Auditory stimuli extend the temporal window of visual integration by modulating alpha-band oscillations

Mengting Xu[a,b†], Biao Han[a,b†], Qi Chen [a,b*], Lu Shen[a,b*]

[a] Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China

[b] School of Psychology, South China Normal University, Guangzhou 510631, China

***Corresponding Authors:**

Lu Shen, PhD **Email:** lu.shen2013@gmail.com

Qi Chen, PhD **Email:** qi.chen27@gmail.com

Author Contributions:

[†]M.X. and B.H. contributed equally to this work.

M.X., B.H., Q.C., and L.S. designed research; M.X., B.H., and L.S. performed research; M.X., B.H., and L.S. contributed new reagents/analytic tools; M.X., B.H., and L.S. analyzed data; and M.X., B.H., Q.C., and L.S. wrote the paper.

Competing Interest Statement: No competing interest.

Keywords: Multisensory perception, fusion illusion, alpha oscillations

Abstract

In the multisensory environment, the interactions between inputs from different sensory modalities are not fully understood. Here, we conducted an electroencephalography (EEG) experiment to investigate how auditory stimuli shape the temporal window of visual integration in human subjects. Participants were presented with two consecutive visual flashes, either accompanied by an auditory beep or without, and were asked to report their perception of one or two flashes. Behaviorally, we found that the introduction of auditory input induced a longer temporal window for integration. Alpha frequency analysis further revealed that the presence of auditory stimuli led to poststimulus alpha frequency degradation, positively correlating with the prolonged temporal window, supporting the idea that alpha oscillations represent the temporal window of visual integration. Further exploration of prestimulus alpha oscillations revealed that auditory stimuli diminished the predictive role of prestimulus alpha frequency while enhancing the predictive role of prestimulus alpha phase in determining perceptual outcomes. To unveil the underlying mechanism, we developed a computational model based on the phase-resetting hypothesis and the perceptual cycle theory, successfully replicating key behavioral and neural findings. Together, our results suggest that concurrent auditory input extends the temporal window of visual integration by resetting the phase of alpha oscillations in the visual cortex, leading to alpha frequency degradation.

Significance Statement

To reduce the complexity of our world, visual perception operates within temporal windows, where stimuli falling within the same temporal window become integrated, as evidenced by alpha-band oscillations. However, it is unclear how these temporal windows can be changed, especially when dealing with multiple senses simultaneously. This EEG study reveals that introducing auditory input prolongs the temporal window of visual integration by reducing alpha frequency. A computational model proposes auditory input-induced phase resetting as the underlying mechanism, influencing the predictive roles of alpha phase and frequency in perception. These findings offer a novel perspective on multisensory processing, providing new insights into the different roles of alpha frequency and phase in shaping temporal perception.

Introduction

In our daily lives, we are constantly bombarded with information from various senses. Our brains face the challenge of organizing and integrating these signals into coherent perceptual experiences, and this becomes more complex due to the interactions between different sensory modalities (1-3). Numerous studies consistently highlight the noteworthy influence of sound signals on visual perception, particularly in the temporal domain. For example, the introduction of sound has been observed to influence the perceived duration of visual stimuli or rate of visual events (4-7). The temporal relationship between visual and auditory stimuli plays a crucial role, either enhancing or degrading visual temporal resolution (8, 9). These findings intriguingly suggest that auditory stimuli may exert a profound influence on the temporal processing of the visual system. However, the neural mechanisms that underlie this cross-modal influence remain elusive.

Temporal windows, a fundamental attribute of the visual system, refer to time intervals during which discrete stimuli interact to influence perception (10). A simple example is the fusion of two brief stimuli into a unified percept when they occur within a certain time range. The rhythmic patterns of alpha oscillations, serve as a potential mechanism for defining this temporal window (11-13). This notion gains support from research demonstrating that the phase and frequency of alpha oscillations shape the temporal integration process. Alpha oscillation phase influences visual signal detectability at the perceptual threshold (14-16), the perceived timing of stimuli (17, 18), and the occurrence of perceptual illusions(19-21). Moreover, emerging evidence highlights a correlation between the frequency of alpha oscillations and temporal integration (3, 12, 13, 22, 23). Notably, this frequency not only correlates with but also causally influences observers' tendency to integrate signals within(13, 22, 23) and across different sensory modalities(3, 12, 24). However, it is important to note that some studies have failed to find a significant correlation between alpha oscillations and temporal parsing of visual signals (25-27). A recent study provide evidence that pre-stimulus alpha frequency does not substantially influence the temporal segmentation of visual signals in visual or audiovisual perception (27). This highlights the need for further research to confirm the relationship between alpha oscillations and the temporal window within and across the senses.

In the context of audiovisual processing, the presentation of two visual flashes with an auditory stimulus leads to an increased likelihood of perceiving them as a single flash, inducing fusion illusions (28). This observation suggests that cross-modal stimulation has the potential to impact the temporal window of visual processing. Neural studies have revealed significant activation changes in the primary visual cortex (V1) when visual stimuli are presented concurrently with incongruent auditory stimuli, which are closely linked to reports of visual illusions(2). Furthermore, research on animals and humans has demonstrated that auditory stimuli can modulate early visual

cortex processing through phase-reset mechanisms(29-32), particularly in the alpha frequency range. Building upon these findings, one potential hypothesis is that cross-modal sound stimuli may influence the temporal window of visual integration by exerting an influence on alpha oscillations.

To test this hypothesis, we conducted an experiment where observers performed a visual flash discrimination task while their EEG data was recorded. We analyzed the frequency and phase of alpha oscillations for two experimental conditions: one with only two visual flashes (F2) and the other with two visual flashes accompanied by an auditory beep (F2B1) (Fig. 1*A* and *B*). Upon comparing different conditions, our investigation revealed that the introduction of auditory stimuli effectively prolonged the temporal window of integration by reducing the instantaneous frequency of alpha oscillations. Notably, in the F2 condition, the instantaneous frequency of alpha oscillations plays a predictive role in perceptual outcomes, while in the F2B1 condition, this predictive role is absent. Conversely, the phase of alpha oscillations exhibits a predictive role in perceptual outcomes in the F2B1 condition, but not in the F2 condition. To elucidate the underlying mechanisms, we introduced a sound-induced phase-resetting model. Our simulation results indicate that the introduction of auditory input resets the phase of alpha oscillations, subsequently extending the temporal window through a reduction in alpha frequency. This modification significantly influences the predictive roles of both phase and frequency.

---------- insert Fig 1 about here ----------

Results

Prior to the main EEG experiment, participants underwent a psychophysical pretest to determine the fusion threshold for each condition. Psychometric curves were fitted to the proportion of perceiving two flash percepts at each of the eight ISIs for the F2 and F2B1 conditions, respectively (Fig. 1*C*). The fusion threshold, representing the ISI at which equal proportions of 1-flash and 2-flash trials were reported, was derived for each condition and participant (*SI Appendix*, Fig. S1). This threshold provides an estimate of the temporal window for integration within the visual modality (F2) and across the audiovisual modality (F2B1). A two-tailed paired t-test revealed that the fusion threshold was significantly shorter ($t_{(33)} = 5.23$, $p = 9.00\times10^{-6}$; Fig. 1*D*) in the F2 (52.68 ± 12.25ms, mean ± SD) than in the F2B1 condition (65.18 ± 13.17ms, mean ± SD), indicating a significant shift in the fusion threshold between the two conditions.

In the main EEG experiment, each condition comprised three types of trials: the explicit 1-flash trials with a short ISI (30ms), the explicit 2-flash trials with a long ISI (100ms), and the bistable trials

with the fusion threshold ISI obtained from the psychophysical procedure. Participants exhibited high accuracy rates in the explicit trials for both the F2 and F2B1 conditions (F2 with short ISI: 89.16 ± 11.81%; F2 with long ISI: 95.96 ± 4.31%; F2B1 with short ISI: 95.48 ± 5.86%; F2B1 with long ISI: 89.52 ± 14.91%, mean ± SD), indicating their ability to accurately discriminate between the explicit percepts in both conditions. Regarding the bistable trials with the threshold ISI, the proportions of 1-flash and 2-flash percepts were comparable in both the F2 ($t_{(33)} = 1.85$, $p = 0.07$, two-tailed) and F2B1 conditions ($t_{(33)} = 1.92$, $p = 0.06$, two-tailed; Fig. 1*E*), suggesting that participants experienced a similar level of perceptual ambiguity in the two conditions.

Auditory input extended the temporal window for visual integration by decreasing occipital alpha frequency. The EEG data showed a clear peak in the alpha-band power and a posterior scalp distribution of alpha power during the peri-stimulus period (-600 to 600 ms relative to the first stimulus onset) for all the participants in both F2 and F2B1 conditions (Fig. 2*A*). These results guided further analysis by confirming the frequency of interest to 8-13hz and the region of interest to the posterior electrodes (Oz, O1, O2, POz, PO1, PO2, PO3, PO4, PO5, PO6).

Building upon our behavioral observations, we noted a significant extension in the temporal window for visual integration when auditory input was introduced. If we assume that alpha frequency serves as an indicator of the temporal window within the visual cortex, it is plausible to hypothesize that alpha frequency in the F2B1 condition might decrease compared to the F2 condition. To test this hypothesis, we performed a direct comparison of poststimulus instantaneous alpha frequency (IAF) between the F2 and F2B1 conditions by collapsing trials across 1-flash and 2-flash precepts while ensuring an equal number of trials in each condition (See Materials and Methods). The instantaneous frequency was calculated as the temporal derivative of the instantaneous phase, which extracted the alpha-band phase angle time series over the occipital electrodes via the Hilbert transform. Consistent with our hypothesis, in the bistable trials, we found that the frequency of F2B1 was significantly lower than that of F2 after the presentation of auditory stimulus (0-280ms relative to the first stimulus onset, $p = 1.00\times10^{-3}$, cluster-based correction, two-tailed; Fig. 2*B*).

