007/s00221-018-5263-x. DOI: 10.1007/s00221-018-5263-x.

[34] Klatzky, R. L., Marston, J. R., Giudice, N. A., Golledge, R. G., & Loomis, J. M. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology: Applied*, 2006, 12(4): 223–232. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1037/1076-898X.12.4.223. DOI: 10.1037/1076-898X.12.4.223.

[35] Whittingham, K. M., Mcdonald, J. S., & Clifford, C. W. G. Synesthetes show normal sound-induced flash fission and fusion illusions. *Vision Research*, 2014, 105: 1–9. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.visres.2014.08.010. DOI: 10.1016/j.visres.2014.08.010.

[36] Brighina, F., Bolognini, N., Cosentino, G., Maccora, S., Paladino, P., Baschi, R., Vallar, G., & Fierro, B. Visual cortex hyperexcitability in migraine in response to sound-induced flash illusions. *Neurology*, 2015, 84(20): 2057–2061. Retrieved on May 22, 2020, from https://sci-hub.im/10.1212/WNL.0000000000001584. DOI: 10.1212/WNL.0000000000001584.

[37] Yalachkov, Y., Bergmann, H. J., Soyda¸s, D., Buschenlange, C., Fadai Motlagh, L. Y., Naumer, M. J., ···Gehrig, J. Cognitive impairment in multiple sclerosis is reflected by increased susceptibility to the sound-induced flash illusion. *Frontiers in Neurology*, 2019, 10: 373. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3389/fneur.2019.00373. DOI: 10.3389/fneur.2019.00373.

[38] Narinesingh, C., Goltz, H. C., & Wong, A. M. F. Temporal binding window of the sound-induced flash illusion in amblyopia. *Investigative Opthalmology & Visual Science*, 2017, 58(3): 1442–1448. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1167/iovs.16-21258. DOI: 10.1167/iovs.16-21258.

[39] Bao, V. A., Doobay, V., Mottron, L., Collignon, O., & Bertone, A. Multisensory integration of low-level information in autism spectrum disorder: Measuring susceptibility to the flash-beep illusion. *Journal of Autism and Developmental Disorders*, 2017, 47(8): 2535–2543. Retrieved on May 22, 2020, from https://sci-hub.im/10.1007/s10803-017-3172-7. DOI: 10.1007/s10803-017-3172-7.

[40] Chan, J. S., Kaiser, J., Brandl, M., Matura, S., Prvulovic, D., Hogan, M., & Naumer, M. J. Expanded temporal binding windows in people with mild cognitive impairment. *Current Alzheimer Research*, 2015, 12(1): 61–68. Retrieved on May 22, 2020, from https://pdfs.semanticscholar.org/2fd7/3be7cd5100d380eab02675bdbd10037a7d79.pdf. DOI: 10.2174/1567205012666141218124744.

[41] Vanes, L. D., White, T. P., Wigton, R. L., Joyce, D., Collier, T., & Shergill, S. S. Reduced susceptibility to the sound-induced flash fusion illusion in schizophrenia. *Psychiatry Research*, 2016, 245: 58–65. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.psychres.2016.08.016. DOI: 10.1016/j.psychres.2016.08.016.

[42] Haß, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., & Dillo, W., ···Szycik, G. Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cognitive Neuropsychiatry*, 2017, 22(2): 145–158. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1080/13546805.2017.1287693. DOI: 10.1080/13546805.2017.1287693.

[43] Gieseler, A., Tahden, M. A. S., Thiel, C. M., & Colonius, H. Does hearing aid use affect audiovisual integration in mild hearing impairment? *Experimental Brain Research*, 2018, 236: 1161–1179. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1007/s00221-018-5206-6. DOI: 10.1007/s00221-018-5206-6.

[44] DeLoss, D. J., Pierce, Russell, S., Andersen, & George, J. Multisensory integration, aging, and the sound-induced flash illusion. *Psychology Aging*, 2013, 28(3): 802–812. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1037/a0033289. DOI: 10.1037/a0033289.

