

Saliency-based Rhythmic Coordination of Perceptual Predictions

Qiaoli Huang and Huan Luo

Abstract

■ Objects, shown explicitly or held in mind internally, compete for limited processing resources. Recent studies have demonstrated that attention samples locations and objects rhythmically. Interestingly, periodic sampling not only operates over objects in the same scene but also occurs for multiple perceptual predictions that are held in attention for incoming inputs. However, how the brain coordinates perceptual predictions that are endowed with different levels of bottom-up saliency information remains unclear. To address the issue, we used a fine-grained behavioral measurement to investigate the temporal dynamics of processing of high- and low-salient visual stimuli, which have equal possibility to occur within

experimental blocks. We demonstrate that perceptual predictions associated with different levels of saliency are organized via a theta-band rhythmic course and are optimally processed in different phases within each theta-band cycle. Meanwhile, when the high- and low-salient stimuli are presented in separate blocks and thus not competing with each other, the periodic behavioral profile is no longer present. In summary, our findings suggest that attention samples and coordinates multiple perceptual predictions through a theta-band rhythm according to their relative saliency. Our results, in combination with previous studies, advocate the rhythmic nature of attentional process. ■

INTRODUCTION

Objects in visual scenes compete for limited processing resources, and selective attention resolves the competition by prioritizing salient or behaviorally relevant objects (Desimone & Duncan, 1995). Recently, a series of studies have revealed the dynamic characteristics of attentional operation, suggesting that multiple locations, features, and objects are indeed sampled in a rhythmic manner (Fiebelkorn, Pinsk, & Kastner, 2018; Helfrich et al., 2018; Ho, Leung, Burr, Alais, & Morrone, 2017; Jia, Liu, Fang, & Luo, 2017; Dugue, Roberts, & Carrasco, 2016; Drewes, Zhu, Wutz, & Melcher, 2015; Huang, Chen, & Luo, 2015; Song, Meng, Chen, Zhou, & Luo, 2014; Fiebelkorn, Saalmann, & Kastner, 2013; Landau & Fries, 2012). These studies support the view that attention involves an oscillation-based temporal mechanism to efficiently represent and monitor multiple objects.

In addition to the traditional top-down task or exogenous cuing manipulations, saliency information, referring to the bottom-up physical distinctiveness of objects compared with the surrounding context (e.g., contrast, color, orientation, shape; Wolfe, O'Neill, & Bennett, 1998; Treisman & Gelade, 1980), provides another efficient strategy for the deployment of attention, indepen-

dent of particular task demands. Saliency information extracted from the visual scene is represented in the form of a “saliency map,” which guides the attentional spotlight to first stay on the most salient location, followed by moving to the next most salient location, resulting in the generation of attentional scan paths (Itti & Koch, 2001; Koch & Ullman, 1985). It is posited that the saliency map might be intrinsically represented via a temporal strategy (Rucci, Ahissar, & Burr, 2018; Jensen, Gips, Bergmann, & Bonnefond, 2014; Jensen, Bonnefond, & VanRullen, 2012), whereby multiple items are sorted in time according to their relative saliency.

Interestingly, this time-based coordination not only operates over objects that are shown explicitly in a visual scene but has also been recently suggested to engage in the organization of perceptual predictions that are kept in internal attention for incoming inputs (Huang et al., 2015). In fact, it is well known that expectation is strongly linked to and even determines perception (de Lange, Heilbron, & Kok, 2018), and therefore the attentional mechanism might generalize to the representation of perceptual predictions. Motivated by these findings, this study investigates the possibility that attention processes perceptual predictions, which are endowed with different saliencies determined solely by their physical distinctiveness, via a temporally dissociated manner.

To address the issue, we employed an orientation discrimination task combined with a time-resolved behavioral measurement approach (Huang et al., 2015; Tomassini, Spinelli, Jacono, Sandini, & Morrone, 2015; Song et al., 2014; Fiebelkorn, Saalmann, et al., 2013;

This article is part of a Special Focus deriving from a symposium at the 2018 annual meeting of Cognitive Neuroscience Society, entitled, “Oscillatory mechanisms underlying human perception and attention.”
Peking University

Landau & Fries, 2012) to examine the fine temporal course of the processing of items that have high or low salient physical properties (i.e., high/low contrast). Crucially, the high- and low-salient stimuli were mixed within the same experimental block and were equally likely to occur, ensuring that the participants had to hold two attentional templates for the incoming inputs (i.e., two perceptual predictions with distinct saliency). Furthermore, we applied a 10-Hz visual entrainment before target appearance so that the brain state could be efficiently reset. This manipulation was motivated by several factors. First, previous work has revealed that a 10-Hz visual entrainment would elicit a robust modulation in the detection performance for subsequent targets (Spaak, de Lange, & Jensen, 2014; Herrmann, 2001). Second, the alpha-band neuronal rhythm, well known to play a central function in attention (Haegens, Luther, & Jensen, 2012; Klimesch, 2012; Händel, Haarmeier, & Jensen, 2011), has been posited to sort unattended items at different phases according to their relative saliency (Jensen et al., 2012, 2014). We compared the time-resolved behavioral performance for the high- and low-salient targets after the 10-Hz entrainment and examined whether they would show distinct temporal profiles.

METHODS

Participants

Fifty-four adults (18–24 years old) participated in the experiments. All participants had normal or corrected-to-normal vision and had no history of psychiatric or neurological disorders. All participants provided written informed consent before the experiments.

Twenty participants took part in Experiment 1. There were eight blocks (112 trials per block) in Experiment 1. We calculated the detection accuracy for each block and for each participant. We excluded the block if its overall accuracy was lower than 55% or higher than 85%. Participants with more than three excluded blocks were not included. This left 17 participants (each with at least 560 trials) in Experiment 1. Nineteen participants took part in Experiment 2. The same criteria were used in Experiment 2, and no participant was excluded in Experiment 2. Fifteen participants took part in Experiment 3. The same criteria were also used in Experiment 3. In addition, because high- and low-contrast stimuli were presented in separate blocks in Experiment 3, we excluded participants who had more than one excluded block for either the high- or low-contrast conditions to ensure balanced trials for the two conditions. This left 13 participants (each with at least 672 trials) in Experiment 3.

Stimuli and Tasks

This experimental paradigm was adapted from Spaak and colleagues (2014). Participants sat in a dark room,

60 cm in front of a CRT monitor (60-Hz refresh rate). Their heads were stabilized using a chin rest. A bright fixation cross was presented at the center of the screen from the beginning of a trial and lasted until the end of the trial. Participants were requested to maintain their fixation on this cross during the whole trial. Five hundred milliseconds after the onset of the cross, two white squares (with sides of 6° [visual angle]) were presented in the left and right visual fields (at 3° eccentricity), respectively. One of the squares flashed periodically at a rate of 10 Hz, accomplished by the square being on for one refresh cycle ($1/60 \text{ sec} \approx 17 \text{ msec}$) and off for five refresh cycles ($5/60 \text{ sec} \approx 83 \text{ msec}$). The other square in the other hemifield also flashed, but with randomly jittered timing between two successive flashes, subject to the constraint that no two flashes could occur in subsequent frames. Both the 10-Hz periodic stimulus train (entrained location) and the temporally jittered stimulus train (nonentrained location) had 26 flashing frames within a trial, and the hemifield in which the two stimulus trains were presented was counterbalanced across trials. After a variable delay (34–238 msec in steps of 34 msec in Experiments 1 and 3, 34–476 msec in steps of 34 msec in Experiment 2; uniform probability across different possible delays), a high- or low-contrast (100% vs. 50%, respectively) Gabor target with near-threshold orientation (duration = 17 msec, diameter = 1°) was presented at either of the two locations previously occupied by the two squares (entrained or nonentrained locations). Participants were asked to determine the orientation of the target (tilted left or right). At the beginning of each block for all the Experiments, 40 additional trials (20 high-contrast trials and 20 low-contrast trials) were presented to calibrate the perceptual threshold for orientation discrimination (tilted left or right) via a 3-down-1-up staircase procedure. The threshold was defined as the orientation at which the discrimination accuracy was 75%. In addition, high- and low-contrast targets were mixed within the same block in Experiments 1 and 2, whereas in Experiment 3, they were presented in separate blocks.