Subsequently, we conducted a between-subject correlation analysis between the poststimulus IAF differences in the F2B1 and F2 conditions and the fusion threshold ISI differences between these conditions, as measured through psychophysics. Interestingly, we observed a significant correlation between the decrease in alpha frequency and the increase in fusion threshold in the F2B1 condition relative to the F2 condition (Spearman correlation: $r = 0.45$, $p = 7.55 \times 10^{-3}$). Specifically, the greater the reduction in alpha frequency induced by the auditory stimulus, the longer the fusion threshold in the F2B1 condition compared to the F2 condition (Fig. 2*C*). These

findings suggest that the concurrent presentation of auditory input may lead to a decrease in alpha frequency, thereby extending the temporal window of visual integration.

Given the difference in fusion thresholds between the F2 and F2B1 conditions, which could potentially contribute to the observed poststimulus differences between the two conditions, we conducted an additional analysis to address this concern. Specifically, we examined the IAF difference between the two conditions in trials with explicit short ISI and long ISI, respectively. In the explicit trials, ISIs were consistent across participants and conditions, with the only distinguishing factor between F2B1 and F2 being the presence of auditory input in F2B1. Notably, our results revealed that the alpha frequency was consistently lower in F2B1 compared to F2 for both explicit conditions (explicit short ISI: 0-180ms relative to the first stimulus onset, p = 0.02, cluster-based correction; explicit long ISI: 0-370ms relative to the first stimulus onset, $p = 1.00\times10^{-3}$, cluster-based correction; Fig. 2*D* and *E*). This control analysis robustly confirmed that the decrease in alpha frequency is indeed attributable to the auditory input, regardless of the variations in stimulus ISI.

---------- insert Fig 2 about here ----------

Auditory input disrupted the predictive role of prestimulus alpha frequency. Extensive research has demonstrated the significant impact of prestimulus alpha frequency on the outcomes of visual integration. In our study, we sought to investigate whether prestimulus alpha frequency exerted a similar influence on the perceptual outcomes in both the F2 and F2B1 conditions. First, we conducted a within-subject comparison of the instantaneous alpha frequency (IAF) variations between different perceptual outcomes of bistable trials in the F2 and F2B1 conditions, respectively. Our results demonstrated that in the F2 condition, the IAF was significantly higher in the 2-flash trials compared to 1-flash trials (Fig. 3*A*), both in the pre-stimulus period (from -330 ms to -30ms relative to the first stimulus onset, p = 0.02, cluster-based correction, two-tailed) and post-stimulus period (from 130ms to 540ms, $p = 4.00\times10^{-3}$, cluster-based correction, two-tailed). These findings are in line with previous research, indicating that slower alpha frequencies are associated with a wider temporal window, thereby increasing the likelihood that two visual stimuli fall within the same window and are integrated as one percept (13). However, in the F2B1 condition, no significant difference in IAF between different perceptual outcomes was observed (Fig. 3*B*).

To further compare the different patterns in perception between the two conditions, we averaged the IAF over the posterior electrodes and the entire time of interest (-600 to 600 ms). These mean

IAFs were then subjected to a 2 (percepts: 1-flash vs. 2-flash) * 2(conditions: F2 and F2B1) repeated-measures ANOVA. The results revealed a significant main effect of percepts, $F_{(1,33)} = 4.29$, $p = 0.046$, indicating that the IAF was significantly higher in the 2-flash percepts compared to the 1-flash percepts. Additionally, there was a marginal significant main effect of conditions, $F_{(1,33)} = 3.40$, $p = 0.07$, suggesting a trend towards different IAFs between the F2 and F2B1 conditions. More importantly, the interaction effect was significant, $F_{(1,33)} = 9.86$, $p = 3.55\times10^{-3}$. Specifically, in the F2 condition, the frequency of 2-flash percepts was significantly higher than 1-flash percepts ($t_{(33)} = 5.23$, $p = 3.87\times10^{-3}$), while in the F2B1 condition, no significant difference between the two percepts was observed ($t_{(33)} = -0.49$, $p = 0.63$; Fig. 3*C*). These results indicate statistically that the influence of IAF on perceptual outcomes differs between the two conditions.

---------- insert Fig 3 about here ----------

Moreover, to address any potential bias in the calculation of instantaneous alpha frequency (IAF) resulting from the 1/f slope effect, we applied a demodulation process with attenuated 1/f characteristics on the time-domain signal (33) and repeated the IAF analysis. The observed pattern of results remained consistent (*SI Appendix*, Fig. S3). In addition, to rule out the possibility that the instantaneous alpha frequency modulation is attributed to the oscillatory power, we performed a statistical test on the instantaneous alpha power within the time and channels of interest between the two percepts for F2 and F2B1 conditions, respectively. No significant differences between the two perceptual outcomes were found in either the F2 or F2B1 conditions (*SI Appendix*, Fig. S3).

Auditory input enhanced the predictive role of prestimulus alpha phase. The spontaneous alpha-band phase has previously been shown to mirror the cyclic alterations in cortical excitability, thereby supporting the concept of perceptual cycles. In light of this, we extended our investigation to explore the influence of prestimulus alpha phase on perceptual outcomes by computing the phase opposition sum (POS), a measure that quantifies the difference in phase distributions between two sets of trials (34, 35). If the phase of spontaneous alpha oscillations prior to the first flash is predictive of the subsequent perceptual outcome, we should observe a strong phase clustering around a certain phase angle for 1-flash trials and distinct phase clustering around another phase angle for 2-flash trials, resulting in a significant value of POS.

In the F2B1 condition, we indeed observed a significant phase opposition between 1-flash and 2-flash trials, spanning from -460 to 0ms relative to the first flash onset, within the frequency range of 8 to 13Hz ($p = 3.03\times10^{-3}$; Fig. 4*A*, cluster-based correction). However, no significant phase effect

manifested in the F2 condition. For demonstration purpose, to further visualize the mean phase angle differences across participants, phases at the time-frequency point with maximal POS effect (at 10.28Hz; -300ms) at electrode Oz were pooled into 16 bins. In line with the POS analysis, the directions of the preferred phase for 1-flash and 2-flash percepts exhibited significant differences across participants in the F2B1 condition ($F_{(1,66)} = 15.31$, $p = 2.18\times10^{-4}$, Fig. 4*B*, upper panel). Conversely, no significant phase difference between the two percepts was detected in the F2 condition ($F_{(1,66)} = 3.30$, $p = 0.07$, Fig. 4*B*, lower panel).

Furthermore, we evaluated whether the observed difference in prestimulus POS patterns between F2B1 and F2 was statistically significant using a permutation procedure. Specifically, we computed the mean POS difference between F2B1 and F2 in the frequency range of 8 to 13hz and time range from -600ms to 0 relative to the first flash onset. Subsequently, we compared this mean POS difference value against a distribution of POS difference values expected by chance (see Materials and Methods; Fig. 4*C*). This analysis confirmed a statistically significant distinction in POS patterns between F2B1 and F2 conditions, $p = 4.60\times10^{-3}$, indicating a prevailing phase modulation effect on perceptual outcomes in the F2B1 condition.

The relationship between the perceptual performance and prestimulus phase was further assessed by categorizing single trials according to the phase at the optimal time-frequency point (-300ms, 10.28Hz, Oz) and calculating the proportion of 2-flash percepts in each of 9 phase bins (Fig. 4*D*). For each participant, the phase angles were shifted so that the phase corresponding to the highest proportion of 2-flash percepts was aligned to a phase angle of zero. Consequently, due to this artificial alignment, the 2-flash proportion is necessarily maximal at a phase angle of zero; therefore, the zero-phase bin was discarded from further analyses. Most importantly, if the prestimulus phase indeed modulated the perceptual outcomes, we expected that the proportions of 2-flash percepts would peak at the optimal phase angle, gradually decreasing as the phases deviate from this optimal angle. Therefore, we tested for the quadratic main effect for phase bins (8 levels) and their interaction with different conditions (F2B1 vs. F2). There was no significant quadratic effect of phase bins ($F_{(1,33)} = 2.47$, $p = 0.13$). Consistent with the observation that phases exerted a more pronounced influence on behavior when auditory input was introduced, a significant quadratic interaction emerged between conditions and phase bins ($F_{(1,33)} = 6.45$, $p = 0.02$). Post hoc tests further revealed a significant phase modulation effect only for F2B1 ($F_{(1,33)} = 7.02$, $p = 0.01$), but not for F2 condition ($F < 1$).

---------- insert Fig 4 about here ----------

Auditory input-induced phase resetting model explains the decrease in alpha frequency and the altered predictive roles of alpha frequency and phase. What could be underlying cause of the decrease in poststimulus alpha frequency observed in F2B1 condition? Could this cause explain the divergent effects of the altered predictive roles of alpha frequency and phase? Previous research has shown that transient auditory stimuli can reset the phase of ongoing oscillations in the visual cortex(31, 36). This, in turn, has a notable impact on multisensory integration processes (32, 37). Importantly, this phase resetting has the potential to disrupt the timing and synchronization of periodic oscillations, which could subsequently lead to a shift in oscillation speed.