[45] Mcgovern, D. P., Eugenie, R., John, S., Martin, M. G. T., & Newell, F. N. The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration. *Frontiers in Aging Neuroscience*, 2014, 6: 250. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3389/fnagi.2014.00250. DOI: 10.3389/fnagi.2014.00250.

[46] Chan, J. S., Wibral, M., Wollstadt, P., Stawowsky, C., Brandl, M., Helbling, S., . . . Kaiser, J. Predictive coding over the lifespan: Increased reliance on perceptual priors in older adults —a magnetoencephalography and dynamic causal modelling study. *BioRxiv*, 2017,178095. Retrieved on May 22, 2020, from https://www.biorxiv.org/content/biorxiv/early/2017/08/18/178095.full.pdf. DOI: 10.1101/178095.

[47] Chan, J. S., Connolly, S. K., & Setti, A. The number of stimulus-onset

asynchronies affects the perception of the sound-induced flash illusion in young and older adults. *Multisensory Research*, 2018, 31(3-4): 175-190. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1163/22134808-00002605. DOI: 10.1163/22134808-00002605.

[48] Hernández, B., Setti, A., Kenny, R. A., & Newell, F. N. Individual differences in ageing, cognitive status, and sex on susceptibility to the sound-induced flash illusion: A large-scale study. *Psychology and Aging*, 2019, 34(7). Retrieved on May 22, 2020, from https://doi.org/10.1037/pag0000396. DOI: 10.1037/pag0000396.

[49] O' Brien, J., Ottoboni, G., Tessari, A., & Setti, A. One bout of open skill exercise improves cross-modal perception and immediate memory in healthy older adults who habitually exercise. *PLoS ONE*, 2017, 12(6): e0178739. Retrieved on May 22, 2020, from https://sci-hub.im/10.1371/journal.pone.0178739. DOI: 10.1371/journal.pone.0178739.

[50] Merriman, N., Whyatt, C., Setti, A., Craig, C., & Newell, F. Successful balance training is associated with improved multisensory function in fall-prone older adults. *Computers in Human Behavior*, 2015, 45: 192-203. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.chb.2014.12.017. DOI: 10.1016/j.chb.2014.12.017.

[51] Cecere, R., Rees, G., & Romei, V. Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology*, 2015, 25(2): 231-235. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.cub.2014.11.034. DOI: 10.1016/j.cub.2014.11.034.

[52] Keil, J. & Senkowski, D. Individual alpha frequency relates to the sound-induced flash illusion. *Multisensory Research*, 2017, 30(6): 565-578. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1163/22134808-00002572. DOI:10.1163/22134808-00002572.

[53] Sun, Liu, Li, Clara, Wang, & Zhang. Effects of repetition suppression on sound induced flash illusion with aging. *Frontiers in Psychology*. 2020, 11: 216. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3389/fpsyg.2020.00216. DOI: 10.3389/fpsyg.2020.00216.

[54] Mahoney, J. R., Li, P. C. C., Ohpark, M., Verghese, J., & Holtzer, R. Multisensory integration across the senses in young and old adults. *Brain Research*, 2011, 1426(17): 43-53. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.brainres.2011.09.017. DOI: 10.1016/j.brainres.2011.09.017.

[55] Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. Enhanced multisensory integration in older adults. *Neurobiology of Aging*, 2006, 27(8): 1155-1163. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.neurobiolaging.2005.05.024. DOI: 10.1016/j.neurobiolaging.2005.05.024.

[56] De Haas, B., Kanai, R., Jalkanen, L., & Rees, G. Grey matter volume in early human visual cortex predicts proneness to the sound-induced flash illusion. *Proceedings of the Royal Society B: Biological Sciences*, 2012, 279(1749): 4955–4961. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1098/rspb.2012.2132. DOI: 10.1098/rspb.2012.2132.