Experiment 1

Experiment 1 had 896 trials, divided into eight blocks (112 trials per block). During each block, the orientation of the Gabor target was fixed at the threshold orientation measured in the pretest. Data in all eight blocks were then pooled in the analysis.

Experiment 2

Experiment 2 had 896 trials, divided into two sessions. The two sessions were conducted on 2 days separated by fewer than 3 days. Each session (~60 min in duration) included four blocks. Experiment 2 followed a procedure similar to that of Experiment 1, except that the delay was

extended from 238 to 476 msec (i.e., in Experiment 1, ranging from 34 to 238 msec in steps of 34 msec; in Experiment 2, ranging from 34 to 476 msec in steps of 34 msec).

Experiment 3

Experiment 3 had 896 trials, divided into eight blocks. Experiment 3 followed a procedure similar to that of Experiment 1, except that the high- and low-contrast targets were presented in separate blocks.

Data Analysis

Behavioral accuracy time courses (hit rate as a function of SOA) were analyzed with MATLAB (The MathWorks) using functions from the Curve Fitting toolbox, EEGLAB toolbox, and CircStat toolbox.

Detrend

The slow trend was estimated using a second polynomial function in each participant separately and was then subtracted from the original time waveform (Huang et al., 2015; Song et al., 2014; Fiebelkorn, Saalman, et al., 2013) to get the detrended time courses.

Sinusoidal Fitting and Permutation Test

The accuracy time course data were fitted using a Fourier series model: $f(t) = A \sin(\omega t + \Phi) + B$, where t is time, ω is frequency, A is amplitude, Φ is the phase of the sinusoidal fit, and B is a constant. To test the statistical significance of the best-fitting frequency, we did a permutation test by shuffling the original hit rates across SOAs and fitted the surrogate data using a sinusoid function with the best-fitting frequency. We repeated this procedure 500 times and obtained a distribution of goodness of fit (R^2), from which the $p \leq .05$ threshold was obtained.

Spectrum Analysis and Permutation Test

We also conducted a spectrum analysis to investigate the spectral characteristics of the accuracy time courses. Specifically, we performed a Fourier transform (FFT) to convert the grand-averaged behavior accuracy time course into the frequency domain (after application of a Hanning window). We did FFT on the grand-averaged time course across participants to investigate the spectral characteristics of the high–low (H–L) difference accuracy time course. To test the statistical significance of the spectrum, we performed a permutation test by shuffling the original hit rates across SOAs and conducted an FFT on surrogate signals. This procedure was repeated 500 times, arriving at a distribution of spectral power for each frequency point, from which the $p \leq .05$ threshold was obtained as a function of frequency. Multiple comparison

correction across frequencies was then applied to the threshold spectrum, by setting the maximum threshold value across all frequency bins as the corrected threshold (Song et al., 2014).

Induced Spectrum Analysis and Permutation Test

To perform the induced spectrum analysis (testing the 10 Hz entrainment effect), instead of using grand average results, we performed FFT on the accuracy time courses for each participant separately, considering that the phase of the elicited 10 Hz fluctuation in behavior might be inconsistent across participants. The power spectrum was then averaged across participants to get the induced spectrum analysis results. To test the statistical significance of the spectrum, we performed a permutation test by shuffling the original hit rates across SOAs and conducted an FFT on surrogate signals. This same induced spectrum analysis was repeated 500 times, arriving at a distribution of spectral power for each frequency point, from which the $p \leq .05$ threshold was obtained as a function of frequency. Multiple comparison correction across frequencies was then applied to the threshold spectrum, by setting the maximum threshold value across all frequency bins as the corrected threshold.

Phase Relationship

To examine the H–L phase relationship, we calculated separately for each participant the phase difference between the high- and low-contrast conditions as a function of frequency from 0 to 15 Hz (sampling frequency of 30 Hz here). Then, we calculated cross-participant coherence in the phase difference values, resulting in an H–L phase difference coherence pattern as a function of frequency. A permutation test was also performed by shuffling the accuracy time series to assess the statistical significance of the H–L phase relationship. After each randomization, we conducted an FFT on surrogate signals; we repeated this procedure 500 times, arriving at a distribution of H–L phase difference coherence patterns for each frequency point, from which the $p \leq .05$ threshold was estimated. Again, multiple comparison correction across frequencies was then applied to the threshold spectrum, by setting the maximum threshold value across all frequency bins as the corrected threshold. Nonuniformity for H–L phase differences at the theta band (3–4 Hz) across participants was then tested using circular statistics (Rayleigh test for nonuniformity for circular data; CircStats toolbox).

RESULTS

Experiment 1

We employed a modified paradigm used in a previous study (Spaak et al., 2014). Specifically, as shown in

Figure 1, after a 500-msec fixation period, participants were presented with 2.5-sec stimulation trains composed of bilateral flashing white squares against a dark gray background. One of the squares flashed periodically at 10 Hz (entrained), and the other flashed according to an arrhythmic sequence with jittered interflash intervals (nonentrained). After the offset of the 2.5-sec stimulation, a Gabor target with high or low contrast appeared in either of the two locations previously occupied by the two squares (entrained or nonentrained locations) after a variable delay (SOA: 34–238 msec in steps of 34 msec). Participants were instructed to judge the orientation of the target (tilted left or right). The orientation of the Gabor stimulus was adjusted in a pretest so that the overall accuracy was approximately 75% in each participant.

Figure 2A plots the overall accuracy performance averaged across all tested SOAs (34–238 msec in steps of 34 msec) for the four conditions (high or low contrast, at the entrained or nonentrained side). As expected, the high-contrast target showed significantly higher hit rates than the low-contrast target (high-contrast target: $73.04 \pm 4.24\%$; low-contrast target: $66.78 \pm 2.99\%$; two-way repeated ANOVA, $df = 16$, main effect of Contrast: $F(1, 16) = 37.93$, $p < .001$). The entrained and nonentrained conditions showed no difference

(entrained: $69.69 \pm 2.58\%$; nonentrained: $70.13 \pm 4.3\%$; two-way repeated ANOVA, $df = 16$, main effect of Entrainment: $F(1, 16) = 0.23$, $p = .64$), indicating that attention was not biased to an entrained versus nonentrained hemifield throughout each trial.