To rigorously test this hypothesis, we introduced a phase-resetting model for examining perceptual processing (Fig. 5). Developed based on perceptual cycles theoretical framework (38), this model utilized simulations to replicate our key observations. In the context of perceptual cycle theory, both frequency and phase play fundamental roles in shaping perceptual outcomes. Frequency governs the temporal resolution of the visual system—higher frequencies favor the perception of two distinct flashes, while lower frequencies tend to amalgamate them into a single event. The oscillatory phase represents brain states fluctuating between high and low cortical excitability (31, 39), thereby influencing perceptual processing intensity. Our model differentiates percepts by assessing whether two flash stimuli fall within a single cycle. Good phases near the peak enhance processing efficiency, resulting in clearer perceptions of 1-flash and 2-flash percepts. Conversely, bad phases near the trough lead to less efficient processing and increased perceptual ambiguity. Fig. 5 visually illustrates our model. When the second flash aligns with a good phase, efficient processing occurs, leading to a clear perception of 1-flash or 2-flash based on whether they fall within the same or different cycles. In contrast, when the second flash aligns with a bad phase, less efficient processing increases perceptual ambiguity, making both interpretations plausible. For the F2B1 condition, the introduction of auditory input, known to induce delayed responses in the visual cortex relative to visual stimuli (40, 41), induces phase resetting of the ongoing oscillations. The specific parameters governing phase resetting are derived from the model proposed by Canavier in 2015 (42), as detailed in the Methods section.

---------- insert Fig 5 about here ----------

Building upon the established model, we simulated 1000 trials under two conditions (F2 vs. F2B1). Initial analysis of the model-generated behavioral data yielded results consistent with real participants, notably indicating a prolonged integration window under the F2B1 condition compared to F2 (Fig. 6*A*). Subsequent analysis of alpha oscillations in the simulated data, using the same methodology as for real data, consistently unveiled a lower alpha frequency under F2B1 compared to F2 (Fig. 6*B*). For the alpha frequency effects in perceptual outcomes, mirroring our real data, in our simulations, we averaged the IAF over the entire time of interest (artificially defined from -600 to 600 ms). The simulation results highlighted a more pronounced modulation of alpha frequency in perceptual outcomes in the F2 compared to the F2B1 condition (Fig. 6*C*).

Furthermore, in line with phase results in our real data, simulated data demonstrated more pronounced phase effects in the F2B1 condition compared to the F2 condition at the onset of flash 1 (Fig. 6*D*). Since the phase at which the second flash occurs determines perceptual clarity between the two percepts in our model, we further presented the phase distribution of the second flash for the F2 and F2B1 conditions, respectively (Fig. 6*E*). Notably, auditory input-induced phase resetting made the second flash more likely to fall into the range of a favorable phase (Fig. 6*E*), leading to higher subjective sensitivity (Fig. 6*F*), compared to the F2 condition.

---------- insert Fig 6 about here ----------

Our own findings lend credence to this hypothesis. We observed a higher alpha inter-trial coherence (ITC) in the F2B1 condition compared to the F2 condition (*SI Appendix*, Fig. S4*A* and *B*). Because an increase in ITC following a stimulus may reflect a stimulus-evoked response, we also computed changes in alpha power after stimulus onset. A stimulus-evoked response, characterized by an additional signal superimposed on ongoing oscillations, would manifest as both an increase in ITC and power following the stimulus, whereas a pure phase-resetting of ongoing oscillations will manifest as an increase in ITC with no accompanying power change or a power decrease (43, 44). Our findings showed no significant difference in alpha power between the two conditions (*SI Appendix*, Fig. S4*C*). Furthermore, we observed no substantial correlation between the magnitude of alpha power and alpha ITC increase after stimulus onset (*SI Appendix*, Fig. S4*D*). Therefore, although we cannot rule out a stimulus-evoked explanation for alpha ITC, we believe the increase in alpha ITC at least partially reflects phase-resetting of spontaneous oscillations in occipital regions.

Discussion

In the present study, we delved into the intricate interplay between auditory stimuli and visual processing, specifically examining the impact on the temporal window of integration. Behaviorally, we uncovered a prolonged fusion threshold in the presence of auditory stimuli, indicative of an extended temporal window for visual perception induced by cross-modal interactions (Fig. 1*C* and *D*). The analysis of alpha oscillations provided valuable insights, demonstrating that the introduction of auditory stimuli resulted in poststimulus alpha frequency degradation, which correlated with the extended window, affirming the role of alpha oscillations in shaping the temporal dynamics of visual processing (Fig. 2). Notably, our findings revealed distinct contributions from prestimulus alpha frequency and phase in predicting perceptual outcomes in unimodal and cross-modal conditions (Fig. 3 and 4). To elucidate the underlying mechanisms, we employed a computational model based on the phase resetting hypothesis and perceptual cycle theory, successfully replicating key behavioral and neural findings (Fig. 5 and 6). The convergence of behavioral, neurophysiological, and computational evidence underscores the intricate interplay between auditory and visual modalities in shaping the temporal dynamics of perception.

Our findings reveal that concurrent auditory input significantly extends the temporal window of visual integration compared to the no-sound condition (Fig. 1*C*). This is in line with consistent findings from previous studies demonstrating the impact of sound signals on the perceived duration or rate of visual stimuli (4-7). Intriguingly, this prolonged temporal window is positively correlated with the decrease in poststimulus alpha frequency induced by auditory stimuli. These findings contribute to the enduring hypothesis associating the human alpha rhythm with temporal windows, positing that higher frequencies correspond to shorter windows and heightened temporal resolution (3, 12, 13, 24). A lengthened integration window, as indicated by a longer alpha cycle, implies that two successive visual flashes are more prone to falling within a single cycle. Consequently, the visual system requires an extended duration to effectively differentiate between the two flashes. This established connection between altered alpha frequency and the resultant temporal window expansion provides a nuanced understanding of how auditory stimulation influences the temporal dynamics of visual perception. Essentially, our findings suggest that auditory stimuli can influence the temporal resolution of visual processing, potentially providing a longer temporal window for the integration of cross-modal information (45).

Building upon these findings, a critical inquiry arises: how does auditory stimulation contribute to a decrease in alpha frequency? Previous fMRI studies on multisensory perception have shown that auditory stimuli could modulate activity in primary visual cortex (2, 46). Neurophysiological evidence further emphasizes that transient auditory stimuli can reset the phase of the ongoing oscillations, particularly in alpha-band, within the visual cortex (32, 36, 47). These insights suggest that cross-

modal auditory stimuli might exert influence on alpha oscillations by resetting the phase in the visual cortex. Our empirical evidence, as indicated by the increased phase coherence in the F2B1 condition according to the ITC analysis, lends support to this hypothesis (Fig. S4*A* and *B*). To unravel the underlying mechanisms, we developed a computational model that integrates the perceptual cycle theory and auditory-induced phase resetting. As expected, the introduction of auditory-induced phase resetting, causing a delay in the originally achieved phase, led to a prolonged temporal integration window and a decrease in post-stimulus alpha frequency. This implies that cross-modal auditory stimuli may disrupt the original trajectory of periodic signals, impacting the integration of incoming inputs in the visual cortex (47-49). Together, our combined empirical and computational evidence provides a comprehensive understanding of how auditory stimuli influence the temporal dynamics of visual perception by altering alpha frequency and temporal integration windows in the visual cortex.

Moreover, through a comparative analysis of prestimulus alpha frequency between two perceptual outcomes, in line with established studies (13, 22, 23, 50), our findings underscore the predictive nature of alpha frequency in the F2 condition. Specifically, higher alpha frequency results in a shorter integration window and more frequent sampling of visual information, leading to accurate perception of stimuli with close temporal proximity. In contrast, a lower frequency results in a longer integration window and slower sampling, leading to fusion illusions (Fig. 3*A*). However, in the F2B1 condition, we did not observe a similar frequency pattern (Fig. 3*B*), suggesting that the introduction of auditory stimulus disrupted the predictive role of alpha frequency in visual integration. This aligns with the recent null effect reported by Buergers and Noppeney (2022) (27). One plausible explanation lies in the auditory stimulus inducing phase resetting, realigning the phase towards a specific direction and resulting in an average decrease of ~0.1Hz in post-stimulus alpha frequency (Fig. 2*B*). This subtle change, although seemingly small (~0.1Hz), significantly impacts the ability to discriminate between perceptual outcomes in the F2 condition (Fig. 3*B*). Even with a slightly faster prestimulus alpha frequency that would typically predict more 2-flash percepts, the auditory-induced decrease in frequency makes distinguishing between 1-flash and 2-flash more challenging. Consequently, the predictive effect of prestimulus alpha frequency diminishes in the presence of auditory stimuli. Our proposed phase resetting model further supports these observations, reproducing distinct frequency patterns between unimodal and cross-modal processing (Fig. 6*C*).

It has been demonstrated that phase is able to transport more units of information per time than power changes(51) or spike counts(52), suggesting its efficacy in encoding rapidly changing stimuli within a brief temporal window (53). Due to potential temporal imprecision in oscillatory analyses, where the stimulus-evoked neural response dominates phase estimates for tens to hundreds of milliseconds during the trial period (34, 54), previous research has employed prestimulus phase as

an uncontaminated proxy for post-stimulus oscillations, revealing its connection to perception. While some studies consistently support the predictive role of prestimulus phase (14, 16, 18, 55), others fail to observe such effects(25, 56-59). Our study further contributes to this ongoing discussion by revealing distinct patterns of prestimulus alpha phase influencing two perceptual outcomes for different conditions. Specifically, in the cross-modal F2B1 condition, the pre-stimulus alpha rhythm exhibited a significant phase opposition effect, whereas this modulation effect was much weaker in the unimodal F2 condition (Fig. 4). This suggests a plausible modification in poststimulus processing dynamics related to phase, thereby amplifying the brain's sensitivity to discriminate between 1-flash and 2-flash percepts. Additionally, our simulations directly explore post-stimulus phase and suggest that the observed variability can be attributed to the cross-modal phase reset mechanism. In the cross-modal F2B1 condition, auditory stimuli induce phase resetting, increasing the likelihood of flash 2 falling in a favorable phase compared to the F2 condition (Fig. 6*E*). This, in turn, leads to better processing and, consequently, more reliable percepts of 1-flash and 2-flash. Consequently, the predictive influence of the pre-stimulus phase becomes more pronounced in the F2B1 condition.

