Title page

Cortical entrainment to hierarchical contextual rhythms recomposes dynamic attending in visual perception

Peijun Yuan[1, 2], Ruichen Hu[1, 2], Xue Zhang[1, 2], Ying Wang[1, 2]*, Yi Jiang[1, 2]*

[1]State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, P. R. China

[2]Department of Psychology, University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, P. R. China

*For correspondence: wangying@psych.ac.cn; yijiang@psych.ac.cn

Abstract

Temporal regularity is ubiquitous and essential to guiding attention and coordinating behavior within a dynamic environment. Previous researchers have modeled attention as an internal rhythm that may entrain to first-order regularity from rhythmic events to prioritize information selection at specific time points. Using the attentional blink paradigm, here we show that higher-order regularity based on rhythmic organization of contextual features (pitch, color, or motion) may serve as a temporal frame to recompose the dynamic profile of visual temporal attention. Critically, such attentional reframing effect is well predicted by cortical entrainment to the higher-order contextual structure at the delta band as well as its coupling with the stimulus-driven alpha power. These results suggest that the human brain involuntarily exploits multiscale regularities in rhythmic contexts to recompose dynamic attending in visual perception, and highlight neural entrainment as a central mechanism for optimizing our conscious experience of the world in the time dimension.

Introduction

Deploying attention over time is crucial for guiding human activities within a rapidly changing environment. However, the constant influx of information goes far beyond our mental capacity, impeding even the most competent human brain from capturing every nuance of the details. How does the human brain surmount such limitations in temporal attention allocation during dynamic information processing?

One feasible solution, as that for spatial attention, is through selection, or by shining an attentional "spotlight" on the most relevant information while filtering out the irrelevant regarding the task demands (Posner, 1980). When it comes to the temporal domain, people tend to utilize regularities in the sensory information flow for directing attention to the moments when a target event is expected to occur (Nobre et al., 2007; Nobre & van Ede, 2018). As a great example, Jones and colleagues have shown in a series of studies that after listening to a rhythmic tone sequence, auditory perception in terms of pitch judgment and time discrimination was more accurate for target tones appearing at the expected than the unexpected time points (Jones et al., 2002; Large & Jones, 1999). Such facilitation effects have been extended to various aspects of visual perception and even across sensory modalities (Bolger et al., 2014; Brochard et al., 2013; Mathewson et al., 2010; Miller et al., 2013; ten Oever et al., 2014), implicating the involvement of a general attentional selection mechanism guided by the regularity in stimulus timing.

In this line of studies, perceptual responses were significantly improved for targets appearing within a rhythmic context but not within an arrhythmic context. These findings can be interpreted by the dynamic attending theory (DAT), which assumes attention as an internal oscillatory activity (or attending rhythm) that can be entrained to rhythmic structures of the exogenous events (Jones, 1976; Jones et al., 1981; Jones & Boltz, 1989; Large & Jones, 1999). In line with this assumption, electrophysiological research in humans and non-human primates have found entrainment of intrinsic neural oscillations to external stimulus rhythms, and regarded such process as an instrument for selective attention (Calderone et al., 2014; Obleser & Kayser, 2019; Schroeder & Lakatos, 2009). Through neural entrainment, neuronal excitability aligns with the occurrence of rhythmic

events, creating "temporal attentional spotlights" that attract the brain's attentional resources towards a string of selected moments (Calderone et al., 2014; Henry & Herrmann, 2014; Lakatos et al., 2008, 2013; Schroeder & Lakatos, 2009).

The synchronization between the internal attending rhythm and the external rhythms allows us to direct attention proactively and enhance perception at the anticipated moments. With regards to forming a coherent perception of the dynamic environment, however, we should not only select information bound to the anticipated time points, but also allocate attentional resources among these points, raising the problem of dynamic attentional deployment over an information stream. For instance, when viewing a rapid serial visual presentation (RSVP) stream, there is a large chance that the observer would miss the second of two temporally proximate targets, as the allocation of attention to the first target hinders the redeployment of mental resources to the second one (Broadbent & Broadbent, 1987). This phenomenon, vividly referred to as the attentional blink (AB)(Chun & Potter, 1995; Raymond et al., 1992), has attracted much interest as it reveals the limitations of attentional allocation and memory processes that may become a bottleneck for conscious awareness (Dux & Marois, 2009; Martens & Wyble, 2010; Shapiro et al., 1997). More intriguingly, as items in the RSVP stream are all rhythmically presented and temporally predictable, the AB effect poses a challenge in dynamic attending that cannot be circumvented solely by the anticipation built upon stimulus timing.

To address this challenge, here we propose that, the brain has to rely, as a complement to the first-order regularity in rhythmic stimulation, on regularities in the higher-order temporal structure of the information stream. More specifically, if the endogenous attentional rhythm could entrain automatically not only to the stimulus rhythm but also to the higher-order structure based on the information content, the deployment of temporal attention might be reconstructed in a way that facilitates target detection in the AB task. To test this hypothesis, we synchronized the original AB stream (stimulation rate at 10 Hz) to a hierarchical contextual stream that possessed a feature-based temporal structure—a 2.5Hz rhythm arising from periodic changes of a physical feature, superimposed on its stimulus rhythm at 10Hz. Using temporal structures defined

by a variety of features (pitch, color, etc.), we provided converging evidence that the structured context, which was task-irrelevant and even from a different modality, could regulate the dynamic deployment of visual attention so as to alleviate the AB effect. To further unravel the neural basis of the observed attentional modulation effect, we conducted an electroencephalogram (EEG) experiment. We are particularly interested in whether neural oscillations can entrain to the contextual temporal structure of stimulus feature along with that of stimulus onset timing, and more critically, whether and how the cortical entrainment to these hierarchical structures mediates the behavioral modulation effect.

Results

Temporal structure of contextual auditory stream recomposes visual attentional deployment

In Experiment 1a, we first explored whether feature-defined temporal structure from a contextual auditory stream could regulate visual attentional deployment during the AB task. If so, the AB effect should be modulated by the positions of the visual targets relative to the rhythmic structure arising from periodic changes of the background sounds (Fig. 1A). Above all, we found a robust AB effect in the short-SOA conditions, no matter whether there were contextual sounds or not. The T2 detection accuracy conditioned on correct T1 response was generally impaired in the short-SOA conditions relative to that in the long-SOA condition, during both the context session (long-SOA: 0.945 ± 0.016 (mean ± se), short-SOA: 0.728 ± 0.034, $t(15) = 5.443$, $p < .001$, Cohen's $d = 1.361$) and the baseline session (long-SOA: 0.950 ± 0.013, short-SOA: 0.702 ± 0.034, $t(15) = 6.720$, $p < .001$, Cohen's $d = 1.680$).

More importantly, looking close at T2 performance in the short-SOA conditions (Fig. 1B), we found T2 was better identified when two targets appeared in two adjacent cycles (between-cycle condition) than within the same cycle defined by the background sounds (within-cycle condition). Notably, such difference was observed only for the context session ($t(15) = 2.947$, $p = .010$, Cohen's $d = 0.737$) but not for the baseline (no sound) session ($t(15) = -0.212$, $p = .835$, Cohen's $d = 0.053$), although the target positions were

completely matched between these two sessions. Meanwhile, only in the between-cycle condition, the contextual sounds enhanced T2 detection accuracy relative to the baseline ($t(15) = 2.287$, $p = .037$, Cohen's $d = 0.572$), while in the within-cycle condition, the performance kept comparable between the context and baseline sessions ($t(15) = -0.271$, $p = .790$, Cohen's $d = 0.068$). The observed dissociation was further confirmed by a two-way repeated-measures ANOVA, which yielded a significant interaction between experimental session (baseline vs. context session) and target position (between- vs. within-cycle, defined by the context) ($F(1, 15) = 7.151$, $p = .017$, $\eta_p^2 = .323$).

Results from Experiment 1a demonstrated that feature-based temporal structure of an auditory stream, though being task-irrelevant, could systematically modulate the allocation of visual attention over the AB stream. Since the temporal structure of the contextual sounds was defined by periodic change of pitch, when two targets were located in distinct cycles as in the between-cycle condition, they were accompanied by different tones, in contrast to that when located within the same cycle they were accompanied by the same tone. It is possible that the contrast of physical stimulation (i.e., pitch) at T1 and T2 could account for the performance improvement in the between-cycle condition. To test this possibility, in Experiment 1b, we matched the pitch of tones at target occurrence with that in Experiment 1a for the between- and within-cycle condition respectively, whereas disrupted feature-based regularity in the temporal structure of the contextual sound sequence (Fig. 1A, bottom). Despite that the sounds paired with the targets were exactly the same as in Experiment 1a, the difference in T2 detection accuracy caused by the contextual sounds was no longer observed ($t(15) = 0.433$, $p = .671$, Cohen's $d = 0.108$), neither was its interaction with experimental session ($F(1, 15) = 2.734$, $p = .119$, $\eta_p^2 = .154$; Fig. 1C). In other words, T2 was identified with similar accuracy across all the conditions in Experiment 1b, suggesting that it is the temporal structure of the contextual sounds, not the pitch difference at target presentation, that accounts for the between-cycle facilitation effect observed in Experiment 1a.