We next examined the behavioral time courses (hit rate as a function of SOA) for the high- and low-contrast targets after the 10-Hz entrainment, using the same analysis procedure from the previous study (Spaak et al., 2014). Specifically, we used the nonentrained time course as a baseline (i.e., as a randomly fluctuating behavioral profile) and subtracted it from the entrained performance, for the high- and low-contrast conditions, respectively. First, as shown in Figure 2B, we did not observe the expected 10-Hz entrainment in the accuracy courses for both the high- and low-contrast targets (but see the 10-Hz activation using induced analysis; Figure 4C). However, the behavioral profiles for the high- and low-contrast targets displayed quite different temporal patterns. Specifically, the high-contrast target (red) showed better performance than the low-contrast target only within short SOAs (<150 msec), peaking at approximately 100 msec (paired t test, $t(16) = 2.285$, $p = .036$), whereas the performance for the low-contrast stimuli (blue) monotonically increased and peaked at approximately 230 msec. Figure 2C plots the H–L difference time

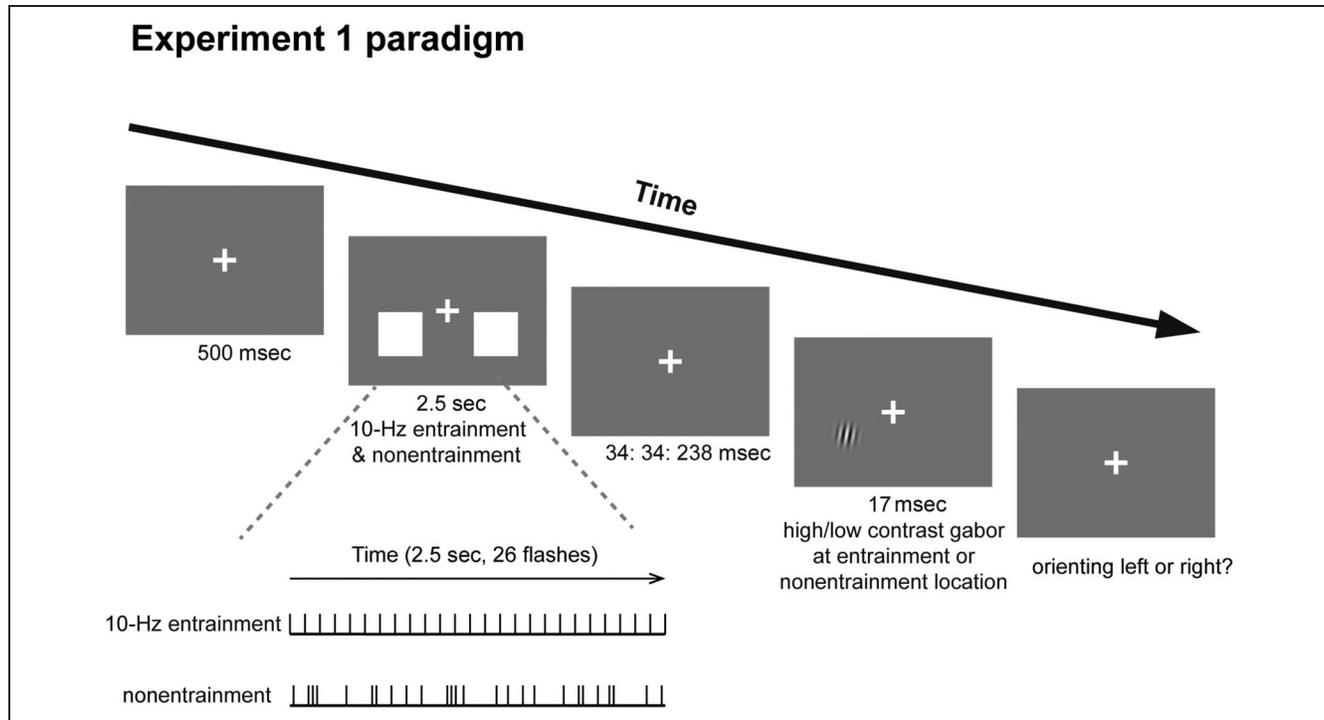


Figure 1. Experiment procedure. Participants were presented with 2.5-sec stimulation trains composed of bilateral flashing white squares against a dark gray background. One of the squares flashed periodically at 10 Hz (entrained), and the other flashed according to an arrhythmic sequence with jittered interflash intervals (nonentrained). After the offset of the 2.5-sec stimulation, a Gabor target with either high or low contrast appeared in either of the two locations previously occupied by the two squares (entrained or nonentrained locations) after a variable delay (SOA: 34–238 msec in steps of 34 msec). Participants were instructed to determine the orientation of the target (tilted left or right). The same experimental design was applied in Experiment 2, but with extended tested SOA (SOA: 34–476 msec in steps of 34 msec). Experiment 3 employed the same design as Experiment 1, except that the high- and low-contrast stimuli were presented in separate blocks.