The observed phase and frequency effects appear to be unrelated to fluctuations in oscillatory amplitude, as our analyses have revealed. This indicates that, at least in the alpha band, the phase and frequency of ongoing neural oscillations may exert an influence on stimulus processing that is distinct from amplitude variations. This finding aligns with prior evidence (18, 22). Moreover, the discernible distinctions in patterns between the phase and frequency of alpha oscillations in cross-modal integration imply that they independently contribute, to some degree, to the modulation of temporal integration processes.

While there is substantial evidence supporting the role of alpha oscillation frequency in characterizing the temporal window for integration (13, 22, 23, 50), the role of phase in the integration process remains unclear. Most studies on the influence of oscillatory phase on perception have primarily focused on analyzing absolute phase angles within an oscillatory cycle at specific moments, which are either favorable or unfavorable for subsequent perception(14, 16, 60, 61). This framework introduces a potential scenario where one stimulus might be presented at a favorable phase, and the other stimulus might be presented at an unfavorable phase, leading to the erroneous perception of only one stimulus, irrespective of whether they fall in the same or different cycles—a contradiction to our behavioral data. Notably, the above-mentioned studies, however, used near-threshold stimuli, raising uncertainties about their applicability to stimuli with clearly suprathreshold intensities, which are presumably perceived independently of specific phase variations (11). In light of this, relying solely on phase quality to elucidate perception becomes constrained.

In contrast, our model takes a more comprehensive approach by integrating the roles of both frequency and phase. Here, frequency dictates the resolution of the visual system, while phase, although not directly determining the perception of stimuli (given that they are all supra-threshold stimuli), influences the extent to which two perceptual states can be distinguished (62). This integrated perspective enriches our understanding of the intricate interplay between frequency and phase in shaping visual perception, offering novel insights into varied research findings. For instance, in a recent study, null results were obtained regarding the influence of alpha frequency on sensitivity in discriminating 1-flash and 2-flash percepts in both unimodal and cross-modal integration (27). However, our model suggests a potential role for alpha phase in perceptual sensitivity.

Please note that our model, based on perceptual cycles and phase reset, offers a speculative hypothesis that necessitates additional empirical evidence for validation. Incorporating studies with variable perceptual clarity or sensitivity would enhance our understanding of the specific role played by phase. Moreover, acknowledging the physiological limitations in our EEG study, we recognize the complexity of phase resetting induced by transient events in the brain. Our model simplifies this intricacy by focusing on sound-induced phase resetting, but further empirical evidence is crucial for validation and refinement. Distinguishing evoked phase locking from pure phase resetting poses challenges, and intracranial recordings with higher resolution could provide a potential solution(63). In our upcoming studies, we plan to utilize intracranial recordings in future studies to address and overcome this challenge, providing further clarity on the role of phase in perceptual processing.

In summary, our study unveils the impact of simultaneous auditory input on the temporal perception of two visual stimuli. The findings indicate that the introduction of auditory stimuli extends the temporal window of integration, with a notable influence on alpha oscillations. We propose that this effect is driven by auditory input-induced phase resetting in visual areas, which extends the duration of the alpha cycle. Consequently, this elongation strengthens the dominant predictive role of the alpha phase while simultaneously reducing the influence of alpha frequency in multisensory processing. In combination with previous studies investigating similar paradigms, our current findings lend further support to the perceptual cycle theory and offer a novel interpretation for the divergent outcomes observed in previous research on frequency and phase.

Materials and Methods

Participants

Thirty-four healthy right-handed participants (23 females; age 20.8 ± 1.9, Mean ± SD) participated in this experiment. The sample size was pre-specified to ensure at least 80% power to detect experimental effect sizes ranging from moderate to large (Cohen's d > 0.5). All participants have normal or corrected-to-normal visual acuity and have no history of neurological or psychiatric disease. Before the experiment, the study was authorized by the Ethics Committee of the School of Psychology, South China Normal University, in line with the Declaration of Helsinki, and each participant provided their informed permission.

Stimuli

The visual stimuli consisted of two identical white circles, each with a radius of 2 degrees, presented against a gray background (RGB:125,125,125). These circles were displayed rapidly one after the other, positioned peripherally below the fixation point at an eccentricity of 5 degrees of visual angle. Each visual stimulus had a duration of 10 milliseconds. The auditory stimulus was a 10-millisecond pure tone with a frequency of 3000 Hz and an approximate sound level of 50 decibels. To provide a spatially congruent experience for the participants, the auditory stimuli were delivered through a pair of loudspeakers situated on either side of the monitor. This arrangement created the perception that the auditory cues originated from the center of the screen, closely aligned with the visual stimuli. Visual stimuli were presented using Presentation software (Neurobehavioral Systems; http://www.neurobs.com) on a 24-inch high-refresh-rate LCD monitor (resolution: 1920 × 1080, refresh rate: 100 Hz, viewing distance 57 cm) and viewed through a chin-and-forehead rest. The monitor was carefully calibrated using an oscilloscope before the experiment. The study consisted of two conditions: the F2 condition, involving the presentation of two visual stimuli, and the F2B1 condition, where the first visual flash was consistently accompanied by an auditory beep.

Psychophysical procedures

To determine each subject's fusion threshold for the bistable condition, each participant completed a psychophysical pretest before the main experiment. Prior to the psychophysical pretest, participants completed a practice block consisting solely of explicit short (30ms) and long trials (100ms) until accuracy reached at least 90%. During the formal psychophysical pretest, the first flash was presented for a duration of 10 ms, after a variable ISI (eight levels: 30, 40, 50, 60, 70, 80, 90, or 100 ms), the second flash was presented for 10 ms as well. Each ISI comprised 35 trials for both F2 and F2B1 conditions. Participants were then asked to perform a two-alternative forced-choice task in which they had to choose between perceiving one flash or two flashes. A logistic

function was used to fit the eight data points (one for each ISI) to a psychometric curve. The fusion threshold ISI, that is, the point at which 1-flash and 2-flash were equally likely to be reported, was calculated by estimating the 50% performance point on the fitted logistic function for each participant. The individual fusion threshold ISI derived from the psychophysical pretest was then utilized as ISI in the bistable trials of the subsequent EEG experiment.

EEG experimental procedures

Participants were instructed to fixate at a central fixation throughout the experiment without moving their eyes. The experimental task was to discriminate one- or two-flash by pressing one of two prespecified buttons on the keyboard using the index and middle fingers of their right hand. The mapping between the two response buttons and the two types of percepts was counterbalanced between participants. Each participant completed 560 experimental trials, including 15% explicit short trials (ISI 30 ms), 15% explicit long trials (ISI 100 ms), and 70% bistable trials (fusion threshold ISI), organized into ten blocks with short breaks in between. Each trial began with a central fixation lasting 500 ms to 1500 ms, followed by a 10 ms presentation of the first visual stimulus, accompanied by an auditory beep or not. After a variable ISI (30 ms, 100 ms, or the participant's individual fusion threshold), the second frame was displayed for 10 ms, and participants were required to respond within 2000 ms.

Recording and preprocessing of EEG data

The electroencephalogram (EEG) was recorded at a sampling rate of 1000 Hz using a Neuronscan SynAmps2 system equipped with 64-channel Quick-Caps and Ag/AgCl electrodes. To capture horizontal and vertical electrooculograms, four extra electrodes were placed around the participants' eyes. The impedances of all electrodes were diligently maintained below 5 kΩ to ensure high signal quality. The signals were online-referenced to an electrode positioned between Cz and CPz. Subsequently, the data underwent offline processing and analysis using FieldTrip (64) (www.fieldtriptoolbox.org) and custom MATLAB scripting. Data were filtered to eliminate interference from 50Hz power lines, down-sampled to 500 Hz, re-referenced to the average reference, epoch from -1300 to 1000 ms. To correct for any baseline drifts, the mean amplitude was subtracted over a time window of -500 to 0 ms relative to the presentation of the first flash. It should be noted that the subtraction of the mean amplitude has no effect on the pre-stimulus alpha oscillations, as it mainly affects only the zero frequency component (35). Before segregating them into different categories, signals containing blinks, muscular movements, and other EEG artifacts were detected and removed through independent component analysis.

Alpha oscillation analysis of EEG data

In our EEG data analysis, our primary focus was on the bistable trials within both the F2 and F2B1 conditions, unless stated otherwise. Fast Fourier Transform (FFT) was applied to generate a power spectrum for all electrodes and participants, ranging from 1 to 30 Hz. In each trial, the power spectrum was computed over the entire period of interest (-600 to 600 ms, relative to the presentation of the first flash). Subsequently, the grand-level power spectrum was obtained by averaging across all trials, electrodes, and participants, thereby identifying the frequency band where the peak frequency, characterized by the highest power, was located. The power topography map was constructed to depict the most prominent frequency band observed in the power spectrum.

Following established methods (35) , we computed instantaneous alpha frequencies (IAF) over time. IAF was determined by selecting occipital electrodes (O1, O2, Oz, PO7, PO5, PO3, POz, PO4, PO6, and PO8) exhibiting maximum alpha power in the posterior region of interest for each participant. Specifically, in the time window of -600 to 600 ms, a plateau-shaped band-pass filter (8 to 13 Hz) with a 15% transition zone was applied to the data. Phase angle time series were extracted using the Hilbert transform, and the temporal derivative of the phase angle time series, indicative of instantaneous frequency (scaled by the sampling rate and 2π), was calculated as the phase varied over time. Previous studies have suggested that noise in the phase angle time series may result in sharp, unphysiological responses in the derivative (13, 65). Therefore, the following disposition was made on the instantaneous frequencies with a median filter of order 10 and a maximum window size of 400 ms. To be more specific, the data were averaged across trials after being median filtered ten times using 10-time windows ranging from 10 to 400 ms. This analysis is theoretically independent of the oscillatory amplitude, except for cases when the amplitude is zero and the phase is indeterminate, as it only considers changes in instantaneous phase. To account for the possibility that differences in trial numbers may lead to discrepancies across conditions, we counterbalanced the number of trials between perceptual outcomes and different conditions. The average instantaneous alpha frequency (IAF) was then calculated for bistable 1-flash and 2-flash trials in F2 and F2B1 conditions.