Generalization of the modulation effect to different cycle frequencies

In Experiment 1a, the auditory context always changed its pitch value every four items, i.e., every 400 ms as one cycle, resulting in rhythmic cycles at 2.5 Hz. In Experiment 1c, we tested whether the modulation effect we observed could be generalized to other cycle frequencies. We set the pitch change rate to 2 Hz (i.e., five items per cycle; Fig. 2A, upper) and 3.3 Hz (i.e., three items per cycle; Fig. 2A, lower). For both context frequencies, the T2 detection performance in the between-cycle condition was significantly higher than that in the within-cycle condition (Fig. 2B; for 2 Hz, $t(15) = 3.478$, $p = .004$, Cohen's $d = 0.869$; for 3.3 Hz, $t(15) = 2.467$, $p = .030$, Cohen's $d = 0.617$), suggesting successful attentional modulation effects. Furthermore, a repeated-measures ANOVA on T2 accuracy revealed only a significant main effect of relative target position (i.e., between- vs. within-cycle) ($F(1, 15) = 23.320$, $p < .001$, $\eta_p^2 = .609$), with a marginally significant main effect of frequency ($F(1, 15) = 4.337$, $p = .055$, $\eta_p^2 = .224$) and no interaction between these two factors ($F(1, 15) = 0.204$, $p = .658$, $\eta_p^2 = 0.013$).

The effect of temporal attention rather than perceptual grouping

As temporal structure of the context was constructed by auditory items sharing the same feature (i.e., pitch), one may argue that perceptual grouping on the basis of similarity (Bregman, 1994), instead of dynamic attending guided by feature-based temporal regularities, contributes to the between-cycle benefit that we observed. To disentangle these factors, in Experiment 1d, we changed the pitch value of tone sequences irregularly to form auditory streams that could be grouped in varying lengths (Fig. 2C, upper). Though temporal grouping was reserved in this setting, no facilitation effect was observed when targets were separated in two distinct groups relative to when they were displayed within the same group (Fig. 2D, Irreg-G). T2 detection performance was comparable in the between- and the within-group conditions ($t(15) = 0.348$, $p = .733$, Cohen's $d = 0.087$).

Compared with Experiments 1a and 1c, the strength of temporal grouping in Experiment 1d might be attenuated due to irregular number of items in each group, which could lead to the lack of behavioral modulation effect. To solve this issue, in Experiment 1e (Fig. 2C, lower), we changed the pitch every four items to keep the rule of temporal

grouping exactly the same as that in Experiment 1a. Nevertheless, we disrupted the regularity of stimulus timing. Such manipulation would have a detrimental impact on dynamic deployment of temporal attention in general, according to the basic assumption of the DAT (Jones et al., 1982; Jones & Boltz, 1989; Large & Jones, 1999). On the other hand, it would have little influence on the grouping effect. Therefore, if temporal attention rather than perceptual grouping is essential to the behavioral modulation effect observed in the current study, we should expect such effect to disappear in Experiment 1e. In line with our speculation, when the stimulus onset timing was randomized, T2 detection performance in the between-cycle condition was no longer improved relative to the within-cycle condition (Fig. 2D, Irreg-T; $t(15) = 0.302$, $p = .767$, Cohen's $d = 0.076$), despite the potential benefit of the grouping effect. Putting together, the absence of context-induced modulation effect in Experiments 1d and 1e consistently supports the idea that temporal grouping without dynamic attending guided by feature- and timing-related regularities in the auditory context is insufficient to cause the behavioral modulation effect.

Temporal regularities in color-defined rhythmic structure recompose visual attentional deployment

Information from the auditory modality, like speech and music, is inherently organized in time and provides rich sources of rhythmic structures that can be proactively tracked by the human brain (Arnal & Giraud, 2012; Doelling & Poeppel, 2015; Haegens & Zion Golumbic, 2018; Zion Golumbic et al., 2012). This suggests a possibility that the role of rhythmic structure in guiding attention is exclusive to auditory context, which may explain the findings from Experiment 1 that temporal structures generated by rhythmic changes of auditory signals in the background automatically modulate the AB effect. To test this idea, we designed Experiment 2 to directly investigate whether temporal structures based on the change of visual properties would exert a similar influence on temporal attentional deployment. In Experiment 2a, we used visual patterns with periodic change in background color as the temporal context while observers were performing the same AB task (Fig. 3A). As a control experiment, Experiment 2b followed the same logic for Experiment 1b, in which we destroyed the structure of the visual context by changing the

background color in random orders, but kept the background color presented with the targets the same as that in Experiment 2a (Fig. 3C).

Similar to findings obtained from Experiment 1a, the interaction between experimental session (baseline vs. context session) and target position (between- vs. within-cycle) was significant in Experiment 2a (Fig. 3B; $F(1, 15) = 5.180$, $p = .038$, $\eta_p^2 = .257$). In the context session only, T2 performance in the between-cycle condition was better than that in the within-cycle condition ($t(15) = 3.538$, $p = .003$, Cohen's $d = 0.885$). Compared with the baseline session, T2 performance was only improved in the between-cycle condition ($t(15) = 2.274$, $p = .038$, Cohen's $d = 0.569$). By contrast, in Experiment 2b, we did not observe a significant facilitation effect in the between-cycle condition compared with the within-cycle condition ($t(15) = -1.176$, $p = .258$, Cohen's $d = 0.294$) or with its counterpart in the baseline session ($t(15) = 0.685$, $p = .504$, Cohen's $d = 0.171$), nor did we observe the interaction between experimental session and target position (Fig. 3D; $F(1, 15) = 1.435$, $p = .250$, $\eta_p^2 = .087$). These findings suggest that the utilization of feature-based temporal regularities in attentional guidance is a fundamental principle that holds true not only for auditory but also for visual processing.

Excluding the impact of structure boundary: evidence from motion context

So far, results from Experiments 1 and 2 have demonstrated a general regulatory effect that feature-based temporal structure from task-irrelevant information recomposed visual attentional allocation during the AB task, which could be exerted within the same or cross different sensory modalities. In both experiments, however, the switch from one feature-based rhythmic cycle to another was always accompanied by an abrupt change in physical features (pitch or color), resulting in explicit boundaries before T2 presentation in the between-cycle but not in the within-cycle condition. This abrupt change may serve as an attentional cue or alerting signal for the upcoming T2, and thus accounts for the improvement of performance in the between-cycle condition. To examine this possibility, in Experiment 3, we introduced a cyclic motion context that possessed feature-based rhythmicity identical to those contextual rhythms in previous experiments (for more details, see Methods) but had no abrupt boundaries between cycles (Fig. 3E). Once again,

we observed significant improvement of T2 performance in the between-cycle condition relative to the within-cycle condition in the cyclic motion session (Fig. 3F; $t(15) = 2.674$, $p = .017$, Cohen's $d = 0.669$), but this was not the case in the random motion session ($t(15) = -0.330$, $p = .746$, Cohen's $d = 0.082$), resulting in a significant interaction between experimental session (baseline vs. context session) and target position (between- vs. within-cycle): $F(1, 15) = 9.253$, $p = .008$, $\eta_p^2 = 0.382$. These results provide compelling evidence that explicit perceptual boundaries are not necessary for the temporal structure in the context to regulate the allocation of attentional resources.

EEG experiment: The role of neural entrainment in regulating attentional deployment

Neural tracking of higher-order temporal structure of contextual rhythms predicts the behavioral modulation effect

To investigate the neural mechanisms underlying the observed context-induced effect, we carried out an EEG experiment using the same task as that in Experiment 1a. First of all, we replicated the behavioral modulation effect that T2 performance was significantly better in the between-cycle condition versus the within-cycle condition, only in the context session (between-cycle: 0.567 ± 0.036, within-cycle: 0.520 ± 0.039, $t(15) = 3.838$, $p = .002$, Cohen's $d = 0.960$) but not in the baseline session (between-cycle: 0.519 ± 0.039, within-cycle: 0.527 ± 0.043, $t(15) = 0.296$, $p = .771$, Cohen's $d = 0.074$). Furthermore, to identify the oscillatory characteristics of EEG signals in response to stimulus rhythms, we examined the FFT spectral peaks by subtracting the mean power of two nearest neighboring frequencies from the power at the stimulus frequency. Power spectrum in Fig. 4A shows several peaks for the context session, with the highest at 10 Hz (compared with zero using one-sample *t*-test, right-tailed, $t(15) = 10.610$, $p < .001$, FDR-corrected for multiple comparisons across frequencies) corresponding to the common stimulation frequency of the visual and auditory streams. More importantly, the second-highest peak appeared at 2.5 Hz ($t(15) = 5.730$, $p < .001$, FDR-corrected), followed by its harmonics at 5 and 7.5 Hz, indicating neural tracking of the feature-defined structure of the auditory context. In contrast with the observation in the context session, we only found significant

power peak at 10 Hz ($t(15) = 9.405$, $p < .001$, FDR-corrected), but not at 2.5 Hz ($t(15) = 0.301$, $p = .384$, FDR-corrected) in the baseline session where contextual rhythms were absent, and the power at 2.5 Hz was significantly weaker than that in the context session ($t(15) = 3.421$, $p = .002$, FDR-corrected).