[57] Balz, J., Roa Romero, Y., Keil, J., Krebber, M., Niedeggen, M., Gallinat, J. & Senkowski, D. Beta/Gamma oscillations and event-related potentials indicate aberrant multisensory processing in schizophrenia. *Frontiers in Psychology*, 2016a, 7: 1896. Retrieved on May 22, 2020, from https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01896/full. DOI: 10.3389/fpsyg.2016.01896.

[58] Balz, J., Keil, J., Romero, Y. R., Mekle, R., Schubert, F., Aydin, S., ···Senkowski, D. GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. *NeuroImage*, 2016b, 125: 724–730. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.cogbrainres.2004.06.004. DOI: 10.1016/j.neuroimage.2015.10.087.

[59] Shams, L., Kamitani, Y., Thompson, S, & Shimojo, S. Sound alters visual evoked potentials in humans. *Neuroreport*, 2001, 12(17): 3849–3852. Retrieved on May 22, 2020, from https://sci-hub.im/10.1097/00001756-200112040-00049. DOI: 10.1097/00001756-200112040-00049.

[60] Mishra, J., Martinez, A., & Hillyard, S. A. Cortical processes underlying sound-induced flash fusion. *Brain Research*, 2008, 1242(4): 102–115. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.brainres.2008.05.023. DOI: 10.1016/j.brainres.2008.05.023.

[61] Kaiser, M., Senkowski, D., Busch, N. A., Balz, J., & Keil, J. Single trial prestimulus oscillations predict perception of the sound-induced flash illusion. *Scientific Reports*, 2019, 9(1): 5983. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1038/s41598-019-42380-x. DOI: 10.1038/s41598-019-42380-x.

[62] Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 2006, 31(3): 1247–1256. Retrieved on May 22, 2020, from https://sci-hub.im/10.1016/j.neuroimage.2006.01.016. DOI: 10.1016/j.neuroimage.2006.01.016.

[63] Watkins, S., Shams, L., Josephs, O., & Rees, G. Activity in human V1 follows multisensory perception. *NeuroImage*, 2007, 37(2): 572–578. Retrieved on May 22, 2020, from, https://sci-hub.im/https://doi.org/10.1016/j.neuroimage.2007.05.027. DOI: 10.1016/j.neuroimage.2007.05.027.

[64] Jiang, Y. & Han, S. H. Perceptual gain and perceptual loss: Distinct neural mechanisms of audiovisual interactions. *International Journal of Magnetic*

Resonance Imaging, 2007, 1: 1749-8023. Retrieved on May 22, 2020, from http://111.205.231.33/static/main/uploads/faculty/han_shi_hui/Jiang%20and%20Han%20jMRI%20reprint%

[65] Bolognini, N., Rossetti, A., Casati, C., Mancini, F., & Vallar, G. Neuromodulation of multisensory perception: A tDCS study of the sound-induced flash illusion. *Neuropsychologia*, 2011, 49(2): 231-237. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.neuropsychologia.2010.11.015. DOI: 10.1016/j.neuropsychologia.2010.11.015.

[66] Maccora, S., Giglia, G., Bolognini, N., Cosentino, G., Gangitano, M., Salemi, G. & Brighina, F. Cathodal occipital tDCS is unable to modulate the sound induced flash illusion in migraine. *Frontiers in Human Neuroscience*, 2019, 13: 247. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3389/fnhum.2019.00247. DOI: 10.3389/fnhum.2019.00247.

[67] Kamke, M. R., Vieth, H. E., Cottrell, D., & Mattingley, J. B. Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage*, 2012, 62(3): 1334-1341. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.neuroimage.2012.05.063. DOI: 10.1016/j.neuroimage.2012.05.063.

[68] Keil, J., Müller, N., Hartmann, T., & Weisz, N. Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cerebral Cortex*, 2013, 24(5): 1278-1288. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1093/cercor/bhs409. DOI: 10.1093/cercor/bhs409.