For the control analysis of instantaneous alpha power, we employed the same filtering and Hilbert transform as utilized in the IAF analysis to extract instantaneous alpha amplitude. The power was then calculated by squaring the amplitude.

Procedure for eliminating 1/f slope effects. To account for the potential effects of the 1/f slope, we reconstructed the time-domain signals that had been demodulated by implementing 1/f attenuation, following the methods outlined in a previous study (33). Specifically, we utilized the Fast Fourier Transform (FFT) to extract the amplitude and phase spectra from each trial of the original signal. Polynomial fits were then applied to the amplitude spectra to obtain the best-fit

values, the inverse of which was multiplied by the amplitude spectrum point-by-point to generate the demodulated amplitude spectrum. To reconstruct the demodulated time-domain signal, we employed the Inverse FFT to combine the phase and amplitude spectra of the original signal into complex values. Next, the demodulated signal underwent a filter with a fixed frequency limit of 8-13 Hz to perform the instantaneous frequency analysis.

Phase opposition analysis. Our primary goal was to investigate whether single-trial perception is influenced by the phase of ongoing oscillatory activity. To achieve this, we utilized the phase opposition sum (POS) method (34) to evaluate the phase difference between the two distinct perceptual outcomes in the F2 and F2B1 conditions. Given that the statistical power of POS is influenced by the absolute trial count within each set of trials (34), and that reliability of POS can be compromised when dealing with unequal 1-flash and 2-flash trial counts (an inherent feature of all phase-based time-frequency analyses methods) (35, 66), we balanced the trial counts for both perceptual outcomes in both F2 and F2B1 conditions.

Time-frequency transformations were first generated over occipital channels using fieldtrip (64). Specifically, we used wavelets with logarithmically spaced cycles (ranging from 2 at the lowest frequency to 7 at the highest frequency) to obtain time-frequency representations (complex numbers) at 60 log-spaced frequency points spanning from 2 to 50 Hz. This yields a complex representation of the amplitude, A, and the phase, φ, for trial j at time t and frequency f:

$$tf_j = A_j(t,f)e^{i\varphi_j(t,f)} \quad [1]$$

The phase of this representation can be extracted by normalizing the complex vector to the unit length:

$$\Phi_j(t,f) = \frac{tf_j(t,f)}{|tf_j(t,f)|} \quad [2]$$

Inter-Trial Phase Coherence (ITC) measures the phase consistency across trials. We calculated the ITC using the method described previously (35):

$$ITC_{tf} = \left|\frac{1}{N}\sum_{n}^{N}\Phi_j(t,f)\right| \quad [3]$$

Where N is the number of trials in one group of trials. Note that the ITC is a scalar rather than a vector and it is always non-negative.

Subsequently, to assess whether trials associated with a specific perceptual outcome (i.e., 1-flash or 2-flash) exhibited differences in phase distribution when compared to trials related to the other perceptual outcome, we computed the POS value using the following formula (34):

$$POS = ITC_A + ITC_B + ITC_{ALL} \quad [4]$$

Where ITC_A and ITC_B are the ITC calculated separately for the two subgroups to be compared (i.e., 1flash or 2flash), and ITC_{ALL} is the ITC calculated across all trials regardless of condition.

Notably, when two conditions exhibit phase clustering at specific, distinct phase angles at a given time and frequency, this signifies that these conditions are phase-clustered with higher ITCs than the total ITC, leading to a higher POS value. Conversely, the POS value will be lower if the different conditions do not display stronger phase clustering or if the phases cluster at approximately the same phase angle.

To rigorously evaluate the significance of phase opposition without assuming any specific probability distribution of the POS values, a non-parametric permutation test was conducted. Initially, the POS values were computed for each point in the time-frequency plane from -600 to 200 ms, ranging from 2 to 50 Hz (in logarithmically spaced cycles) for each occipital electrode and subject, and then averaged over all occipital electrodes and subjects. The surrogate POS values were obtained by randomly assigning trials to one or the other condition for each subject while maintaining a constant number of trials in each condition, and then recalculating the grand mean POS values. The p-value of each time-frequency point was calculated as the proportion of grand mean surrogate POS values that exceeded the empirically observed grand mean POS. To ensure statistical rigor, 100,000 surrogates were employed, and a p-value of 1×10^{-5} was assigned to the points without more extreme POS values in the surrogates.

Statistical analyses

Cluster-based correction was applied when multiple time-frequency points (Fig. 4*A*) or time points (Fig. 2*B*, 2*D-E*, and 3*A*) were tested. For multiple comparison correction of time-frequency points, elements that passed a threshold corresponding to a p-value of 0.05 were marked in each surrogate. Neighboring marked elements were identified as clusters. The count of suprathreshold samples within a cluster was used to define the cluster size, and the largest cluster size was entered into a distribution of cluster size, which was expected under the null hypothesis (35). The p-value of the cluster was calculated as the proportion of the largest cluster size that exceeded the empirically observed largest cluster size. Regarding multiple comparison correction for time points, paired t-statistics were initially computed between different conditions across participants. Elements that passed a threshold of $P < 0.05$ (two-tailed) were marked to identify temporal clusters.

The t-values within a connected cluster were summed up to yield a cluster-level statistic. The p-value of the cluster was determined using a similar shuffling process within each participant.

Simulations of EEG data

Simulations of EEG data were conducted utilizing MATLAB (MathWorks Inc.). The fundamental representation of an oscillator with constant parameters is expressed as:

$$g(t) = \psi\, cos[2\pi f t + \theta]\,. \qquad [5]$$

For each trial, signal (y_t) changes over time according to:

$$y_t = y_{t-1} + \Delta g(t-1) + \varepsilon_{t-1}. \qquad [6]$$

Where ε follows a normal distribution. Aligning with the phase reset hypothesis, which posits that a salient stimulus shifts the phase of ongoing neural oscillation to its optimal excitability phase(31, 54, 67-69), we strategically adjust the stimulus occurrence to synchronize with the good phase range for both experimental conditions. Specifically, we randomly set the initial phase within the good phase range of [-1/2π, 1/2π].

For the F2B1 condition, specifically designed to incorporate sound-induced phase resetting, a phase delay of 25 ms was introduced to the signal after the presentation of a sound stimulus at 0 ms. This delay was randomly varied within the range of 0 to 50 ms.

To simulate behavioral and EEG results, we generated 1000 trials for F2 and F2B1 conditions, respectively, and then implemented eight-level ISIs, ranging evenly from 30ms to 100ms. The perceptual outcome depended on the temporal alignment of the two flashes and the phases at which the second flash occurred. Specifically, subjective perceptual outcome was determined based on whether the two flashes occurred within the same alpha cycle or different. The phase at which the second flash occurred played a crucial role in the clarity of differentiating between 1-flash and 2-flash percepts, thereby directly affecting the reported percepts. We assigned a clarity value of 1 for favorable phases [-1/2π, 1/2π] and a clarity value of 0.5 for unfavorable phases [-π, -1/2π] and [1/2π, π]. Then we calculated the proportion of reported 2-flash percepts for each ISI. Applying a logistic function, we fitted the proportions of 2-flash percepts to determine the fusion threshold for each condition. Subsequently, the thresholds were used as ISI in F2 and F2B1 condition for further frequency and phase analysis, followed the methodology outlined above for the analysis of real EEG data.

Moreover, we calculated the subjective sensitivity (*d'*) for the F2 and F2B1 conditions, following the signal detection theory:

$$d' = z(HR) - z(FAR), \quad [7]$$

where the hit rate (HR) represents the proportion of cases where the subjective perceptual outcome is 2-flash (i.e. the two flashes fall in the different alpha cycles) and the reported outcome is also 2-flash. The false alarm rate (FAR) is the proportion of cases where the subjective perceptual outcome should be 1-flash (i.e. the two flashes fall in the same alpha cycle), but the reported outcome is 2-flash percept.