The significant enhancement of EEG power at 2.5 Hz clearly demonstrates that the brain can entrain to the higher-order structure defined by changes in an auditory feature (i.e. pitch) of the contextual stream. Consistent with previous studies, we also observed a wide range of individual variation in such cortical tracking of contextual rhythms (Grahn & McAuley, 2009; Kranczioch, 2017; Nozaradan et al., 2016). Could such variation predict one's ability to extract and utilize the feature-based structure at the neural level, and thus explain the individual differences in the attentional modulation effect? To explore this possibility, we calculated the Pearson correlation between the magnitude of the neural entrainment effect and the behavioral modulation index (BMI) using a cluster-based permutation test. In the context session, we identified two significant clusters showing positive correlation between power at 2.5 Hz and individuals' behavioral effect—one in the parieto-occipital region (Fig. 4B; P5, PO7, PO5, PO3; $r = .587$, $p = .008$, right-tailed) and the other in the frontal area (F3, F1, FZ, FC3, FC1, FCZ, C1, CZ; $r = .681$, $p = .002$). By contrast, no significant clusters were found in the baseline session ($p > .05$).

To further examine the role of brain activity phase-locked with the rhythmic context, we also analyzed the inter-trial phase coherence (ITPC) of EEG signals. Consistent with the power spectrum, ITPC in the context session peaked at 2.5 and 10 Hz (Fig. 4C), suggesting a hierarchical entrainment effect elicited by both feature-based and time-based regularities. By contrast, ITPC in the baseline session only peaked at 10 Hz, mirroring the stimulation rate of the visual stream, and the ITPC at 2.5 Hz was significantly weaker than that in the context session ($t(15) = 4.652$, $p < .001$, FDR corrected). Critically, only in the context session, the 2.5 Hz ITPC was positively correlated with the behavioral modulation index, yielding two significant clusters in the parieto-occipital area (Fig. 4D; P7, P5, PO7, PO5, PO3, O1: $r = .612$, $p = .006$) and the frontal area (FPZ, FP2, AF4, F2, F4, F6; $r = .672$, $p = .002$). Taken together, the results of power and ITPC jointly demonstrate that the better one's brain oscillations entrain to the higher-order temporal structure of

289 the contextual rhythms, the larger attentional enhancement one may exhibit in the
290 between-cycle condition over the within-cycle condition.

291 *T2-related alpha power reflects the attentional modulation*

292 Alpha oscillations have been considered to play a crucial and even causal role in temporal
293 attention, particularly in the AB effect (Hanslmayr et al., 2011; Klimesch, 2012). As the AB
294 phenomenon is characteristic of its stimulation frequency approximately at 10 Hz within
295 the alpha band, the brain can be in a resonant state with the AB stream at the same
296 frequency. It has been demonstrated that an increase in alpha power at the stimulus
297 frequency indicated attentional orienting to the stimulus stream, providing an on-line
298 measure of attentional allocation over the RSVP stream (Müller & Hübner, 2002). On the
299 other hand, enhanced alpha power in the AB task has also been shown to be associated
300 with correct T2 detection (Janson et al., 2014; Keil et al., 2006). Motivated by these
301 findings, we investigated whether alpha activity related to T2 processing could reflect the
302 attentional modulation in our study. We calculated alpha power around stimulation
303 frequency (9.5–10.5 Hz) within the time window of 0–100 ms after T2 onset, and found
304 two significant clusters for the context session—one in the left parieto-occipital region
305 (Fig. 5A; T7, C5, C3, TP7, CP5, CP3, P5, P3, PO5, PO3, O1) and the other in a right-lateralized
306 region (AF4, F2, F4, FC4, FC6, FT8, C4, C6, T8, CP4, CP6, TP8, P8), both showing stronger
307 alpha power in the between-cycle condition than in the within-cycle condition (for the left
308 cluster, $t(15) = 3.570$, $p = .0014$; for the right cluster, $t(15) = 3.631$, $p = .0012$, right-tailed,
309 cluster-based permutation test). This increase in T2-related alpha power, which could be
310 regarded as a sign of stimulus-driven neural activity, agrees well with the observations
311 that more attentional resources are deployed to T2 and thus higher accuracy was achieved
312 in the between-cycle condition than in the within-cycle condition.

313 *Cross-frequency coupling between delta phase and alpha power correlates with the*
314 *attentional modulation effect*

315 Examinations on delta-band entrainment effect and T2-related alpha power both reveal
316 behavioral relevance in our study. This leads to a natural question of whether the observed

attentional modulation effect is implemented through a coordinative process between neural oscillations at delta and alpha bands. To address this question, we analyzed cross-frequency coupling between delta phase and alpha power, which has been found to support the attentional selection between competing stimuli (Gomez-Ramirez et al., 2011; Wilson & Foxe, 2020; Wöstmann et al., 2016). We conducted the analysis in two clusters whose neural responses in both the delta band (the ITPC at 2.5 Hz) and the alpha band (T2-related alpha power) had an established link with the attentional modulation effect: one in the parieto-occipital region (P5, PO3, PO5, O1) and the other in the frontal region (AF4, F2, F4). We calculated the modulation index (MI) of phase-amplitude coupling (PAC) between delta (1.5–3.5 Hz) and alpha band (7–13 Hz) for each cluster. The MI was stronger in the between-cycle condition than in the within-cycle condition, while the effect reached significance only in the parieto-occipital region (Fig. 5B; $t(15) = 2.432$, $p = .028$) but not in the frontal region ($t(15) = 1.459$, $p = .165$). More importantly, this contrast effect of delta-alpha PAC showed a positive correlation with the attentional modulation effect on behavioral performance, which was also restricted to the parieto-occipital region (Fig. 5C; $r = .660$, $p = .005$) and not found in the frontal region ($r = .154$, $p = .569$). To further confirm the association between the delta-alpha PAC and the observed attentional modulation effect, we did a cluster-based permutation test, which again yielded a positively significant cluster in the parieto-occipital region (PO7, PO5, PO3, O1, OZ; $r = .697$, $p = .003$). These results, combined with the findings from single-band analyses, indicate that cortical tracking of hierarchical temporal structures of the auditory context, as well as the coordination of such cortical tracking effects in delta and alpha bands, may play a vital role in reconstructing the deployment of visual attentional in the AB task.

Discussion

Temporal attention guided by time- and feature-based regularities

Dynamic information flows, such as speech and music, are composed of rhythmic structures nested across multiple timescales (Ding et al., 2016; Gross et al., 2013; Koelsch et al., 2013; Peelle & Davis, 2012). These hierarchical structures are organized in time

346 based on regularities in stimulus timing, that is, when sensory signals are emitted (time-
347 based), as well as regularities in information content, that is, how physical or semantic
348 features of the sensory inputs change over time (feature-based). Accrued evidence
349 suggests that temporal structures formed by time-based regularities are effective in
350 directing attention and enhance information selection at the expected time points (Jones
351 et al., 2002; Nobre et al., 2007; Nobre & van Ede, 2018). Yet the current study
352 demonstrates the role of feature-based temporal structures in recomposing temporal
353 attention deployment, which optimizes the distribution of attentional resources over two
354 temporally proximate targets in the AB task.

355 We modified the standard AB paradigm by introducing a contextual stream whose
356 physical property changed periodically to form perceivable, but unattended rhythmic
357 cycles in the background. Although this feature-based temporal structure was task-
358 irrelevant, it modulated the deployment of attentional resources along the AB stream, as
359 indicated by higher T2 detection performance when the two targets were located in
360 different cycles than in the same cycle. More intriguingly, this modulation effect was
361 observed no matter whether the contextual stream was from the auditory (Experiment 1)
362 or the visual (Experiment 2) modality. These findings provide clear evidence that
363 temporal structures defined by periodic changes of physical features in a dynamic context
364 can automatically reconstruct the temporal distribution of visual attention.

365 In the current study, the rhythmic cycles in the contextual stream consisted of a set of
366 temporally grouped items, some with abrupt changes in physical features across the cycle
367 boundaries. Could the attentional modulation effect be achieved purely on the basis of
368 transient perceptual boundaries or temporal grouping? Findings from several control
369 experiments do not agree with these assumptions. In Experiment 3, the rhythmic cycles
370 of contextual rhythms were defined by cyclic motion without any abrupt changes at the
371 boundaries. Even in this case, the cyclic motion yielded a significant attentional
372 modulation effect, excluding the possibility that the observed effect was caused simply by
373 perceptual changes of the background. In addition, results from Experiments 1d and 1e
374 further confirm that temporal attention guided by temporal regularities rather than
375 perceptual grouping is key to the reduced AB effect. On the one hand, simple grouping

without feature-based temporal regularities had little influence on T2 detection (as in Experiment 1d, the feature-based grouping was irregular). On the other hand, when we disrupted time-based regularities by using stochastic stimulus timing, the attentional modulation effect also vanished, even though the rule of feature-based grouping remained in force (as in Experiment 1e, every four identical tones constituted one group). Jointly, these findings point to a mechanism of temporal attentional guidance independent of transient perceptual cues and simple perceptual grouping.