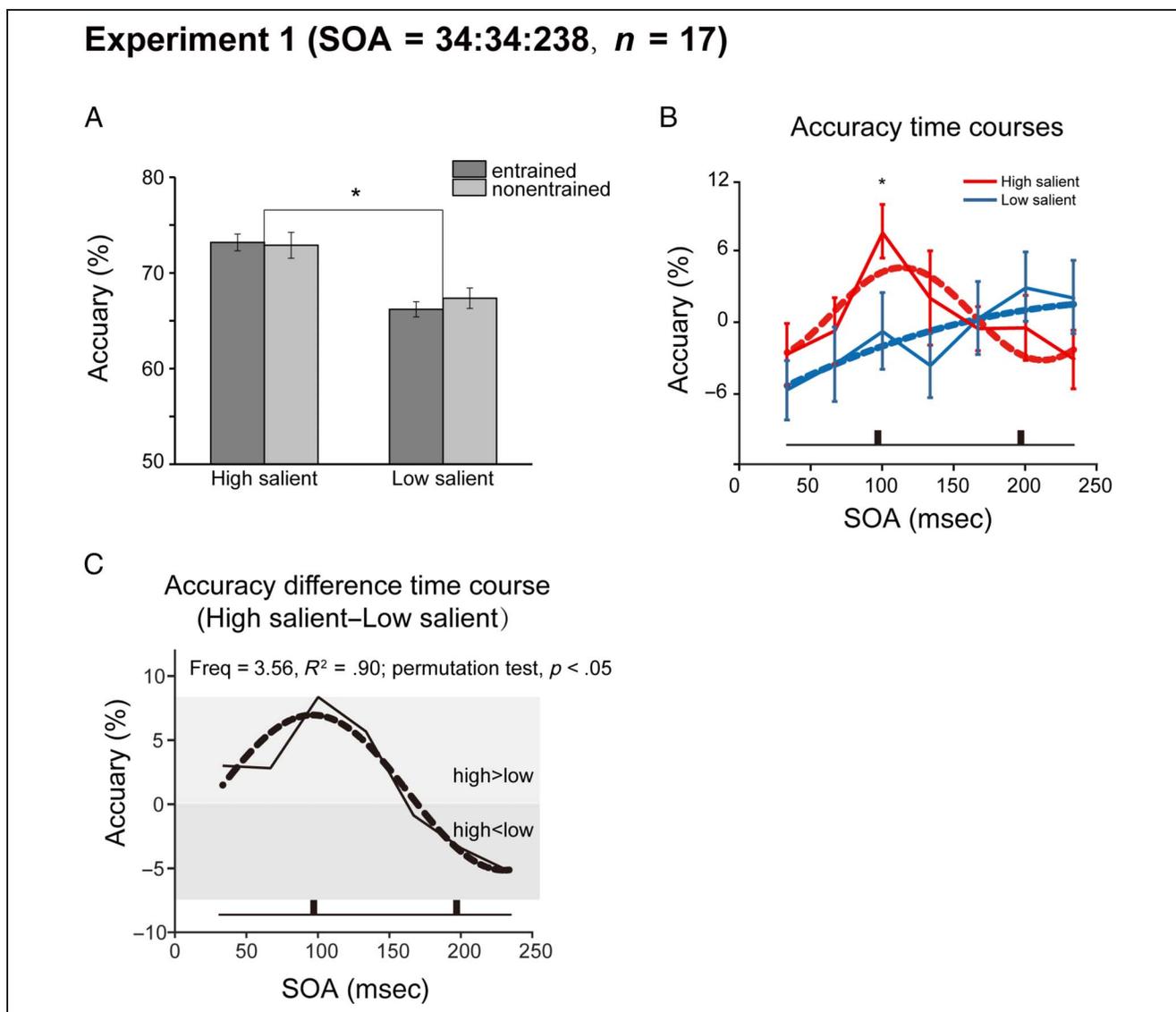


Figure 2. Results for Experiment 1. (A) Grand-averaged detection accuracy across all the tested SOAs (34–238 msec in steps of 34 msec) for the four conditions (high or low salient, at the entrained [dark gray] or nonentrained [light gray] side). Error bars represent ± 1 SEM. (B) Grand-averaged accuracy time courses as a function of SOA, for the corrected (entrained–nonentrained) high-salient (red) and low-salient (blue) conditions, with the best sinusoidal fits in dashed lines. Error bars represent ± 1 SEM. (C) Grand-averaged accuracy time course as a function of SOA, for the difference between the corrected high- and low-salient conditions, with the best sinusoidal fit as a dashed line.

course, which reveals a peak (~ 100 msec) followed by a trough (~ 230 msec), indicating that the high-contrast target was processed approximately 130 msec earlier than the low-contrast target. We further fitted a sinusoidal curve (dashed line) to the H–L difference waveform, and the best-fitting frequency was 3.56 Hz (permutation test by shuffling original hit rates across SOAs and fitting the surrogate data using a 3.56-Hz sinusoid, $p \leq .05$).

Thus, perceptual predictions with different saliencies (i.e., high and low contrast here) seem to be processed at different latencies, such that attention first dwells on the most salient perceptual prediction, followed by examining the next most salient perceptual hypothesis held in mind. Note that, here, the high- and low-contrast stimuli

were not presented simultaneously within the same display, and our results thus could not be accounted for by the saliency map theory (Itti & Koch, 2001; Koch & Ullman, 1985).

Experiment 2

After observing a saliency-based sequential profile in Experiment 1, we next assessed the behavioral courses within an extended time range (0–500 msec; SOAs in Experiment 2: 34–476 msec in steps of 34 msec). Our main interest was to see whether the observed one-cycle pattern (Figure 2C) would continue by starting another cycle (rhythmic sampling hypothesis) or would just decay to baseline.

First, the overall performance (i.e., hit rates averaged across all tested SOAs) in Experiment 2 was similar to that in Experiment 1. The high-contrast target showed higher hit rates than the low-contrast target (high-contrast target: $73.24 \pm 3.93\%$; low-contrast target: $67.17 \pm 3.66\%$; two-way repeated ANOVA, $df = 18$, main effect of Contrast: $F(1, 18) = 70.79, p < .001$). The entrained and nonentrained conditions were not different (entrained: $69.69 \pm 2.58\%$; nonentrained: $70.13 \pm 4.3\%$; two-way repeated ANOVA, $df = 18$, main effect of Entrainment: $F(1, 18) = 0.23, p = .64$), again suggesting that attention was not biased to the entrained versus non-entrained location throughout each trial. We then used the same analysis procedure as that in Experiment 1 to

examine the behavioral time courses for the high- and low-contrast stimuli at SOAs from 0 to 500 msec.

Figure 3A plots the accuracy time courses (entrained–nonentrained) for the high-contrast (red) and low-contrast (blue) conditions. First, both the high- and low-contrast conditions showed a theta-band (4–6 Hz) rhythmic profile (sinusoidal fitting; high-contrast condition: $\text{freq} = 5.7 \text{ Hz}, R^2 = .394$, permutation test, $p = .07$; low-contrast condition: $\text{freq} = 4.54, R^2 = .522$, permutation test, $p < .05$). However, as shown in Figure 3B, the detrended H–L difference waveform (see the detrending details in the Methods section and the slow trend in the Supplementary Figure¹), instead of returning to baseline after the first 250 msec, displayed a