Data and code availability statement

The data and code utilized in the study are openly accessible on the website (https://osf.io/2hsqw/). They have been made available for reuse or distribution in the public domain, as per the criteria set forth by the institute and in compliance with the ethical committee's endorsement.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (32000785，32000741, 32071052), the Guangdong Natural Science Foundation (2021A1515011185, 2021A1515011100, 2020A1515110223)

References

1. L. Shams, Y. Kamitani, S. Shimojo, Visual illusion induced by sound. *Cognitive brain research* **14**, 147-152 (2002).
2. S. Watkins, L. Shams, S. Tanaka, J.-D. Haynes, G. Rees, Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* **31**, 1247-1256 (2006).
3. J. Cooke, C. Poch, H. Gillmeister, M. Costantini, V. Romei, Oscillatory properties of functional connections between sensory areas mediate cross-modal illusory perception. *Journal of Neuroscience* **39**, 5711-5718 (2019).
4. J. T. Walker, K. J. Scott, Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance* **7**, 1327 (1981).
5. J. W. Gebhard, G. H. Mowbray, On discriminating the rate of visual flicker and auditory flutter. *The American journal of psychology* **72**, 521-529 (1959).
6. T. Shipley, Auditory flutter-driving of visual flicker. *Science* **145**, 1328-1330 (1964).
7. R. B. Welch, L. D. DutionHurt, D. H. Warren, Contributions of audition and vision to temporal rate perception. *Perception & psychophysics* **39**, 294-300 (1986).
8. R. Fendrich, P. M. Corballis, The temporal cross-capture of audition and vision. *Perception & Psychophysics* **63**, 719-725 (2001).
9. C. Scheier, Sound alters visual temporal resolution. *Invest. Ophthalmol. Visual Sci.* **40**, 4169 (1999).
10. J. Samaha, V. Romei, Alpha-band frequency and temporal windows in perception: a review and living meta-analysis of 27 experiments (and counting). *Journal of Cognitive Neuroscience*, 1-15 (2023).
11. T. J. Baumgarten, A. Schnitzler, J. Lange, Beta oscillations define discrete perceptual cycles in the somatosensory domain. *Proceedings of the National Academy of Sciences* **112**, 12187-12192 (2015).
12. R. Cecere, G. Rees, V. Romei, Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology* **25**, 231-235 (2015).
13. J. Samaha, B. R. Postle, The speed of alpha-band oscillations predicts the temporal resolution of visual perception. *Current Biology* **25**, 2985-2990 (2015).
14. N. A. Busch, J. Dubois, R. VanRullen, The phase of ongoing EEG oscillations predicts visual perception. *Journal of neuroscience* **29**, 7869-7876 (2009).
15. N. A. Busch, R. VanRullen, Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences* **107**, 16048-16053 (2010).

16. K. E. Mathewson, G. Gratton, M. Fabiani, D. M. Beck, T. Ro, To see or not to see: prestimulus α phase predicts visual awareness. *Journal of Neuroscience* **29**, 2725-2732 (2009).
17. R. Chakravarthi, R. VanRullen, Conscious updating is a rhythmic process. *Proceedings of the National Academy of Sciences* **109**, 10599-10604 (2012).
18. A. Milton, C. W. Pleydell-Pearce, The phase of pre-stimulus alpha oscillations influences the visual perception of stimulus timing. *Neuroimage* **133**, 53-61 (2016).
19. T. Rohe, A.-C. Ehlis, U. Noppeney, The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications* **10**, 1907 (2019).
20. R. Gulbinaite, B. İlhan, R. VanRullen, The triple-flash illusion reveals a driving role of alpha-band reverberations in visual perception. *Journal of Neuroscience* **37**, 7219-7230 (2017).
21. L. Ronconi, N. N. Oosterhof, C. Bonmassar, D. Melcher, Multiple oscillatory rhythms determine the temporal organization of perception. *Proceedings of the National Academy of Sciences* **114**, 13435-13440 (2017).
22. B. Han, Y. Zhang, L. Shen, L. Mo, Q. Chen, Task demands modulate pre-stimulus alpha frequency and sensory template during bistable apparent motion perception. *Cerebral Cortex* (2022).
23. L. Shen, B. Han, L. Chen, Q. Chen, Perceptual inference employs intrinsic alpha frequency to resolve perceptual ambiguity. *PLoS biology* **17**, e3000025 (2019).
24. J. Keil, D. Senkowski, Individual alpha frequency relates to the sound-induced flash illusion. *Multisensory Research* **30**, 565-578 (2017).
25. M. Ruzzoli, M. Torralba, L. M. Fernández, S. Soto-Faraco, The relevance of alpha phase in human perception. *Cortex* **120**, 249-268 (2019).
26. M. J. Gray, T. A. Emmanouil, Individual alpha frequency increases during a task but is unchanged by alpha‐band flicker. *Psychophysiology* **57**, e13480 (2020).
27. S. Buergers, U. Noppeney, The role of alpha oscillations in temporal binding within and across the senses. *Nature human behaviour* **6**, 732-742 (2022).
28. T. S. Andersen, K. Tiippana, M. Sams, Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research* **21**, 301-308 (2004).
29. C. Kayser, C. I. Petkov, N. K. Logothetis, Visual modulation of neurons in auditory cortex. *Cerebral Cortex* **18**, 1560-1574 (2008).
30. J. Keil, N. Müller, T. Hartmann, N. Weisz, Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cerebral Cortex* **24**, 1278-1288 (2014).
31. P. Lakatos, C.-M. Chen, M. N. O'Connell, A. Mills, C. E. Schroeder, Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* **53**, 279-292 (2007).

32. M. R. Mercier *et al.*, Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage* **79**, 19-29 (2013).
33. J. Samaha, M. X. Cohen, Power spectrum slope confounds estimation of instantaneous oscillatory frequency. *NeuroImage* **250**, 118929 (2022).
34. R. VanRullen, How to evaluate phase differences between trial groups in ongoing electrophysiological signals. *Frontiers in neuroscience* **10**, 426 (2016).
35. M. X. Cohen, *Analyzing neural time series data: theory and practice* (MIT press, 2014a).
36. P. Lakatos *et al.*, The leading sense: supramodal control of neurophysiological context by attention. *Neuron* **64**, 419-430 (2009).
37. M. R. Mercier *et al.*, Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. *Journal of Neuroscience* **35**, 8546-8557 (2015).
38. R. VanRullen, Perceptual cycles. *Trends in cognitive sciences* **20**, 723-735 (2016).
39. A. K. Engel, P. Fries, W. Singer, Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience* **2**, 704-716 (2001).
40. R. Land, G. Engler, A. Kral, A. K. Engel, Auditory evoked bursts in mouse visual cortex during isoflurane anesthesia. *PLoS One* **7**, e49855 (2012).
41. R. A. Oeur, M. Palaniswamy, M. Ha, M. Fernandez-Corazza, S. S. Margulies, Regional variations distinguish auditory from visual evoked potentials in healthy 4 week old piglets. *Physiological Measurement* **44**, 025006 (2023).
42. C. C. Canavier, S. Wang, L. Chandrasekaran, Effect of phase response curve skew on synchronization with and without conduction delays. *Frontiers in neural circuits* **7**, 194 (2013).
43. S. Makeig, S. Debener, J. Onton, A. Delorme, Mining event-related brain dynamics. *Trends in cognitive sciences* **8**, 204-210 (2004).
44. A. S. Shah *et al.*, Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral cortex* **14**, 476-483 (2004).
45. Y. Noguchi, Individual differences in beta frequency correlate with the audio–visual fusion illusion. *Psychophysiology* **59**, e14041 (2022).
46. S. Watkins, L. Shams, O. Josephs, G. Rees, Activity in human V1 follows multisensory perception. *Neuroimage* **37**, 572-578 (2007).
47. V. Romei, J. Gross, G. Thut, Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current biology* **22**, 807-813 (2012).

48. H. Luo, Z. Liu, D. Poeppel, Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS biology* **8**, e1000445 (2010).
49. A. Wutz, N. Weisz, C. Braun, D. Melcher, Temporal windows in visual processing:“prestimulus brain state” and “poststimulus phase reset” segregate visual transients on different temporal scales. *Journal of Neuroscience* **34**, 1554-1565 (2014).
50. A. Wutz, D. Melcher, J. Samaha, Frequency modulation of neural oscillations according to visual task demands. *Proceedings of the National Academy of Sciences* **115**, 1346-1351 (2018).
51. P. G. Schyns, G. Thut, J. Gross, Cracking the code of oscillatory activity. *PLoS biology* **9**, e1001064 (2011).
52. M. A. Montemurro, M. J. Rasch, Y. Murayama, N. K. Logothetis, S. Panzeri, Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current biology* **18**, 375-380 (2008).
53. R. VanRullen, N. A. Busch, J. Drewes, J. Dubois, Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Frontiers in psychology* **2**, 60 (2011).
54. A. M. Harris, Phase resets undermine measures of phase-dependent perception. *Trends in Cognitive Sciences* (2023).
55. W. R. Gruber *et al.*, Alpha phase, temporal attention, and the generation of early event related potentials. *Neuroimage* **103**, 119-129 (2014).
56. C. Keitel, M. Ruzzoli, L. Dugué, N. A. Busch, C. S. Benwell (2022) Rhythms in cognition: The evidence revisited. (Wiley Online Library), pp 2991-3009.
57. C. S. Benwell, A. Coldea, M. Harvey, G. Thut, Low pre‐stimulus EEG alpha power amplifies visual awareness but not visual sensitivity. *European Journal of Neuroscience* **55**, 3125-3140 (2022).
58. C. S. Benwell *et al.*, Prestimulus EEG power predicts conscious awareness but not objective visual performance. *eneuro* **4** (2017).
59. R. M. van Diepen, M. X. Cohen, D. Denys, A. Mazaheri, Attention and temporal expectations modulate power, not phase, of ongoing alpha oscillations. *Journal of cognitive neuroscience* **27**, 1573-1586 (2015).
60. B. S. W. Ng, T. Schroeder, C. Kayser, A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *Journal of Neuroscience* **32**, 12268-12276 (2012).
61. T. Neuling, S. Rach, S. Wagner, C. H. Wolters, C. S. Herrmann, Good vibrations: oscillatory phase shapes perception. *Neuroimage* **63**, 771-778 (2012).

62. S. Ten Oever, T. Meierdierks, F. Duecker, T. A. De Graaf, A. T. Sack, Phase-coded oscillatory ordering promotes the separation of closely matched representations to optimize perceptual discrimination. *Iscience* **23** (2020).
63. A. R. Bauer, S. Debener, A. C. Nobre, Synchronisation of Neural Oscillations and Cross-modal Influences. *Trends Cogn Sci* **24**, 481-495 (2020).
64. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* **2011**, 1-9 (2011).
65. M. X. Cohen, Fluctuations in oscillation frequency control spike timing and coordinate neural networks. *Journal of Neuroscience* **34**, 8988-8998 (2014b).
66. M. Vinck, M. van Wingerden, T. Womelsdorf, P. Fries, C. M. Pennartz, The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *Neuroimage* **51**, 112-122 (2010).
67. J. D. Thorne, M. De Vos, F. C. Viola, S. Debener, Cross-modal phase reset predicts auditory task performance in humans. *Journal of Neuroscience* **31**, 3853-3861 (2011).
68. P. Lakatos *et al.*, An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of neurophysiology* **94**, 1904-1911 (2005).
69. P. Lakatos, G. Karmos, A. D. Mehta, I. Ulbert, C. E. Schroeder, Entrainment of neuronal oscillations as a mechanism of attentional selection. *science* **320**, 110-113 (2008).