It is worth noting that the attentional modulation effect did not occur in the absence of regular stimulus timing. In other words, the feature-based regularities should work in tandem with the time-based regularities to reconstruct the dynamics of visual temporal attention, at least under the current experimental settings. This finding is consistent with the emerging view concerning the role of a diversity of temporal structures in guiding adaptive behavior (Nobre & van Ede, 2018). It has been suggested by studies using auditory materials, mostly in speech and music perception, that temporal regularities embedded in information content can act along with the time-based anticipation in attentional guidance (Doelling & Poeppel, 2015; Morillon et al., 2016; Peelle & Davis, 2012; Zion Golumbic et al., 2012). Our findings extend these studies by establishing a mechanism in visual temporal attention that is guided by regularities in feature-defined structures on top of the anticipation based on stimulation timing.

The roles of dynamic attentional deployment in reducing attentional blink and boosting awareness

The AB phenomenon represents a bottleneck of conscious awareness pertaining to the temporal resolution of visual attention. It is well known for its robustness that even long repetitive training cannot eliminate the AB effect (Braun, 1998). Some studies have demonstrated attenuated AB magnitude, as manifested in increased T2 detectability, by enhancing T2 salience with color-salience training (Choi et al., 2012), emotional arousal (Keil & Ihssen, 2004), or concurrent sounds (Olivers & Van der Burg, 2008). Another line of research has also reported improved T2 performance when explicitly cueing the target-onset-asynchrony (TOA) on a trial-by-trial basis (Martens & Johnson, 2005) or

manipulating the predictability of target onset (Tang et al., 2014; Visser et al., 2015). Despite implementing different approaches, all these studies tried to manipulate certain aspects of T2, regarding either its salience or predictability in time. By contrast, in our study, the salience of targets and temporal expectations about T2 onset were comparable across all experimental conditions. The only difference between the within- and between-cycle conditions was the positions of the two targets relative to the feature-defined temporal structure. Under this situation, items in the RSVP stream were no longer encoded in isolation, but treated as a part of a structured information flow that could be organized by periodic changes in the context. In particular, when T1 and T2 were separated in different cycles, the temporal relations between them were reframed, which might at least partially reduce the competition between the targets, thus improving the resolution of visual temporal attentional and boosting the conscious access to T2. Instead of emphasizing the role of a given target or a certain time point, our findings highlight the significance of attentional deployment as a dynamic process in regulating visual awareness and the AB effect, which is modulated by temporal structures of the entire information flow.

Neural entrainment to hierarchical contextual rhythms modulates dynamic attending in visual perception

Neural oscillations can be entrained to external rhythms across different frequencies (Calderone et al., 2014; Escoffier et al., 2015; Henry et al., 2014; Mathewson et al., 2012; Schroeder et al., 2010; Schroeder & Lakatos, 2009; Thut & Gross, 2011), allowing the brain to encode dynamic information with multiplexed rhythmic structures across different timescales (Fontolan et al., 2014; Lakatos et al., 2005; O'Connell et al., 2015). A common example of this comes from studies of speech processing. The linguistic structure possesses a temporal hierarchy—from smaller phonetic elements to larger syllabic and phrasal units, which accordingly elicit neural entrainment at multiple frequency bands (Arnal & Giraud, 2012; Zion Golumbic et al., 2012). There is growing evidence that cortical tracking of the higher-order structures plays a vital role in speech and music comprehension (Ding et al., 2016; Doelling & Poeppel, 2015; Gross et al., 2013; Koelsch et

al., 2013; Peelle & Davis, 2012). In our EEG study, we demonstrate an analogous entrainment effect that not only keeps track of the original AB stream at 10 Hz but also represents the higher-order feature-based structure of contextual rhythms at 2.5 Hz. More importantly, the magnitude of the 2.5-Hz entrainment effect is significantly correlated with the strength of the attentional modulation effect. The scalp topographic map of correlation is lateralized and restricted to the left parietal region, which was found to be associated with temporal attention (Bolger et al., 2014; Coull & Nobre, 1998). These findings are in good accordance with the assumption that the cortical tracking of feature-based contextual structure is critical to the redeployment of attentional resources over the AB stream and may lead to the behavioral modulation effect.

The AB paradigm is characterized by its stimulation frequency approximately at 10 Hz within the alpha band. In our experiment, the 10-Hz power after T2 is stronger in the between-cycle condition than in the within-cycle condition, which probably because the increased attentional resources delivered to T2 enhance the stimulus-evoked neural responses in the between-cycle condition (Janson et al., 2014; Keil et al., 2006). Further analysis reveals that, in the left parieto-occipital cluster that exhibits phase-locked neural responses to feature-based structures of the contextual rhythms and a T2-related increment in alpha power, there is a phase-amplitude coupling between the delta and alpha oscillations. Moreover, the strength of this delta-alpha coupling effect predicts the effect of higher-order temporal structures on dynamic attentional allocation at the individual level. These findings corroborate the idea that neural entrainment to a slower external rhythm may serve as a mechanism of attentional selection, with the phase of delta oscillation regulating the excitability of neural activity in the alpha band (Gomez-Ramirez et al., 2011; Wilson & Foxe, 2020; Wöstmann et al., 2016).

Taken together, findings from the current study have cast new light on the classic theory of DAT and its neural implementation. The DAT assumes attention to be inherently oscillatory and can be driven by the timing pattern of external events (Jones, 1976; Jones et al., 1982; Jones & Boltz, 1989; Large & Jones, 1999). By taking advantage of temporal regularities of isochronous or rhythmic events, attentional synchrony can be established and thus improve perceptual accuracy and elevate response speed. Our study extends the

DAT to more general cases of dynamic information processing at both the behavioral and the neural levels. Primarily, our behavioral observations suggest that to utilize regularities in a hierarchical temporal structure, the internal attentional oscillation may not only align with first-order rhythmic structures based on stimulus timing, but also with higher-order rhythmic structures defined by content-based changes of the information flow. Such a dynamic attending process necessitates the synergy between time- and content-based regularities, which could be implemented by neural entrainment to the higher-order temporal structure and its coordination with the cortical tracking of the stimulus rhythm through cross-frequency coupling.

Conclusion

In summary, the current study emphasizes the role of feature-defined contextual rhythms in reconstructing the deployment of visual attention along dynamic information streams. This work enriches our knowledge, as raised at the beginning of this article, about how we optimize the limited mental capacity to process successive inputs from this ever-changing world. Taking the AB phenomenon as an example, we provide a new perspective on visual temporal attention research—when examining the perception of complex dynamic information, temporal context on multiple timescales should be taken into consideration because it provides a meaningful hierarchical temporal frame for attentional deployment. This temporal frame, implemented by neural entrainment, may serve to organize attentional resources in a prospective manner and help construct our conscious experience of the world in the dimension of time.

Materials and Methods

Participants

A total of one hundred and forty-four volunteers (aged from 18 to 30 years, 69 females) were recruited and paid for their participation in the current study. One hundred and twenty-eight participated in the behavioral Experiments 1a-1e, 2a-2b, and 3 (16 for each experiment, with participants' gender balanced), and 16 (5 females) in the EEG experiment. All participants had normal or corrected-to-normal vision and normal hearing and were naïve to the purpose of the experiment. Considering the individual differences in the AB effect, only participants who exhibited a typical AB effect (i.e., an impairment of T2 accuracy at short lags compared with that at long lags) during a pre-screening session were asked to take part in the formal experiments. All participants provided written informed consent in accordance with experimental procedures and protocols approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences.

Stimuli

The rapid visual serial presentation (RSVP) stream used in the AB task consisted of 16 items (except in Experiments 1c and 1d). Among these items, one or two were the targets (capital letters selected from the alphabet, excluding B, D, O, I, M, Q, S, W, and Z), and the remaining were distractors (one-digit numbers, 1 and 0 excluded, without repetitions between any two of four successive digits). The items were displayed for 83 ms each and were separated by 17 ms blank intervals (except Experiment 1e), generating a 10 Hz rhythm based on stimulus presentation (see Fig. 1A, top). Each item subtended 0.47°×0.57° of visual angle and was displayed in white within a gray square (3°×3°) located at the center of a black screen. In each experiment, a contextual stream, which contained the same number of items as the AB stream but was organized by a feature-defined structure, was presented in synchronization with the AB stream. Stimuli were generated and displayed using MATLAB (The MathWorks Inc., Natick, MA) with the Psychophysics toolbox extension (Brainard, 1997). Visual stimuli were presented on a 21-inch CRT monitor with a viewing distance of 55 cm in a dim room. Auditory stimuli were delivered

binaurally over Bose QC3 headphones with the volume set to a comfortable listening level.

Procedures

Behavioral Experiments

In all experiments, participants were explicitly instructed to ignore the contextual events and focused attention on the AB task. Participants initiated each trial by pressing the enter key. A white fixation cross appeared for 600 ms at the center of the screen, followed by the presentation of an AB stream (along with an auditory/visual stream in the context session). After the last item disappeared, the central fixation turned blue to remind the participant to report the identities of the target(s) in the order they detected them by typing on the keyboard.