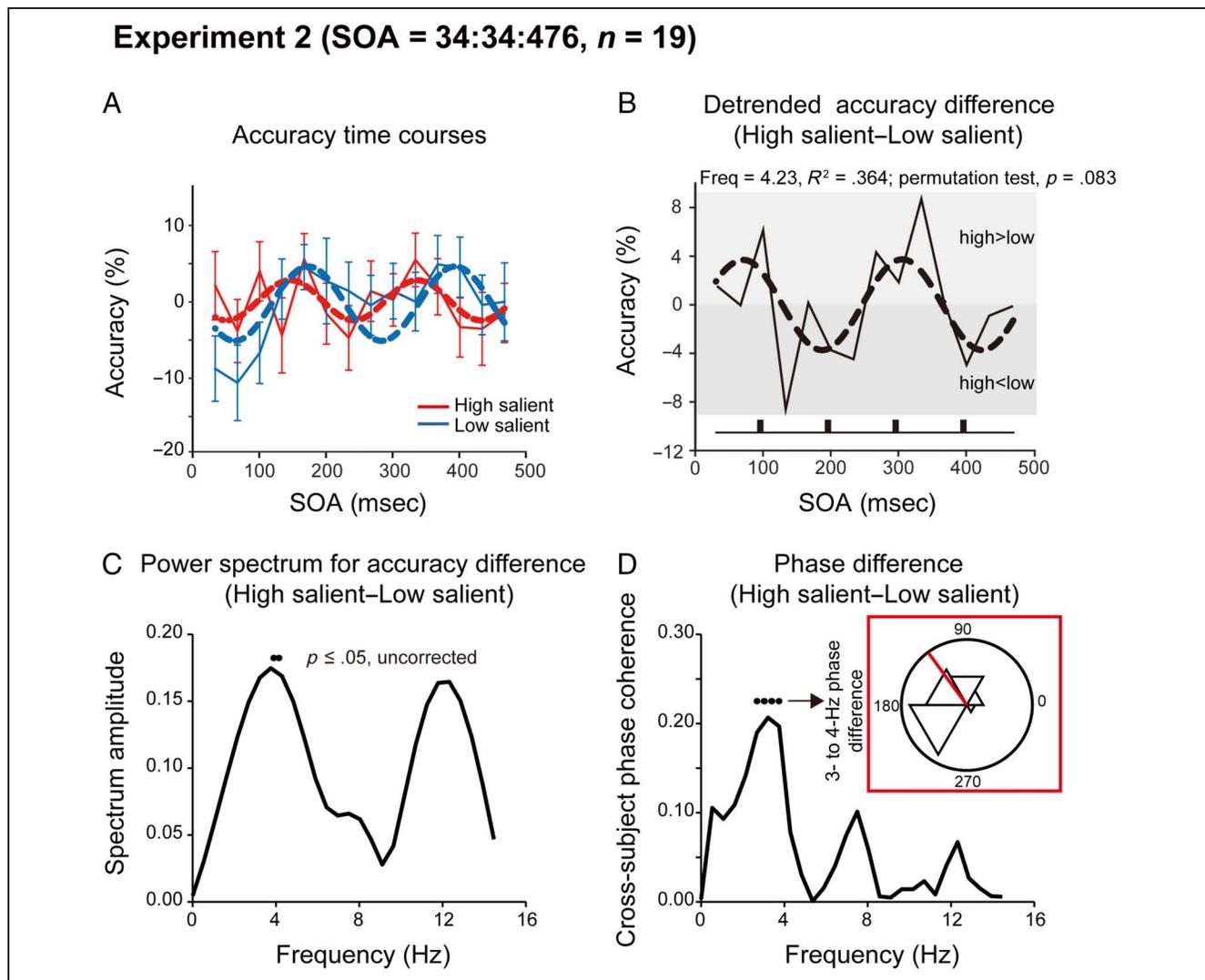


Figure 3. Results for Experiment 2. (A) Grand-averaged accuracy time courses as a function of SOA, for the corrected (entrained–nonentrained) high-salient (red) and low-salient (blue) conditions, with the best sinusoidal fits as dashed lines. Error bars represent $\pm 1 \text{ SEM}$. (B) Detrended grand-averaged accuracy time courses as a function of SOA, for the difference between the corrected high- and low-salient conditions, with the best sinusoidal fit as a dashed line (see the results before detrending in the Supplementary Figure). (C) Power spectrum of results shown in B as a function of frequency. Black dots indicate frequency points showing statistical significance (permutation tests, $p \leq .05$). (D) Intersubject phase coherence in the phase difference between the corrected high- and low-salient conditions as a function of frequency. The red inset figure shows the polar plots for the distribution ($n = 19$) of H–L phase difference in the theta band as well as the mean phase (red line).

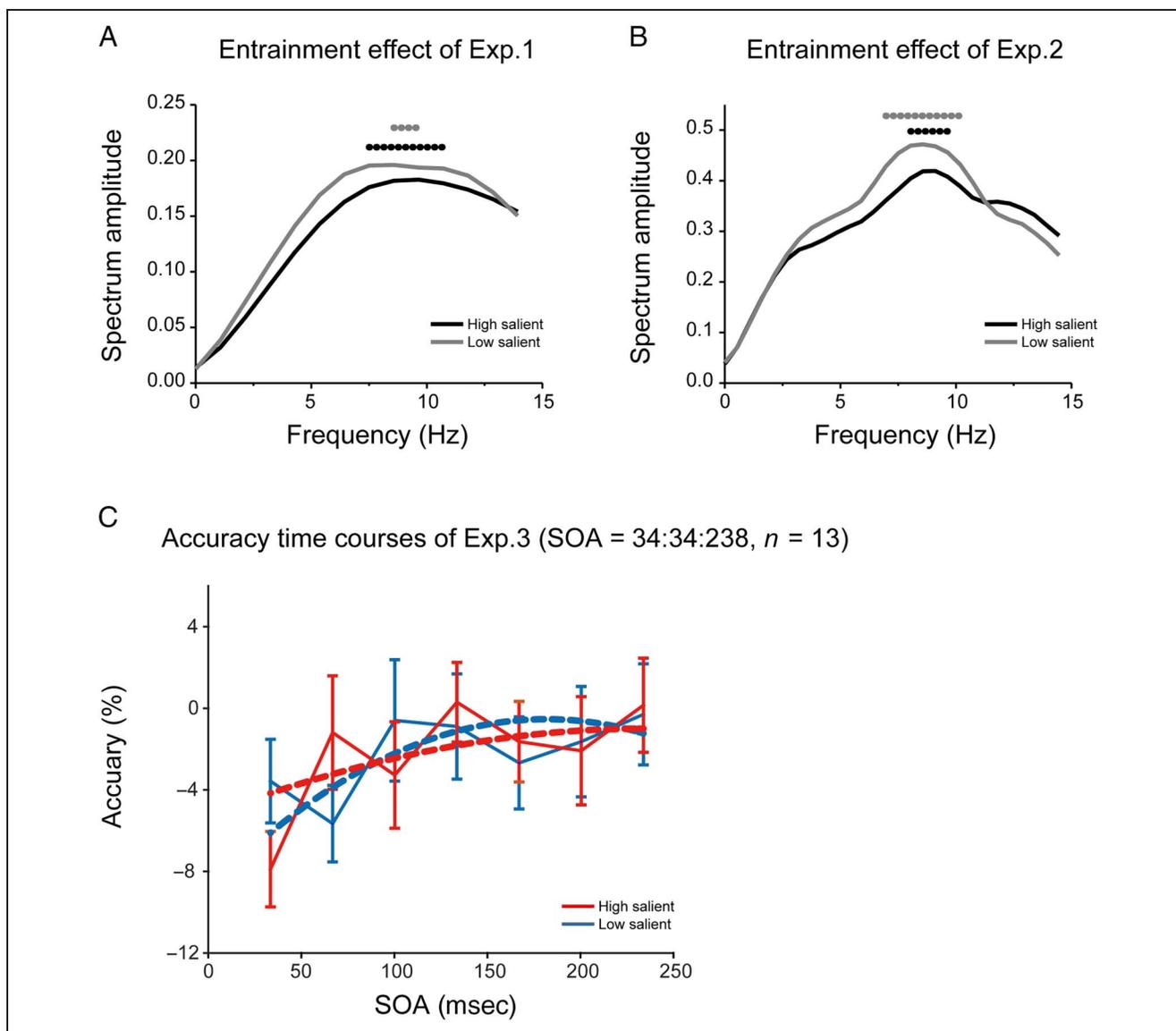


Figure 4. Entrainment effects and Experiment 3 results. Induced spectral analysis (performing spectral analysis for each participant separately and then averaged across participants) was used to examine the 10-Hz entrainment effect in Experiments 1 and 2. (A) Induced analysis results of Experiment 1. Spectrum of the accuracy time courses for the corrected high-salient (black) and low-salient (gray) conditions. Black (high-salient) and gray (low-salient) dots indicate frequency points showing significant activations (permutation tests, $p \leq .05$). (B) Induced analysis results of Experiment 2. (C) Experiment 3 results. Experiment 3 was the same as Experiment 1, except that the high- and low-contrast targets were presented in separate blocks. Grand-averaged accuracy time courses as a function of SOA, for the corrected (entrained–nonentrained) high-salient (red) and low-salient (blue) conditions, with the best sinusoidal fits as a dashed line. Error bars represent ± 1 SEM. Exp. = Experiment.

periodic pattern containing another cycle of peaks and troughs within the second 250-msec range (also see the results before detrend in the Supplementary Figure). The best-fitting sinusoid (dashed line) for the difference waveform was approximately 4.23 Hz (permutation test, $p = .08$), still in the theta-band frequency range as revealed in Experiment 1. Further spectrum analysis (Figure 3C) of the H–L difference accuracy time course also revealed a significant peak of approximately 4 Hz (permutation test, $p \leq .05$), similar to the sinusoidal fitting result, suggesting that a theta-band oscillation mediates high- and low-salient perceptual predictions alternatively.