[69] Tang, X., Wu, J., & Shen, Y. The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, 2016, 61: 208-224. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.neubiorev.2015.11.002. DOI: 10.1016/j.neubiorev.2015.11.002.

[70] Zhou, W., Jiang, Y., He, S., & Chen, D. Olfaction modulates visual perception in binocular rivalry. *Current Biology*, 2010, 20(15): 1356-1358. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.cub.2010.05.059. DOI: 10.1016/j.cub.2010.05.059.

[71] Quak, M., London, R. E. & Talsma, D. A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, 2015, 9: 197. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3389/fnhum.2015.00197. DOI: 10.3389/fnhum.2015.00197.

[72] Barutchu A, Sahu A, Humphreys G W, et al. Multisensory processing in event-based prospective memory. *Acta Psychologica*, 2019, 192: 23-30. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.actpsy.2018.10.015. DOI: 10.1016/j.actpsy.2018.10.015.

[73] Brunetti, R., Indraccolo, A., Mastroberardino, S., Spence, C., & Santangelo, V. The impact of cross-modal correspondences on working

memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, 2017, 43(4): 819-831. Retrieved on May 22, 2020, from https://ora.ox.ac.uk/objects/uuid:dd7d7a78-016a-4cc9-aeeb-f070e31a3401/download_file?file_format=pdf&safe_filename=The%2BImpact%2Bof%2BCrossmodal%2BCo DOI: 10.1037/xhp0000348.

[74] Xie, Y., Xu, Y., Bian, C., & Li, M. Semantic congruent audiovisual integration during the encoding stage of working memory: An ERP and sLORETA study. *Scientific Reports*, 2017, 7(1): 5112. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1038/s41598-017-05471-1. DOI:10.1038/s41598-017-05471-1.

[75] Marques, L. M., Lapenta, O. M., Costa, T. L., & Boggio, P. S. Multisensory integration processes underlying speech perception as revealed by the McGurk illusion. *Language, Cognition and Neuroscience*, 2016, 31(9): 1115-1129. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1080/23273798.2016.1190023. DOI: 10.1080/23273798.2016.1190023.

[76] Samuel, A. G. Speech perception. *Annual Review of Psychology*, 2011, 62(1): 49-72. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1146/annurev.psych.55.090902.142028. DOI: 10.1146/annurev.psych.55.090902.142028.

[77] Olasagasti, I., Bouton, S., & Giraud, A. L. Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex*, 2015, 68: 61-75. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.cortex.2015.04.008. DOI: 10.1016/j.cortex.2015.04.008.

[78] Magnotti, J. F., & Beauchamp, M. S. The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review*, 2015, 22(3): 701-709. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.3758/s13423-014-0722-2. DOI: 10.3758/s13423-014-0722-2.

[79] 罗霄骁, 康冠兰, 周晓林. McGurk 效应的影响因素与神经基础. 心理科学进展, 2018, 26(11):1935-1951. (LUO Xiaoxiao, KANG Guanlan, ZHOU Xiaolin. The influential factors and neural mechanisms of McGurk effect. *Advances in Psychological Science*, 2018, 26(11): 1935-1951. Retrieved on May 22, 2020, from http://journal.psych.ac.cn/xlkxjz/CN/Y2018/V26/I11/1935#1. DOI: 10.3724/SP.J.1042.2018.01935.)

[80] Cuppini, Magosso, Bolognini , Vallar , & Ursino. A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage*, 2014, 92: 248-266. Retrieved on May 22, 2020, from https://sci-hub.im/https://doi.org/10.1016/j.neuroimage.2014.02.001. DOI: 10.1016/j.neuroimage.2014.02.001.

Author Contributions:

WANG Aijun: Conceptualization, Writing; HUANG Jie: Literature review, Reference management; LU Feifei: Literature review; HE Jiaying: Literature review; TANG Xiaoyu: Framework development; ZHANG Ming: Conceptualization, Final manuscript revision.

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

Source: ChinaXiv —Machine translation. Verify with original.