Experiment 1a had a baseline session followed by a context session. In the baseline session, participants viewed only the AB stream and performed the typical AB task. To induce the AB effect, the second target (T2) in the AB stream was located at the second lag of the first target (T1) with a short stimulus onset asynchrony (SOA) of 200 ms, as the magnitude of AB effect is most robust around the second and the third lags. In contrast with the short-SOA condition, we introduced a long-SOA condition where T2 always appeared at the 8th Lag of T1 and could rarely be missed. To measure the false alarm rate, we also included catch trials in which only one target was displayed. The context session had the same settings and task as the baseline session, except that a task-irrelevant auditory stream was presented in synchronization with the original RSVP stream. Specifically, the auditory stream was composed of 16 tones, each aligned with the onset of a visual item and displayed for 30 ms. The tone sequence changed its pitch from high (2000 Hz) to low (1200 Hz) or vice versa every four items (corresponding to 400 ms), generating 4 auditory cycles (i.e., 4-4-4-4) at a rate of 2.5 Hz (Fig. 1A, middle). To examine the regulation effect of such pitch-defined rhythmic structures, we created two experimental conditions specifically for the short-SOA trials, by varying the positions of T1 and T2 relative to the contextual cycles. In the “between-cycle condition”, T1 and T2 were located in two adjacent cycles; and in the “within-cycle condition”, the two targets were located in the same cycle. To reduce observers’ anticipation about the timing of T1

543 onset across trials, we introduced various T1 positions while keeping T2 located within
544 the middle two cycles. Each session had 120 experimental trials (40 trials for the between-
545 cycle, within-cycle, and long SOA condition each) and 20 catch trials. These trials were
546 divided into four equal blocks, with randomized trial order within each block.

547 Experiment 1b-1e adopted the same procedure as Experiment 1a but with the
548 following exceptions. In Experiment 1b, as shown in Fig. 1A (bottom), we abolished the
549 feature-based structure of the contextual streams by pseudo-randomizing the auditory
550 tone sequences while keeping the pitch of tones at target locations the same as that in
551 Experiment 1a. In Experiment 1c, we changed the temporal structure of the contextual
552 streams by altering their pitch change rate, generating two types of auditory sequences:
553 one with 4 five-tone cycles displayed at 2Hz (i.e., 5-5-5-5, see Fig. 2A, upper), and the other
554 with 5 three-tone cycles at 3.3 Hz (i.e., 3-3-3-3-3, see Fig. 2A, lower). For both frequency
555 conditions, T2 was located in the next to last or third from last cycles. In Experiment 1d,
556 we varied the length of chunks in the contextual streams, generating auditory sequences
557 with four cycles of different lengths (e.g., 5-2-4-3) but always having 4 tones in the third
558 cycle where the second target appeared (see Fig. 2C, upper). In Experiment 1e, the feature-
559 based structure remained while the rhythm from stimulus timing was removed (see Fig.
560 2C, lower). Specifically, the tone pitch changed every four items just as in Experiment 1a,
561 whereas the stimulus onset asynchrony (SOA) of each visual item was selected randomly
562 from a predetermined uniform distribution (50, 67, 83, 100, 100, 117, 133, 150 ms) to
563 keep the total presentation time identical to that in Experiment 1a. In both Experiment 1d
564 and 1e, T2 was always the second item in the 3rd cycle for the between-cycle condition
565 and the last item in the 3rd cycle for the within-cycle condition.

566 Experiments 2a and 2b had a design similar to that of Experiments 1a and 1b, except
567 that we replaced the auditory context with a visually presented contextual stream that
568 possessed color-defined temporal structure. Specifically, in the context session of
569 Experiment 2a, the color of the background square changed from green to red or vice
570 versa at the same tempo as that for contextual tones in Experiment 1a (Fig. 3A, upper).
571 And in Experiment 2b, the background color changed in arrhythmic patterns (Fig. 3C,
572 upper). Luminance of the two colors was matched for each observer with a chromatic

573 flicker fusion procedure before the experiments.

574 Experiment 3 consisted of an experimental session with a structured context as that
575 in Experiment 2a and a control session with a random context as that in Experiment 2b.
576 In the experimental session, the contextual rhythm was created by cyclic motion patterns
577 in the background (Fig. 3E, upper). Specifically, a blue right-angle (width = 0.38°, side
578 length = 1.5°), initiating from one corner (the upper-left or the upper-right, balanced
579 between blocks) of the background square, rotated clockwise at the same pace as the AB
580 stream. In this way, one cycle of rotation corresponded to the appearance of four items
581 (i.e., 400 ms), forming a 2.5 Hz structure based on the motion cycles. In the control session,
582 no cyclic motion pattern remained but the right-angle shifted to a random quadrant under
583 the constraint of identical initial quadrant in each 'cycle' (Fig. 3E, lower).

584 Note that in all these experiments, we also labelled the conditions in baseline and
585 control sessions as "within-cycle" or "between-cycle", just to indicate that these conditions
586 shared the same absolute target positions with the corresponding conditions in the
587 context session. This design was adopted to control for any potential influence of the
588 absolute position of a target within the AB stream. Specifically, for each experimental
589 condition (within- or between-cycle), we matched the absolute positions of T1 and T2
590 between the context session and the baseline session without a context (Experiments 1–
591 2), or between the experimental session and the control session with a random context
592 (Experiments 1 & 3).

593 **EEG Experiment**

594 The procedure of the EEG experiment was mostly identical to that of Experiment 1a except
595 for the following modifications. Black items were presented on a gray background and the
596 item size was 0.59°×0.78°. In each trial, the fixation duration was 1000 ms and each item
597 was displayed for 100 ms with no blank interval. After response, there was a 1.2–1.5 s
598 blank interval. Each subject completed 3 baseline blocks followed by 6 experimental
599 blocks with the auditory context. Each block consisted of 40 trials, with 17 short-SOA trials
600 in each of the between- and within-cycle condition, and the remaining 6 as the catch trials,
601 run in a random order.

EEG recording

A SynAmps² Neuroscan amplifier system (Compumedics Ltd, Abbotsford, Australia) was used for data acquisition. EEG signals were recorded continuously from 64 Ag/AgCl electrodes mounted on an elastic cap according to the extended 10–20 system, with a reference electrode placed between Cz and CPz. Vertical and horizontal eye movements were monitored with two bipolar EOG electrode pairs positioned above and below the left eye and on the outer canthus of each eye. Data were acquired at a sampling rate of 1000 Hz with an online 0.05-100 Hz band-pass filter (notched at 50 Hz). Electrode impedances were kept below 8 kΩ for all electrodes.

EEG data analysis

Preprocessing

Data preprocessing and analysis was performed using EEGLAB toolbox (Delorme & Makeig, 2004) and FieldTrip (Oostenveld et al., 2011) in combination with custom MATLAB scripts. EEG recordings were down-sampled offline to 500 Hz, high-pass filtered at 0.3 Hz, and then segmented into 2200 ms trials from -600 to 1600 ms relative to the onset of the AB stream. Ocular artifacts were then identified and removed using the ADJUST algorithm (Mognon et al., 2011) based on independent component analysis (ICA). Segments with voltage deflections greater than 75 uV were rejected. Residual artifacts were checked by visual inspection. On average, 90 trials remained for each condition and each individual. The segmented data were re-referenced to the average potential of all electrodes excluding the mastoid and EOG electrodes.

Power analysis

The preprocessed EEG signals were first corrected by subtracting the average activity of the entire stream for each epoch, and then averaged across trials for each condition, each participant, and each electrode. Then signals from stream onset were zero-padded and fast Fourier transformed, yielding amplitude and phase estimation at a frequency resolution of 0.5 Hz. Power spectra was calculated as the squared amplitude and then converted to decibel scale (i.e., $10*\log_{10}$). To remove unrelated background noises from

the frequency response of stimulus rhythms, for each frequency, the mean power at two nearest neighboring frequencies was subtracted from the power at that center frequency. The subtracted power at each frequency was then averaged across all channels (excluding M1, M2, VEO, HEO, CB1 and CB2) and compared with zero using one-sample *t* test to determine whether neural oscillations were entrained to temporal structures of the stimulus rhythms. Multiple comparisons across frequencies were controlled by the false discovery rate (FDR, $p < .05$) procedure.

Phase locking analysis

Inter-trial phase coherence (ITPC) serves to indicate the consistency with which intrinsic neural oscillations were phase-locked to the external rhythms over trials. We first obtained phase estimation from spectral decomposition for each single trial based on fast Fourier transform, and then calculated ITPC as follows:

$$\mathrm{ITPC}(f) = \left|\frac{1}{n}\sum_{k=1}^{n}\left(\frac{F_k(f)}{|F_k(f)|}\right)\right| \qquad (1)$$

where, for *n* trials, $F_k(f)$ is the spectral estimate of trial k at frequency *f*, and || represents the complex norm.

T2-related alpha power

In order to measure the neural activity time-locked to T2 at alpha band, time-frequency analysis was performed by convolving single-trial data with a complex Morlet tapered wavelet using the *newtimef* function of EEGLAB. To optimize the trade-off between temporal and frequency resolution, the length of wavelets increased linearly from 1 cycle at the lowest frequency (2 Hz) to 7.5 cycles at the highest frequency (30 Hz, in increments of 0.5 Hz), resulting in power estimates from -321 to 1321 ms around stream onset. For each frequency, power at each time point was first averaged across trials and then divided by the average activity in baseline period from -300 to -200 ms and log-transformed to decibels.