Figures

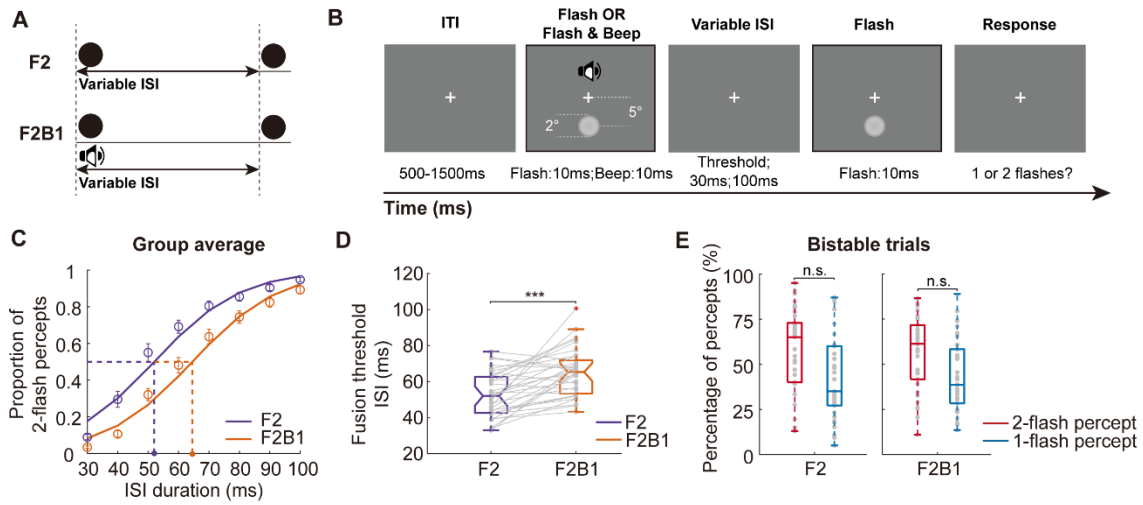

Fig. 1. Paradigm and behavioral results. (A) Schematic illustration of the F2 and F2B1 conditions. (B) Illustration of one trial sequence. On each trial, the first visual stimulus presented below the fixation, either accompanied by an auditory beep or without it (the sound icon in the figure is not part of the actual experiment). After a variable ISI, the second visual stimulus was presented. Participants were required to report whether they perceived one or two flashes. Participants were more likely to report perceiving one flash with a short ISI (30 ms) and two flashes with a long ISI (100 ms). Perception became bistable during trials where the ISI matched fusion threshold, individually determined for each participant, resulting in approximately equal proportions of 1-flash and 2-flash reports. (C) Psychometric curves represent the best fit of the average probability of perceiving the two flashes plotted as a function of different ISIs in the F2 and F2B1 conditions. Error bars represent ±1 SEM. (D) Fusion threshold IFIs derived from the psychophysical procedure for each condition. (E) Percentage of 1-flash and 2-flash percepts for the bistable trials in the F2 and F2B1 conditions of the main EEG experiment. ***$p < 0.001$.

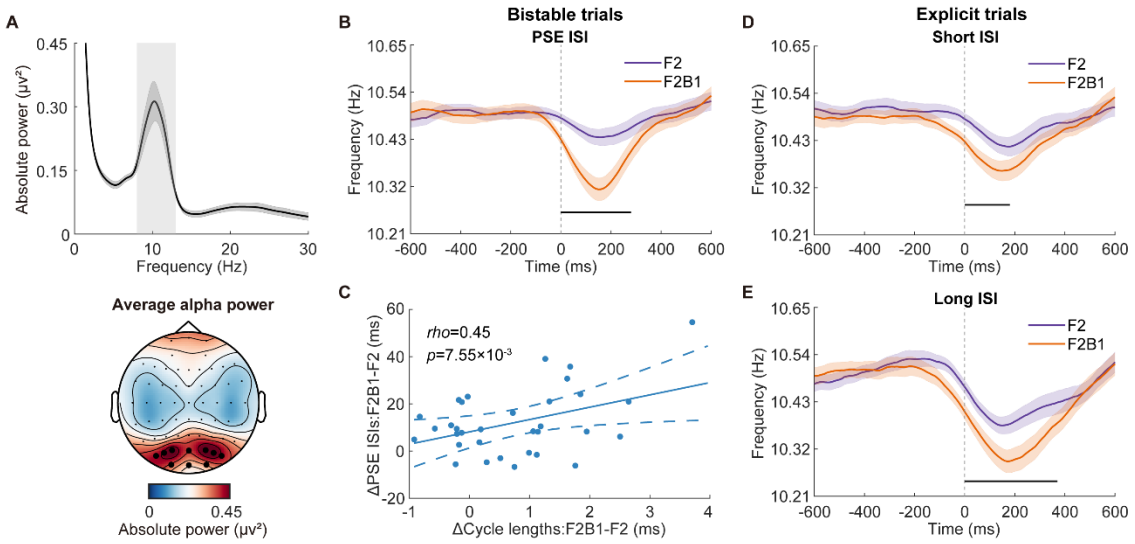

Fig. 2. The relationship between decreased post-stimulus IAFs and perceptual precision. (A) *Upper panel*: the power spectrum obtained from all bistable trials (from -600 to 600 ms relative to the onset of the first flash) in F2 and F2B1 conditions, collapsed across all electrodes in all participants, clearly revealed a distinct peak in the alpha band. The gray shading represents ±1 within-subjects SEM. The light gray rectangle represents the chosen alpha frequency band (8-13Hz). *Lower panel*: Topography of absolute alpha-band power (8-13Hz) recorded from 64-channel EEG, averaged over F2 and F2B1 conditions, showed a clear occipital scalp distribution. Black dots indicated the chosen posterior channels. (B) IAFs for F2 and F2B1 bistable trials averaged over the occipital channels. The within-subject findings of the IAFs revealed that the post-stimulus alpha frequency of F2B1 decreased more than that of F2. (C) Between-subject correlation was conducted to examine the relationship between the post-stimulus IAFs difference of F2B1 and F2 in the main EEG experiment and the threshold ISIs difference of F2B1 and F2 in the psychophysical pretest. 95% confidence intervals around the linear fit line are shown by the dashed line. (D) Result identical to (B), but for explicit short ISI trials. (E) Result identical to (B), but for explicit long ISI trials.

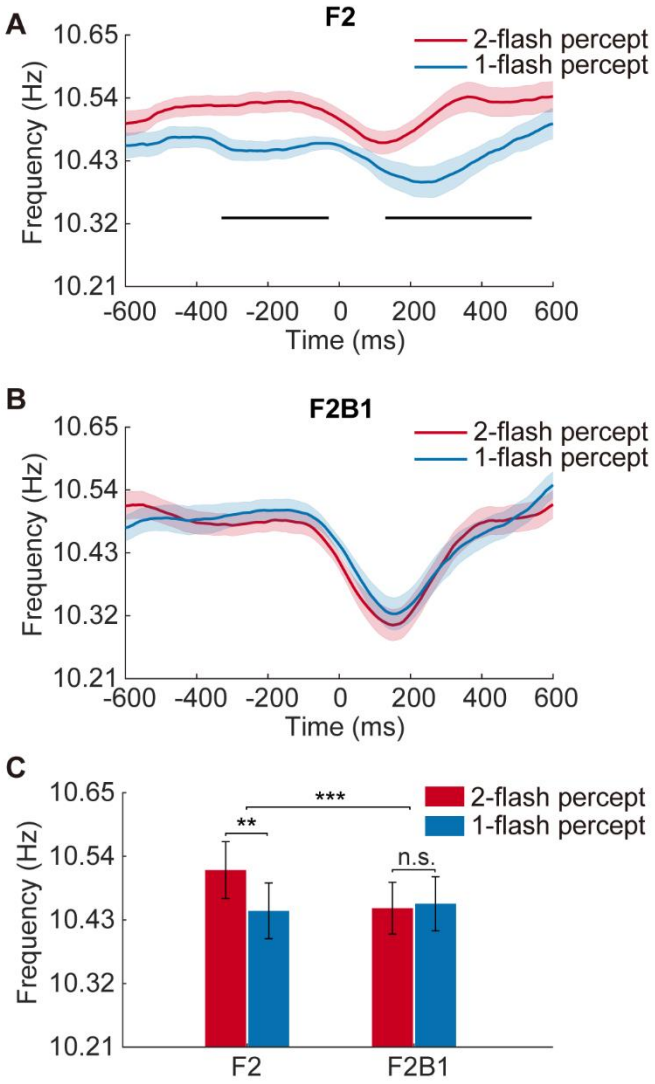

Fig. 3. Relationship between the instantaneous alpha frequency (IAF) and perception. (A) IAFs over time were averaged over posterior channels (shown in A) for different perceptual outcomes in the F2 condition. Shaded areas show ±1 within-subjects SEM. Significant time epochs are indicated with black lines ($p < 0.05$; permutation test; cluster corrected). (B) Similar to (A), but for F2B1 condition. (C) IAFs during the time of interest (-600 to 600 ms) were averaged separately for F2 and F2B1 conditions with different perceptual outcomes. The error bars represent ±1 SEM. $^{**}p < 0.01$. $^{***}p < 0.001$.