Finally, we examined the phase relationship between the high- and low-contrast conditions, by calculating their phase difference in each participant respectively and assessing how reliable the H–L phase difference was across participants. As shown in Figure 3D, the intersubject phase coherence was significant at approximately 4 Hz (permutation test, $p \leq .05$), and the phase difference between the high- and low-contrast conditions centered at approximately 126.3° (inset of Figure 3D; Rayleigh test, $p = .01$), again supporting an out-of-phase alternating relationship between the high- and low-contrast performances. Thus, perceptual predictions associated with different levels of saliency are organized via a theta-band

rhythmic course and are optimally processed in different phases within each theta-band cycle.

10-Hz Entrainment Effects

Previous studies have revealed a 10-Hz periodic modulation of behavioral performance after 10-Hz entrainment (Spaak et al., 2014). Thus, we examined whether our results also contained a 10-Hz entrainment effect. Rather than employing an evoked analysis (performing the spectrum analysis on the grand-averaged results across participants), we conducted an induced analysis by analyzing the frequency contents of the high- and low-contrast behavior performances (entrained–nonentrained) in each participant respectively, considering that the phase of the elicited 10-Hz fluctuation in behavior might be inconsistent across participants. As illustrated in Figure 4A–B, both Experiments 1 and 2 showed significant activations at approximately 10 Hz (permutation test, $p \leq .05$), suggesting that the 10-Hz entrainment did result in a periodic modulation of subsequent behavioral performance, consistent with previous findings (Hickok, Farahbod, & Saberi, 2015; Spaak et al., 2014). Recent

studies have shown that alpha-band stimulation largely drives the endogenous alpha-band neuronal rhythms that differ across individuals (Samaha & Postle, 2015). Therefore, the same 10-Hz external entrainment might induce alpha-band entrainment with a different phase shift in each participant, leading to the observed inconsistent alpha phase across participants.

Experiment 3

After establishing that the high- and low-contrast stimuli were alternatively processed in a theta-band rhythm (Figures 2C and 3B), we further explored whether the dissociated temporal profiles were because of their different physical properties (i.e., high- and low-contrast stimuli) (Gollisch & Meister, 2008) or indeed derived from the mutual competition between the two perceptual predictions (Huang et al., 2015). To address the issue, rather than mixing high- and low-contrast targets within the same block as in Experiments 1 and 2, we presented them in separate experimental blocks in Experiment 3. If the physical properties determine the temporal profiles (e.g., high contrast earlier than low contrast), we would

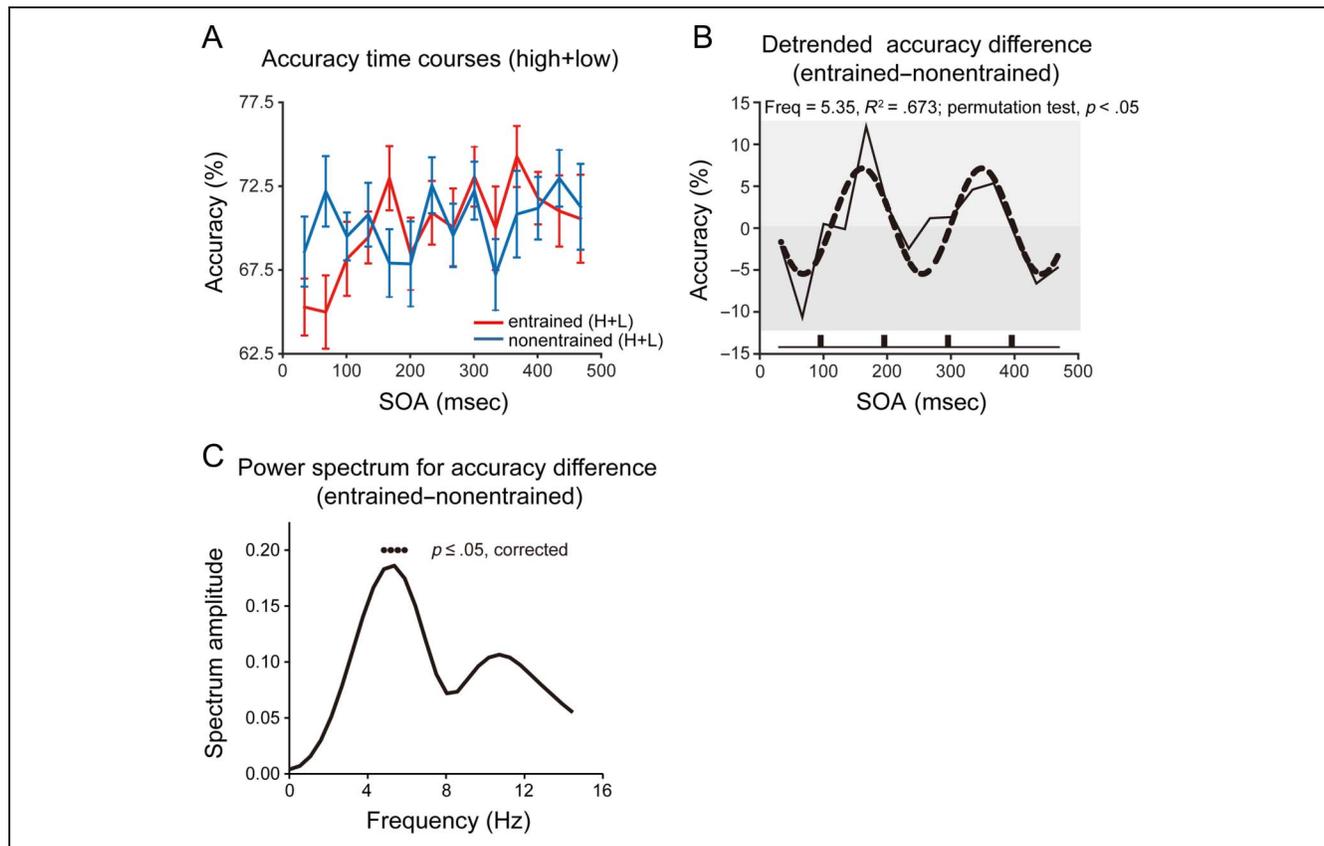


Figure 5. Temporal dynamics in spatial attention. (A) Grand-averaged accuracy time courses as a function of SOA, for the entrained (red) and nonentrained (blue) conditions (grouping the high- and low-contrast conditions). Error bars represent $\pm 1 SEM$. (B) Detrended grand-averaged accuracy time courses as a function of SOA, for the difference between the entrained and nonentrained conditions, with the best sinusoidal fit as a dashed line. (C) Power spectrum of results shown in B as a function of frequency. Black dots indicate frequency points showing significant activations (permutation tests, $p \leq .05$).

expect to find similar rhythmic profiles as before (Figure 2B). In contrast, if it is the mutual competition between perceptual predictions that causes the different temporal profiles, the high- and low-contrast conditions would not show the temporal dissociation, because they are now divided into different blocks, in which participants would anticipate one rather than two perceptual expectations.