Delta-alpha phase-amplitude coupling analysis

The modulation index (MI) of phase-amplitude coupling (PAC) was used to measure the coordinative modulation between the phase of ongoing oscillations in delta band (1.5–3.5 Hz) and the power in alpha bands (7–13 Hz) at each electrode. First, the low-frequency phase at delta band (f_p) and high-frequency amplitude at alpha band (f_a) were estimated by filtering each epoch with a Butterworth bandpass filter and then applying the Hilbert transform. The broad bandwidth of alpha band (7–13 Hz) was determined to be wide enough to contain the side-bands of the modulating frequency at f_p (2.5 Hz) (Dvorak & Fenton, 2014; Seymour et al., 2017). Next, the modulation index of PAC was quantified using the mean-vector length method first introduced by Canolty et al. (Canolty et al., 2006). As shown in formula (2), for each epoch, the MI values were calculated by combining low-frequency phase and high-frequency amplitude into complex time series and then taking the length of the average vector within the selected time window (500–1300 ms relative to stream onset), which corresponded to the middle two cycles of the contextual stream. The first and last 400 ms of the stream was discarded to avoid the edge artifacts after bandpass filtering. The resulting MI values were then averaged across trials for each condition.

$$\mathrm{MI} = \left| \frac{1}{N} \sum_{n=1}^{N} A_H(n) e^{i(\Phi_L(n))} \right| \qquad (2)$$

where MI is estimated for a single trial with length of N samples or time points, $A_H(n)$ is the amplitude of higher-frequency at time point n, $\Phi_L(n)$ is the phase of lower-frequency at time point n, and || represents the complex norm.

Correlation analysis

To examine whether the above EEG indices were associated with the observed attentional modulation effect, we correlated these EEG indices with individual's behavioral modulation index (BMI), which was determined by the following formula:

$$\mathrm{BMI} = \frac{P_{BET} - P_{WIT}}{P_{BET} + P_{WIT}} \qquad (3)$$

where P_{BET} and P_{WIT} were the accuracy rate of T2 identification in the between-cycle and the within-cycle conditions in the context session, respectively.

Cluster-based permutation test

To identify clusters of channels that are significant in each statistical test, we used the cluster-based permutation test, which was first stated by Maris and Oostenveld (Maris & Oostenveld, 2007) and used in a number of previous studies (Doelling & Poeppel, 2015; Spaak et al., 2014). Firstly, cluster-level statistics are calculated as the sum of channel-specific test statistics within every cluster. Then, the maximum of the cluster-level statistics is taken as the actual test statistic. Finally, the significance probability of the maximum cluster-level statistic is evaluated under the permutation distribution obtained with the Monte Carlo method in which the permutation cluster-level statistic is calculated by randomly swapping the conditions in participants 1000 times.

Acknowledgements

This research was supported by grants from the National Natural Science Foundation of China (31830037 and 31771211), the Strategic Priority Research Program (XDB32010300) and the Youth Innovation Promotion Association (2018116) of the Chinese Academy of Sciences, the National Key Research and Development Project (2020AAA0105600), and the Fundamental Research Funds for the Central Universities.

Competing interests

The authors declare that no competing interests exist.

References

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Bolger, D., Coull, J. T., & Schön, D. (2014). Metrical Rhythm Implicitly Orients Attention in Time as Indexed by Improved Target Detection and Left Inferior Parietal Activation. *Journal of Cognitive Neuroscience*, *26*(3), 593–605. https://doi.org/10.1162/jocn_a_00511

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Braun, J. (1998). Vision and attention: The role of training. *Nature*, *393*(6684), 424–425. https://doi.org/10.1038/30875

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

Brochard, R., Tassin, M., & Zagar, D. (2013). Got rhythm… for better and for worse. Cross-modal effects of auditory rhythm on visual word recognition. *Cognition*, *127*(2), 214–219. https://doi.org/10.1016/j.cognition.2013.01.007

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*(6), 300–309. https://doi.org/10.1016/j.tics.2014.02.005

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2006). High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science*, *313*(5793), 1626–1628. https://doi.org/10.1126/science.1128115

Choi, H., Chang, L.-H., Shibata, K., Sasaki, Y., & Watanabe, T. (2012). Resetting capacity limitations revealed by long-lasting elimination of attentional blink through training. *Proceedings of the National Academy of Sciences*, *109*(30), 12242–12247. https://doi.org/10.1073/pnas.1203972109

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109–127. https://doi.org/10.1037/0096-1523.21.1.109

Coull, J. T., & Nobre, A. C. (1998). Where and When to Pay Attention: The Neural Systems for Directing Attention to Spatial Locations and to Time Intervals as Revealed by Both PET and fMRI. *The Journal of Neuroscience*, *18*(18), 7426–7435. https://doi.org/10.1523/JNEUROSCI.18-18-07426.1998

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of

hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, 201508431. https://doi.org/10.1073/pnas.1508431112

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, *71*(8), 1683–1700. https://doi.org/10.3758/APP.71.8.1683

Dvorak, D., & Fenton, A. A. (2014). Toward a proper estimation of phase–amplitude coupling in neural oscillations. *Journal of Neuroscience Methods*, *225*, 42–56. https://doi.org/10.1016/j.jneumeth.2014.01.002

Escoffier, N., Herrmann, C. S., & Schirmer, A. (2015). Auditory rhythms entrain visual processes in the human brain: Evidence from evoked oscillations and event-related potentials. *NeuroImage*, *111*, 267–276. https://doi.org/10.1016/j.neuroimage.2015.02.024

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nature Communications*, *5*, 4694. https://doi.org/10.1038/ncomms5694

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., & Foxe, J. J. (2011). Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrocorticographic Investigation. *The Journal of Neuroscience*, *31*(50), 18556–18567. https://doi.org/10.1523/JNEUROSCI.2164-11.2011

Grahn, J. A., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage*, *47*(4), 1894–1903. https://doi.org/10.1016/j.neuroimage.2009.04.039

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biol*, *11*(12). https://doi.org/10.1371/journal.pbio.1001752

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, *86*, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Hanslmayr, S., Gross, J., Klimesch, W., & Shapiro, K. L. (2011). The role of alpha oscillations in temporal attention. *Brain Research Reviews*, *67*(1–2), 331–343. https://doi.org/10.1016/j.brainresrev.2011.04.002

Henry, M. J., & Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Perception*, *2*, 62–86. https://doi.org/10.1163/22134468-00002011

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*, *111*(41), 14935–14940. https://doi.org/10.1073/pnas.1408741111

Janson, J., De Vos, M., Thorne, J. D., & Kranczioch, C. (2014). Endogenous and Rapid Serial Visual Presentation-induced Alpha Band Oscillations in the Attentional Blink.

Journal of Cognitive Neuroscience, *26*(7), 1454–1468. https://doi.org/10.1162/jocn_a_00551

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355. https://doi.org/10.1037/0033-295X.83.5.323

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491. https://doi.org/10.1037/0033-295X.96.3.459

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, *32*(3), 211–218. https://doi.org/10.3758/BF03206225

Jones, M. R., Kidd, G., & Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1059–1073. https://doi.org/10.1037/0096-1523.7.5.1059

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. *Psychological Science*, *13*(4), 313–319. https://doi.org/10.1111/1467-9280.00458

Keil, A., & Ihssen, N. (2004). Identification Facilitation for Emotionally Arousing Verbs During the Attentional Blink. *Emotion*, *4*(1), 23–35. https://doi.org/10.1037/1528-3542.4.1.23

Keil, A., Ihssen, N., & Heim, S. (2006). Early cortical facilitation for emotionally arousing targets during the attentional blink. *BMC Biology*, *4*(1), 23. https://doi.org/10.1186/1741-7007-4-23

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*, *110*(38), 15443–15448. https://doi.org/10.1073/pnas.1300272110

Kranczioch, C. (2017). Individual differences in dual-target RSVP task performance relate to entrainment but not to individual alpha frequency. *PLOS ONE*, *12*(6), e0178934. https://doi.org/10.1371/journal.pone.0178934

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The Spectrotemporal Filter Mechanism of Auditory Selective Attention. *Neuron*, *77*(4), 750–761. https://doi.org/10.1016/j.neuron.2012.11.034

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology*, *94*(3), 1904–1911. https://doi.org/10.1152/jn.00263.2005

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119. https://doi.org/10.1037/0033-295X.106.1.119

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martens, S., & Johnson, A. (2005). Timing attention: Cuing target onset interval attenuates the attentional blink. *Memory & Cognition*, *33*(2), 234–240. https://doi.org/10.3758/BF03195312

Martens, S., & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews*, *34*(6), 947–957. https://doi.org/10.1016/j.neubiorev.2009.12.005

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition*, *115*(1), 186–191. https://doi.org/10.1016/j.cognition.2009.11.010

Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., & Gratton, G. (2012). Making Waves in the Stream of Consciousness: Entraining Oscillations in EEG Alpha and Fluctuations in Visual Awareness with Rhythmic Visual Stimulation. *Journal of Cognitive Neuroscience*, *24*(12), 2321–2333. https://doi.org/10.1162/jocn_a_00288