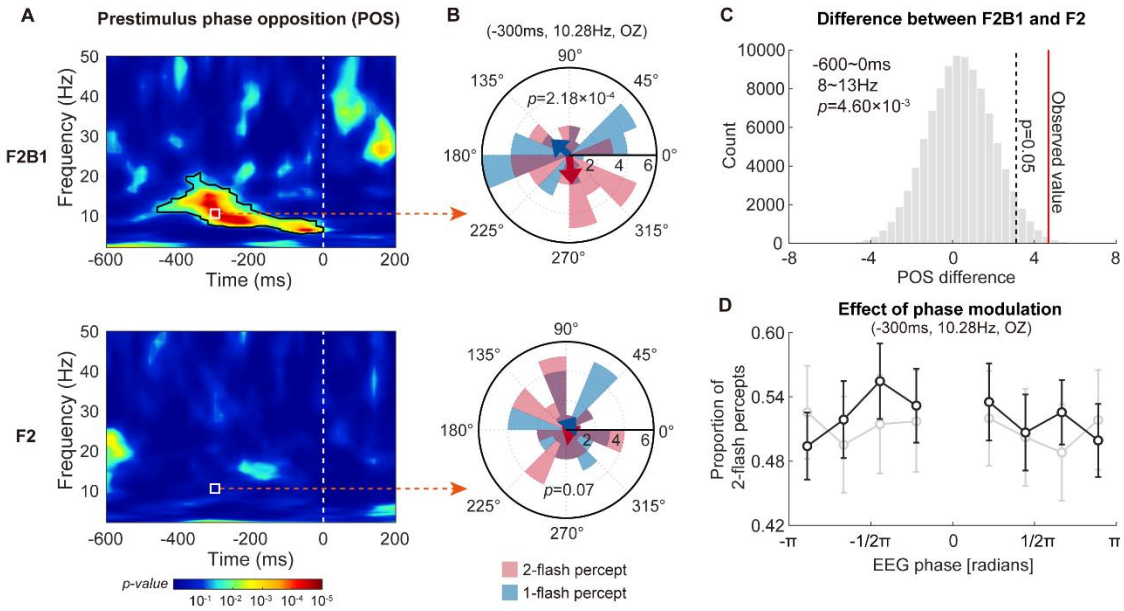

Fig.4. Prestimulus alpha phase difference between 1-flash and 2-flash trials. (A) Time-frequency representation of p values (for occipital regions), computed as a proportion of surrogate phase opposition values (distribution of phase opposition values expected under null hypothesis) that exceeded empirically observed phase opposition values for the F2B1 (*upper panel*) and F2 conditions (*lower panel*), respectively. The outlined area indicates significant effects corrected for multiple comparisons using cluster-based correction. The white square indicates the time-frequency point with strongest POS effect in the F2B1 condition. (B) The circular histograms of mean phase angles (at 10.28Hz; -300ms) at the occipital channel Oz during 1-flash and 2-flash trials across participants for the F2B1 (*upper panel*) and F2 conditions (*lower panel*), respectively. The horizontal black line indicates the number of participants with the mean phase angle in each bin. The direction of the arrows corresponds to the mean phase angle across participants, and the length of the arrows indicates the extent of phases clustering around the mean. (C) Statistical analysis of the differential phase effects between F2B1 and F2 condition within the entire prestimulus period (-600~0ms) and alpha-band range (8~13Hz). The red line indicates the observed mean POS difference between F2B1 and F2 within this cluster. The gray bars represent the surrogate distribution of mean POS difference expected by chance for this cluster, estimated from shuffled data. Dashed black line indicates the 95th percentile value of the surrogate distribution. (D) The relationship between the proportion of two-flash percepts and pre-stimulus phase (at 10.28Hz; -300ms) at the occipital channel Oz. Single trials were categorized into 9 phase bins, centered on the preferred phase bin with the highest proportion of 2-flash percepts for each participant (the central bin is removed due to artificial shifting). The error bars represent ±1 SEM.

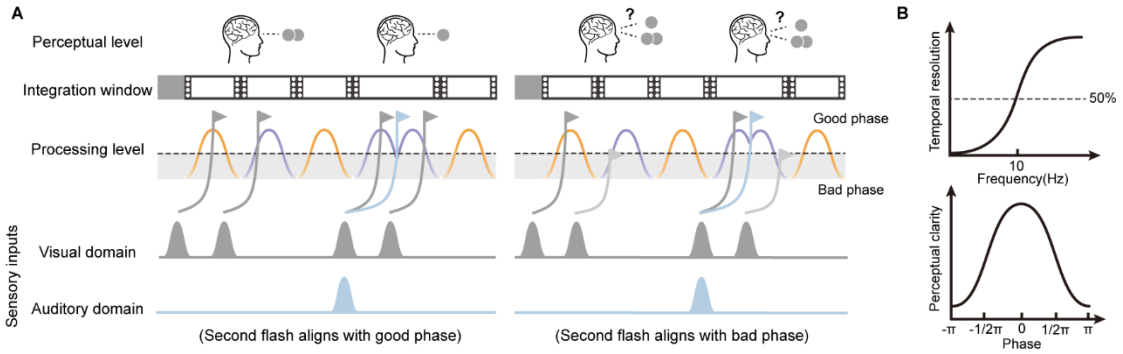

Fig. 5. Hypothesized phase-resetting model of perceptual processing. (A) The perceptual outcomes depend on the temporal alignment of two stimuli and the phase at which the second flash occurs. When the second flash coincides with good phases, around the peak of the oscillation, effective processing occurs, and the perceptual outcomes are determined based on whether they fall within the same temporal window defined by the alpha cycle. In cases where the two visual stimuli fall in different temporal windows, a 2-flash percept is reported. The introduction of auditory stimuli concurrently with visual stimuli induces phase resetting, elongating the temporal window. Consequently, the second flash, originally falling in different temporal windows, aligns with the first flash, resulting in a 1-flash percept. Conversely, when the second flash aligns with bad phases, around trough of the oscillation, effective processing is hindered, leading to ambiguity in 1-flash and 2-flash percepts, irrespective of their temporal alignment. (B) *Upper*: The frequency of the alpha rhythm is anticipated to represent temporal resolution, with higher alpha frequencies correlating with an increased proportion of 2-flash percepts. *Lower*: The phase of the alpha rhythm is expected to dictate perceptual clarity. In instances where the second flash falls in bad phases (e.g., [-π, -1/2π] and [1/2π, π]), perceptual clarity is low, making it challenging to distinguish between 1-flash and 2-flash percepts. Conversely, when the second flash aligns with good phases (e.g., [-1/2π, 1/2π]), perceptual clarity is high, and the perception of 1-flash or 2-flash depends on whether the two stimuli fall within the same alpha cycle or not.

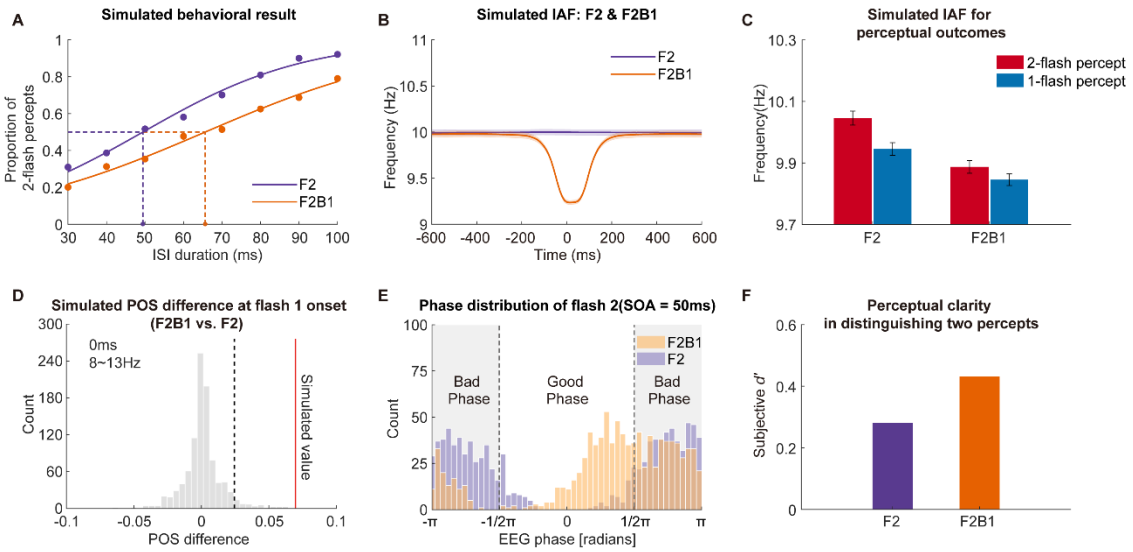

Fig.6. Simulation results. (A) Psychophysical results. Psychometric curves depict the best-fit average probability of perceiving two flashes plotted against different ISIs in simulated F2 and F2B1 conditions. (B) IAFs for simulated F2 and F2B1 trials. (C) IAFs, averaged during the time of interest (-600 to 600 ms), separately for simulated F2 and F2B1 trials with different perceptual outcomes. Error bars represent ±1 within-trials SEM. (D) Statistical analysis of the differential phase effects between F2B1 and F2 condition at stimulus onset (0 ms), prior to the auditory input-induced phase-resetting. The red line indicates the mean POS difference between simulated F2B1 and F2 trials. Dashed black line indicates the 95th percentile value of the surrogate distribution. (E) Phase distribution at the onset of flash 2 (ISI = 50 ms; following the phase-resetting effect). (F) Subjective perceptual sensitivity in distinguishing between 1-flash and 2-flash percepts for simulated F2 and F2B1 trials.