The overall hit rate (accuracy averaged across all tested SOAs) for the high-contrast targets was still higher than that for the low-contrast ones (two-way repeated ANOVA, $F(1, 12) = 6.492, p = .026$), consistent with Experiments 1 and 2. However, as shown in Figure 4C, which plots the behavioral time courses (entrained–nonentrained) and the best sinusoidal fitting (dashed line) for the high-contrast (red) and low-contrast (blue) targets, the two conditions showed a similar monotonically increasing pattern, different from the rhythmic profiles observed before.

Thus, the theta-band temporal coordination of targets is not simply attributable to their different physical properties but occurs when competing perceptual predictions coexist within the same block.

Behavioral Oscillation in Spatial Attention

In each trial of this study, a visual target would appear at either the entrained or nonentrained locations. Our original analysis, by following the procedure elaborated in previous work (Spaak et al., 2014), regarded the nonentrained performance as a baseline to be subtracted from the entrained one. Meanwhile, several studies have demonstrated that attention alternates between two locations at a theta rhythm (Song et al., 2014; Fiebelkorn, Saalman, et al., 2013; Landau & Fries, 2012). Therefore, we would expect to find theta-band behavioral oscillation when comparing the results for the entrained and nonentrained locations.

To improve the signal-to-noise ratio, we grouped the high- and low-contrast conditions and examined the behavioral temporal course for the entrained (red) and nonentrained (blue) locations (Figure 5A). We used the same analysis procedure as we did for the high- and low-contrast conditions. As shown in Figure 5B, the detrended entrained–nonentrained difference waveform revealed a 5-Hz periodic pattern (sinusoidal fitting, $\text{freq} = 5.35 \text{ Hz}$, $R^2 = .673$; permutation test, $p < .05$). Spectrum analysis (Figure 4C) further confirmed a significant peak of approximately 5 Hz (permutation test, $p \leq .05$), supporting that attention indeed samples the two locations in a theta-band rhythm, consistent with previous findings (Song et al., 2014; Fiebelkorn, Saalman, et al., 2013; Landau & Fries, 2012).

Therefore, two time-based processes occur in the present experiment. Attention samples locations rhythmically (i.e., appeared at either entrained or nonentrained locations), and low- and high-saliency predictions elicit theta-band rhythms with systematic phase lag.

DISCUSSION

Using a fine-grained behavioral measurement, we investigated how the brain coordinates multiple perceptual predictions that vary in saliency. We first revealed that high- and low-salient predictions are processed in a temporally dissociated manner (i.e., optimally processed in different time windows; Experiment 1). When extending the tested time range, we further demonstrated that perceptual hypotheses with different levels of saliency are organized via a theta-band rhythmic course (i.e., optimally processed in different phases within each theta-band cycle; Experiment 2). Finally, such periodic behavioral profiles were no longer present when high- and low-salient targets were presented in separate blocks and thus could not constitute a competition relationship (Experiment 3). Taken together, our findings support that attention processes and coordinates multiple perceptual predictions, which are endowed with different levels of bottom–up saliency, through a theta-band rhythm.

A series of recent studies have shown that attention samples multiple locations, features, and objects rhythmically (Dugue et al., 2016; Drewes et al., 2015; Huang et al., 2015; Song et al., 2014; Fiebelkorn, Saalman, et al., 2013; Landau & Fries, 2012), by using exogenous cues or top–down attentional tasks to manipulate attentional distribution across locations or objects. Several studies have also revealed the neural mechanism for this rhythmic processing (Fiebelkorn et al., 2018; Helfrich et al., 2018; Spyropoulos, Bosman, & Fries, 2018; Jia et al., 2017; Landau, Schreyer, Van Pelt, & Fries, 2015). Here, we demonstrate that the rhythmic processing also occurs in a saliency-based context. It is well known that the saliency information extracted from a visual scene is represented as a “saliency map,” which provides an efficient control strategy for attentional deployment (Zhang, Zhaoping, Zhou, & Fang, 2012; Itti & Koch, 2001; Koch & Ullman, 1985). Specifically, it directs attention to first stay on the most salient location in a winner-take-all fashion, followed by the shift to the next most salient location. Here, we have also shown a latency-based coding of saliency information, which is conceptually in line with the saliency map theory. However, it is noteworthy that our results could not simply be accounted for by this theory, because the high- and low-salient stimuli have never been displayed simultaneously within the same scene. Furthermore, we revealed a theta-band rhythmic switching when assessing performance at longer temporal lags, implying that the saliency-based processing is not only a one-round course as the saliency map theory suggests but also involves a multirun temporal development (Buschman & Kastner, 2015; Huang et al., 2015; Chen et al., 2014; Enns & Di Lollo, 2000).

Jensen and colleagues have proposed a time-based saliency representation model (Jensen et al., 2012, 2014) in which posterior inhibitory alpha-band neuronal

oscillation serves as a mechanism for prioritizing and ordering unattended visual input based on their saliency. Specifically, as inhibition gradually lowers within an alpha cycle, unattended objects are sequentially encoded in accordance with their saliency, resulting in consecutive activations within each alpha cycle. Our findings that multiple objects with different saliency are coordinated via a rhythmic process are thus partly in line with this model. Meanwhile, instead of an alpha-band oscillatory profile as proposed by the model, we found that the theta-band plays a central function. A possible reason is that, in our tasks, all the targets are task relevant and need to be attended to, whereas the previous model is mainly proposed for representations of unattended items.

Most importantly, instead of examining how attention samples a multitude of locations or objects that are presented explicitly in the same display, here, our results support that the internally generated perceptual hypotheses about incoming inputs are also organized and coordinated through a rhythm-based time framework according to their relative saliency. In fact, prediction is known to serve a central function in influencing and even determining perception (de Lange et al., 2018). Perceptual prediction has been found to elicit neuronal activities in a very temporally precise way, independent of physical inputs (Ekman, Kok, & de Lange, 2017). Most relevantly, a recent study, by assessing the time-resolved visual priming behavioral performance, has revealed an alternating theta-band rhythmic course for the congruent and incongruent conditions, suggesting that multiple perceptual predictions are conveyed in a rhythmic manner to be compared with the physical inputs and that the hypotheses are harmonized in time by occupying different phases of the theta-band oscillations (Huang et al., 2015). Thus, the rhythm-based temporal organization not only operates over objects that are shown explicitly in visual displays but also engages in coordinating multiple perceptual predictions that are kept in internal attention.

Furthermore, we do not think that eye movements could account for these results. First, previous studies have provided extensive evidence that behavioral oscillation revealed in attentional behavioral performances is not because of eye movement (Huang et al., 2015; Song et al., 2014; Landau & Fries, 2012). In addition, recent studies have shown that microsaccades, in the presence of central fixation, are closely associated with attention and could even reveal oscillatory profiles (Hafed, 2013; Bosman, Womelsdorf, Desimone, & Fries, 2009). Thus, temporal dynamics might be an inherent property of attentional processing.