Miller, J. E., Carlson, L. A., & McAuley, J. D. (2013). When What You Hear Influences When You See Listening to an Auditory Rhythm Influences the Temporal Allocation of Visual Attention. *Psychological Science*, *24*(1), 11–18. https://doi.org/10.1177/0956797612446707

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal Prediction in lieu of Periodic Stimulation. *Journal of Neuroscience*, *36*(8), 2342–2347. https://doi.org/10.1523/JNEUROSCI.0836-15.2016

Müller, M. M., & Hübner, R. (2002). Can the Spotlight of Attention Be Shaped Like a Doughnut? Evidence From Steady-State Visual Evoked Potentials. *Psychological Science*, *13*(2), 119–124. https://doi.org/10.1111/1467-9280.00422

Nobre, A. C., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470. https://doi.org/10.1016/j.conb.2007.07.006

Nobre, A. C., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34–48. https://doi.org/10.1038/nrn.2017.141

Nozaradan, S., Peretz, I., & Keller, P. E. (2016). Individual Differences in Rhythmic Cortical Entrainment Correlate with Predictive Behavior in Sensorimotor Synchronization. *Scientific Reports*, *6*, 20612. https://doi.org/10.1038/srep20612

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

O'Connell, M. N., Barczak, A., Ross, D., McGinnis, T., Schroeder, C. E., & Lakatos, P. (2015). Multi-Scale Entrainment of Coupled Neuronal Oscillations in Primary Auditory Cortex. *Frontiers in Human Neuroscience*, 655.

https://doi.org/10.3389/fnhum.2015.00655

Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, *1242*, 191–199. https://doi.org/10.1016/j.brainres.2008.01.070

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. https://doi.org/10.1155/2011/156869

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–176. https://doi.org/10.1016/j.conb.2010.02.010

Seymour, R. A., Rippon, G., & Kessler, K. (2017). The Detection of Phase Amplitude Coupling during Sensory Processing. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00487

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. https://doi.org/10.1016/S1364-6613(97)01094-2

Spaak, E., Lange, F. P. de, & Jensen, O. (2014). Local Entrainment of Alpha Oscillations by Visual Stimuli Causes Cyclic Modulation of Perception. *The Journal of Neuroscience*, *34*(10), 3536–3544. https://doi.org/10.1523/JNEUROSCI.4385-13.2014

Tang, M. F., Badcock, D. R., & Visser, T. A. W. (2014). Training and the attentional blink: Limits overcome or expectations raised? *Psychonomic Bulletin & Review*, *21*(2), 406–411. https://doi.org/10.3758/s13423-013-0491-3

ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia*, *63*, 43–50. https://doi.org/10.1016/j.neuropsychologia.2014.08.008

Thut, G., & Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Frontiers in Perception Science*, *2*, 170. https://doi.org/10.3389/fpsyg.2011.00170

Visser, T. A. W., Ohan, J. L., & Enns, J. T. (2015). Temporal cues derived from statistical patterns can overcome resource limitations in the attentional blink. *Attention, Perception, & Psychophysics*, *77*(5), 1585–1595. https://doi.org/10.3758/s13414-

015-0880-y

Wilson, T. J., & Foxe, J. J. (2020). Cross-frequency coupling of alpha oscillatory power to the entrainment rhythm of a spatially attended input stream. *Cognitive Neuroscience*, *11*(1–2), 71–91. https://doi.org/10.1080/17588928.2019.1627303

Wöstmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, *113*(14), 3873–3878. https://doi.org/10.1073/pnas.1523357113

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, *122*(3), 151–161. https://doi.org/10.1016/j.bandl.2011.12.010

Figures

Figure 1

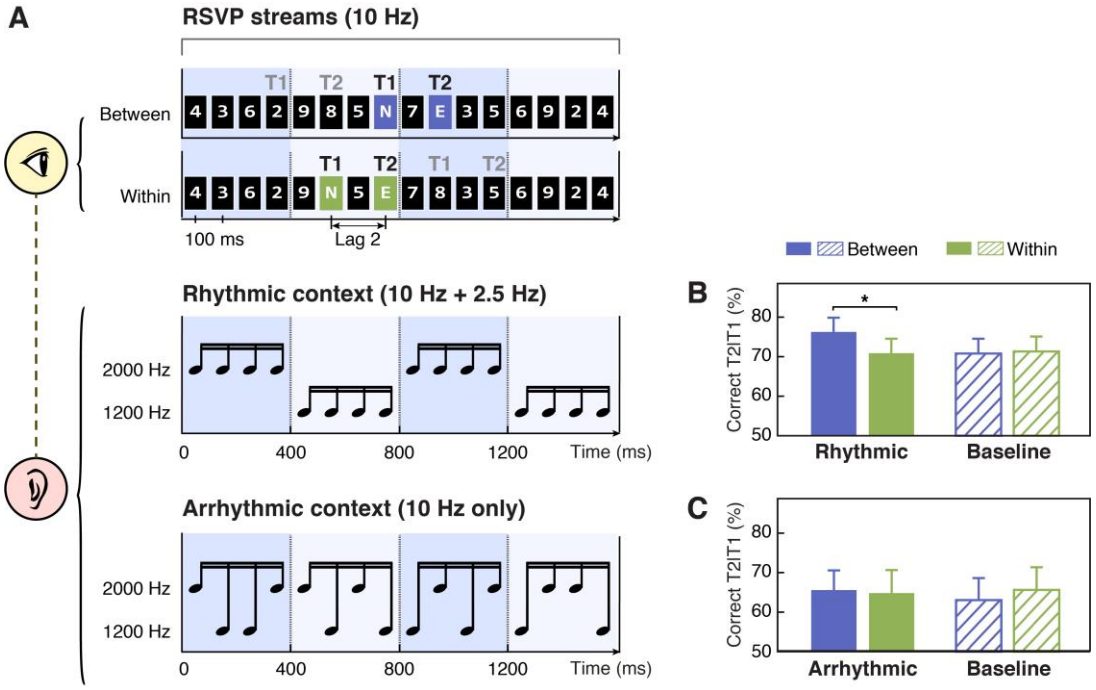
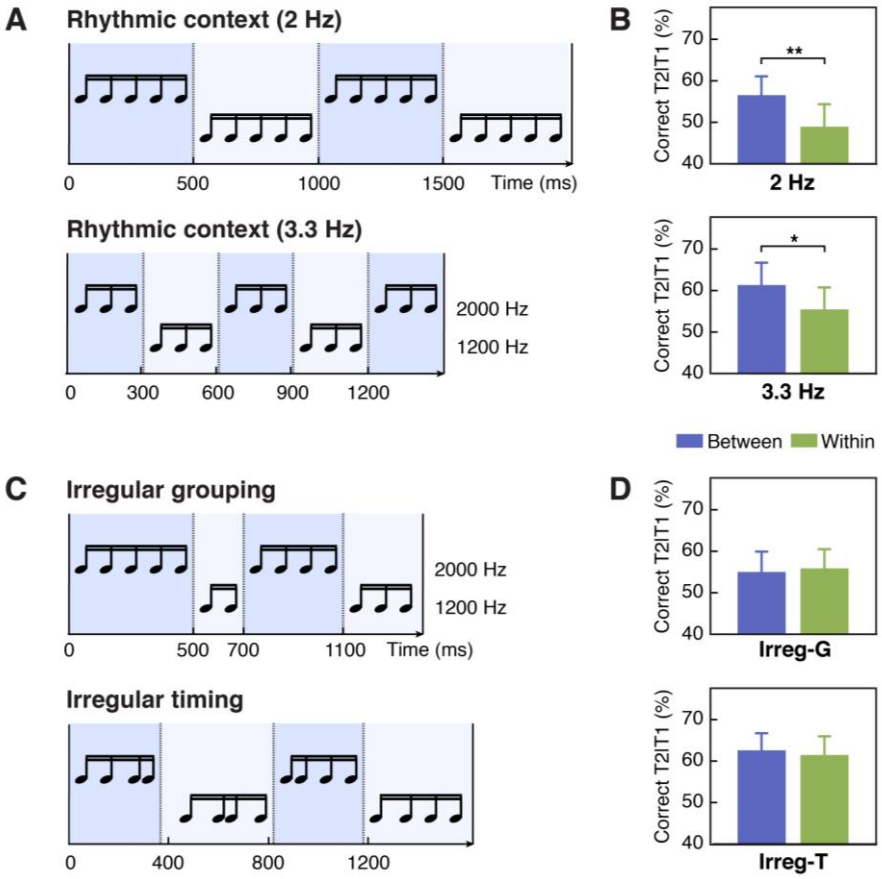


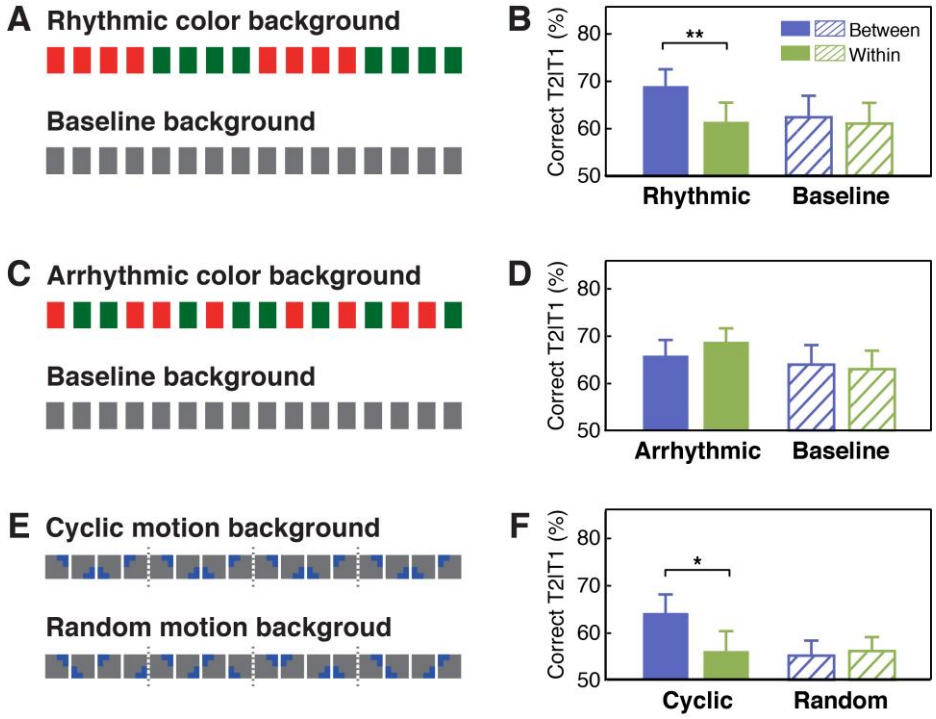
Fig. 1. Schematics of stimuli and results for Experiments 1a and 1b. (A) In the AB task, participants were presented with rapid serial visual presentation (RSVP) streams at 10 Hz (top). Each stream contained two capital letter targets embedded in fourteen number distractors. Black and gray "T1" and "T2" denote two alternative options for target locations in the short-SOA conditions. These targets were located either in two adjacent cycles (the between-cycle condition, displayed on violet background for illustration only) or within the same rhythmic cycle (the within-cycle condition, displayed on green background for illustration only) defined by a rhythmic auditory context in Experiment 1a (middle). Arrhythmic context was used as a control in Experiment 1b (bottom). (B & C) T2 detection accuracy conditioned on correct T1 response for the experiments using rhythmic and arrhythmic contexts. Note that in the baseline (visual-only) session, the labels of "between" and "within" were used to refer to the conditions where the two targets shared the same absolute positions with their corresponding conditions in the context (audiovisual) session. Error bars represent 1 SEM; * p<0.05.

950 **Figure 2**



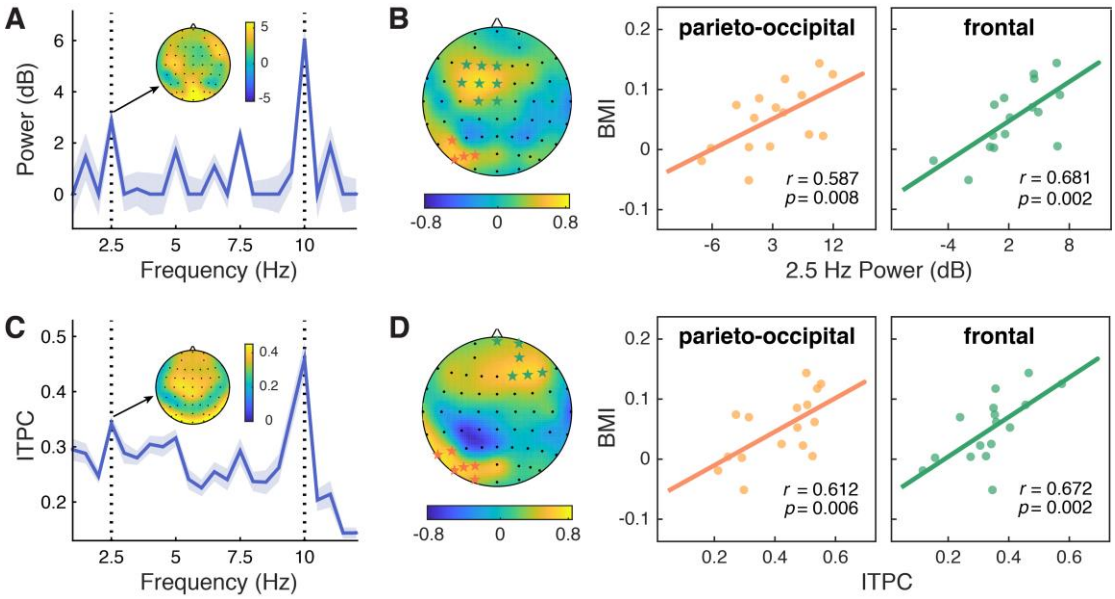
951 **Fig. 2.** Stimuli and results for Experiments 1c, 1d & 1e. (A) Contextual tone sequences with
952 pitch changed every 5 tones (2 Hz, upper) and every 3 tones (3.3 Hz, lower) in Experiment
953 1c. (B) T2 performance in short-SOA conditions for 2-Hz(upper) and 3.3-Hz (lower)
954 sequence in Experiment 1c. (C) The auditory context was grouped irregularly into four
955 chunks with different numbers of tones (G-irregular) in Experiment 1d (upper) and into
956 four regular chunks (four tones in each) but with irregular onset timing (T-irregular) in
957 Experiment 1e (lower). (D) T2 performance in Experiment 1d (upper) and 1e (lower).
958 Error bars represent 1 SEM; * p<0.05, ** p<0.01.

959 **Figure 3**



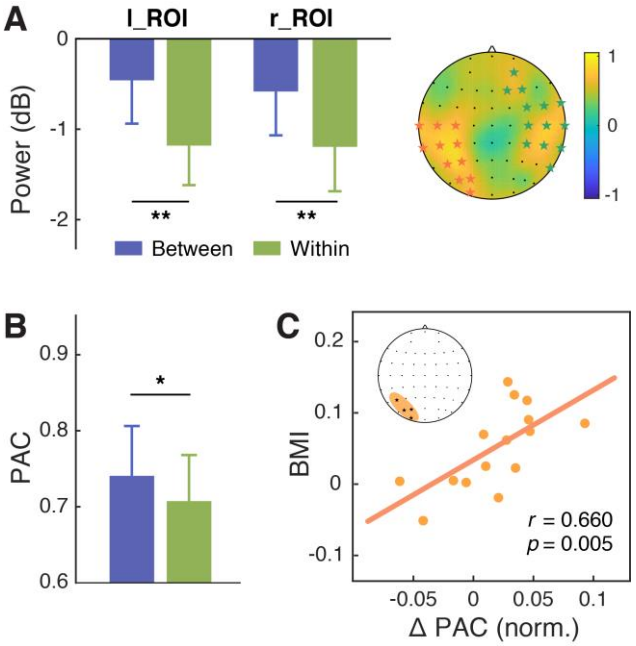
960 **Fig. 3.** Stimuli and results for Experiments 2 and 3 using the visual contexts. (A) The visual
961 context with or without periodic changes in the background color and (B) the T2
962 performance in Experiment 2a. (C) The visual context with or without the background
963 color changed irregularly and (D) the T2 performance in Experiment 2b. (E) Contextual
964 rhythms defined by cyclic/random motion at a constant speed and (F) the T2 performance
965 in Experiment 3. Error bars represent 1 SEM; * p<0.05, ** p<0.01.
966

967 **Figure 4**



968 **Fig. 4.** Neural entrainment to contextual rhythms and its correlation with the attentional
969 modulation effect. (A) The power spectrum of EEG signals averaged across all epochs and
970 channels. For each frequency, power was normalized by subtracting the mean power of
971 the two nearest neighboring frequencies from the power of the center frequency. Shaded
972 areas indicate standard errors of the mean. (B) The 2.5-Hz power entrainment effect in
973 the parieto-occipital cluster (middle, in orange) and the frontal cluster (right, in green),
974 as in indicated in the scalp topographic map (left), significantly correlated with the
975 behavioral modulation index (BMI). (C & D) Analysis of inter-trial phase coherence (ITPC)
976 results yielded similar patterns to that for power.
977

978 **Figure 5**



979 **Fig. 5.** Modulation effect of the alpha power and its coupling with the delta phase. (A) T2-
980 related alpha power averaged within the time window of 0–100 ms relative to the T2
981 onset was significantly higher in the between-cycle condition than in the within-cycle
982 condition in a left parieto-occipital cluster (starred in orange) and a right-lateralized
983 cluster (starred in green). (B) The modulation index of phase-amplitude coupling (PAC)
984 between the delta and alpha bands was higher for the between-cycle condition than for
985 the within-cycle condition, and (C) the difference in normalized PAC strength could
986 predict the BMI across individuals. Shadowed area in the topographic plot indicates the
987 cluster showing significant behavioral relevance in both delta- and alpha-band activities.
988 Error bars represent 1 SEM; * p<0.05, ** p<0.01.