By applying a 10-Hz luminance entrainment before the target, we observed a theta-band profile, consistent with previous attention findings (Huang et al., 2015; Song et al., 2014; Fiebelkorn, Saalman, et al., 2013). Theta-band rhythm is well known to mediate many

cognitive processes, including sensory stream processing (Fiebelkorn, Snyder, et al., 2013; Kayser, Ince, & Panzeri, 2012; Luo & Poeppel, 2007, 2012; Luo, Liu, & Poeppel, 2010), memory (Luo, Tian, Song, Zhou, & Poeppel, 2013; Lisman & Idiart, 1995), and spatial navigation by cyclically modulating cortical excitability and segmenting inputs into appropriate chunks in time (Giraud & Poeppel, 2012; Kayser et al., 2012). Our results thus constitute new evidence supporting that a multitude of perceptual predictions that are held in internal attention are also represented via a theta-band dynamic course. Specifically, the high-salient prediction that is associated with the most excited neuronal representation will discharge first during each theta-band attentional chunk, followed by the second most excitable item (i.e., here the low-salient target). Another theta-band cycle then begins by monitoring possible new incoming targets and temporally sorting them according to their relative saliency. Our results demonstrate that this temporal organization profile vanishes when the two hypotheses are not maintained simultaneously. Because of the short durations tested in Experiment 3, however, further studies are needed to systematically examine the idea.

In conclusion, we demonstrate that perceptual hypotheses with different levels of saliencies are organized via a theta-band rhythm and are optimally processed in different phases within each cycle. Such periodic behavioral profiles result from the competition between predictions rather than their physical distinctiveness. Our results, in combination with previous attention sampling studies, provide support for the rhythmic nature of attention.

Acknowledgments

We thank Dr. Nai Ding and Ms. Ying Fan for helpful comments. This work was supported by National Nature Science Foundation of China grants 31522027 and 31571115 and Beijing Municipal Science & Technology Commission grant Z181100001518002 to L. H.

Reprint requests should be sent to Huan Luo, School of Psychological and Cognitive Sciences, Peking University IDG/McGovern Institute for Brain Science, Peking University, 52 Haidian Road, Beijing 100087, China, or via e-mail: huan.luo@pku.edu.cn.

Note

1. Supplementary material for this paper can be retrieved from www.psy.pku.edu.cn/szdw/qzjy/lh/index.htm.

REFERENCES

- Bosman, C. A., Womelsdorf, T., Desimone, R., & Fries, P. (2009). A microsaccadic rhythm modulates gamma-band synchronization and behavior. *Journal of Neuroscience*, *29*, 9471–9480.
- Buschman, T. J., & Kastner, S. (2015). From behavior to neural dynamics: An integrated theory of attention. *Neuron*, *88*, 127–144.

- Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., & Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron*, *82*, 682–694.
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Drewes, J., Zhu, W., Wutz, A., & Melcher, D. (2015). Dense sampling reveals behavioral oscillations in rapid visual categorization. *Scientific Reports*, *5*, 16290.
- Dugue, L., Roberts, M., & Carrasco, M. (2016). Attention reorients periodically. *Current Biology*, *26*, 1595–1601.
- Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*, 15276.
- Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*, 345–352.
- Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018). A dynamic interplay within the frontoparietal network underlies rhythmic spatial attention. *Neuron*, *99*, 842.e8–853.e8.
- Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Current Biology*, *23*, 2553–2558.
- Fiebelkorn, I. C., Snyder, A. C., Mercier, M. R., Butler, J. S., Molholm, S., & Foxe, J. J. (2013). Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage*, *69*, 126–137.
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517.
- Gollisch, T., & Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science*, *319*, 1108–1111.
- Haegens, S., Luther, L., & Jensen, O. (2012). Somatosensory anticipatory alpha activity increases to suppress distracting input. *Journal of Cognitive Neuroscience*, *24*, 677–685.
- Hafed, Z. M. (2013). Alteration of visual perception prior to microsaccades. *Neuron*, *77*, 775–786.
- Händel, B. F., Haarmeier, T., & Jensen, O. (2011). Alpha oscillations correlate with the successful inhibition of unattended stimuli. *Journal of Cognitive Neuroscience*, *23*, 2494–2502.
- Helfrich, R. F., Fiebelkorn, I. C., Szczepanski, S. M., Lin, J. J., Parvizi, J., Knight, R. T., et al. (2018). Neural mechanisms of sustained attention are rhythmic. *Neuron*, *99*, 854.e5–865.e5.
- Herrmann, C. S. (2001). Human EEG responses to 1–100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, *137*, 346–353.
- Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science*, *26*, 1006–1013.
- Ho, H. T., Leung, J., Burr, D. C., Alais, D., & Morrone, M. C. (2017). Auditory sensitivity and decision criteria oscillate at different frequencies separately for the two ears. *Current Biology*, *27*, 3643.e3–3649.e3.
- Huang, Y., Chen, L., & Luo, H. (2015). Behavioral oscillation in priming: Competing perceptual predictions conveyed in alternating theta-band rhythms. *Journal of Neuroscience*, *35*, 2830–2837.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.
- Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*, 200–206.
- Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences*, *37*, 357–369.
- Jia, J., Liu, L., Fang, F., & Luo, H. (2017). Sequential sampling of visual objects during sustained attention. *PLoS Biology*, *15*, e2001903.
- Kaysner, C., Ince, R. A., & Panzeri, S. (2012). Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS Computational Biology*, *8*, e1002717.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*, 606–617.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.
- Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*, 1000–1004.
- Landau, A. N., Schreyer, H. M., Van Pelt, S., & Fries, P. (2015). Distributed attention is implemented through theta-rhythmic gamma modulation. *Current Biology*, *25*, 2332–2337.
- Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7 plus/minus 2 short-term memories in oscillatory subcycles. *Science*, *267*, 1512–1515.
- Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*, e1000445.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*, 1001–1010.
- Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: Evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology*, *3*, 170.
- Luo, H., Tian, X., Song, K., Zhou, K., & Poeppel, D. (2013). Neural response phase tracks how listeners learn new acoustic representations. *Current Biology*, *23*, 968–974.
- Rucci, M., Ahissar, E., & Burr, D. (2018). Temporal coding of visual space. *Trends in Cognitive Sciences*, *22*, 883–895.
- Samaha, J., & Postle, B. R. (2015). The speed of alpha-band oscillations predicts the temporal resolution of visual perception. *Current Biology*, *25*, 2985–2990.
- Song, K., Meng, M., Chen, L., Zhou, K., & Luo, H. (2014). Behavioral oscillations in attention: Rhythmic alpha pulses mediated through theta band. *Journal of Neuroscience*, *34*, 4837–4844.
- Spaak, E., de Lange, F. P., & Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *Journal of Neuroscience*, *34*, 3536–3544.
- Spyropoulos, G., Bosman, C. A., & Fries, P. (2018). A theta rhythm in macaque visual cortex and its attentional modulation. *Proceedings of the National Academy of Sciences, U.S.A.*, *115*, E5614–E5623.
- Tomassini, A., Spinelli, D., Jacono, M., Sandini, G., & Morrone, M. C. (2015). Rhythmic oscillations of visual contrast sensitivity synchronized with action. *Journal of Neuroscience*, *35*, 7019–7029.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception & Psychophysics*, *60*, 140–156.
- Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural activities in v1 create a bottom-up saliency map. *Neuron*, *73*, 183–192.