EJN European Journal of Neuroscience   FENS   WILEY

The impact of phase entrainment on auditory detection is highly variable: Revisiting a key finding

Yue Sun[1] | Georgios Michalareas[1] | David Poeppel[1,2,3,4]

[1]Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

[2]Department of Psychology, New York University, New York, New York, USA

[3]Max Planck-NYU Center for Language, Music, and Emotion (CLaME), New York, New York, USA

[4]Ernst Strüngmann Institute for Neuroscience, Frankfurt, Germany

Correspondence
Dr. Yue Sun, Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt, Germany.
Email: yue.sun@ae.mpg.de

Edited by: Christian Keitel

Abstract

Ample evidence shows that the human brain carefully tracks acoustic temporal regularities in the input, perhaps by entraining cortical neural oscillations to the rate of the stimulation. To what extent the entrained oscillatory activity influences processing of upcoming auditory events remains debated. Here, we revisit a critical finding from Hickok et al. (2015) that demonstrated a clear impact of auditory entrainment on subsequent auditory detection. Participants were asked to detect tones embedded in stationary noise, following a noise that was amplitude modulated at 3 Hz. Tonal targets occurred at various phases relative to the preceding noise modulation. The original study ($N = 5$) showed that the detectability of the tones (presented at near-threshold intensity) fluctuated cyclically at the same rate as the preceding noise modulation. We conducted an exact replication of the original paradigm ($N = 23$) and a conceptual replication using a shorter experimental procedure ($N = 24$). Neither experiment revealed significant entrainment effects at the group level. A restricted analysis on the subset of participants (36%) who did show the entrainment effect revealed no consistent phase alignment between detection facilitation and the preceding rhythmic modulation. Interestingly, both experiments showed group-wide presence of a non-cyclic behavioural pattern, wherein participants' detection of the tonal targets was lower at early and late time points of the target period. The two experiments highlight both the sensitivity of the task to elicit oscillatory entrainment and the striking individual variability in performance.

KEYWORDS
auditory perception, brain rhythm, neural oscillations, phase locking, replication

1 | INTRODUCTION

There is ample empirical evidence showing that the brain tracks temporal regularities in the acoustic input, perhaps by synchronizing slow rhythmic brain activity to the temporal regularities in stimuli (Ahissar et al., 2001; Arnal et al., 2015; Doelling et al., 2014; Howard & Poeppel, 2010; Peelle, 2012; Peelle et al., 2013). It has been proposed that neural oscillations, which reflect rhythmic fluctuations of local neuronal excitability (Buzsáki & Draguhn, 2004; Lakatos et al., 2005), play an instrumental role in capturing rhythmic structures in the acoustic input (Schroeder & Lakatos, 2009; Zion Golumbic et al., 2013). Such synchronization, otherwise referred to as 'entrainment', involves the temporal adjustment of the period and/or the phase of neural oscillations to dominant rhythms in the acoustic input (Poeppel & Teng, 2020; Thut et al., 2011), such that the high-excitability phase of the oscillations is temporally aligned with critical acoustic events in the input signal (Giraud & Poeppel, 2012; Kayser et al., 2009; Lakatos et al., 2005, 2008; Schroeder & Lakatos, 2009). Furthermore, it has been demonstrated that entrained oscillatory activity can be sustained for a brief time after the termination of the rhythmic stimulation (Lakatos et al., 2013). Cumulatively, the findings suggest a mechanism through which rhythmicity in the preceding acoustic input is internalized by the human brain to optimize auditory processing of upcoming signals (Lakatos et al., 2019).

Recent behavioural and neural studies have provided evidence for the impact of oscillatory entrainment after rhythmic stimulation (e.g., Hickok et al., 2015; Kösem et al., 2018). For instance, several studies show that the rate of a preceding acoustic rhythm regulates the 'sampling rate' of the auditory system and thereby alters participants' perception of upcoming ambiguous stimuli (Bosker & Ghitza, 2018; Dilley & Pitt, 2010; Kösem et al., 2018). Other studies have tested whether the phase of entrained oscillations influences the processing of upcoming signals (e.g., Hickok et al., 2015; Jones et al., 2002). These studies typically used a paradigm in which they presented an acoustic target with equal probabilities across a range of time points after the offset of a rhythmic stimulus. If oscillatory entrainment to an acoustic rhythm outlasts the stimulation period and influences the perception of a following target, one should expect to observe the same type of rhythmicity in the temporal fluctuation of participants' performance on the target. For instance, in a series of studies, Jones et al. showed that exposure to a sequence of periodically presented pure tones facilitated participants' processing of an upcoming target tone that occurred at the time predicted by the rhythm of the preceding sequence

(Barnes & Jones, 2000; Jones et al., 2002, 2006). These findings demonstrated a specific form of the entrainment effect wherein the rhythmic carrier provides a direct prediction for the timing of an upcoming auditory event and thereby facilitates processing of target stimuli that occur at the predicted timing (in phase) as opposed to those out of phase. Other forms of the entrainment effect have been revealed by studies that did not use the approach of direct temporal probing (Farahbod et al., 2020; Hickok et al., 2015). For instance, Hickok et al. (2015) exposed participants to a broadband noise stimulus whose amplitude was sinusoidally modulated at 3 Hz. This exposure period was followed by a target period during which a brief 1-kHz tone target was presented with equal probability across nine temporal positions covering two cycles of the preceding amplitude modulation. They showed that participants' detection of the tonal target at different temporal positions fluctuated cyclically at the same rate as the noise modulation. Interestingly, the cyclic pattern of target detectability showed an anti-phase relation with the noise modulation, with detection accuracy reaching local maxima when the noise signal would have approached the minimum amplitude if the modulation had continued. The authors argued that the observed entrainment effect cannot be explained by a direct temporal prediction mechanism but reflects the impact of specific phase of the slow neural oscillations on perception of the upcoming stimuli.[1]

These studies provide evidence for the impact of oscillatory entrainment to a preceding rhythmic signal on upcoming perceptual processes, but the robustness of these effects remains an active topic of debate (see Haegens & Zion Golumbic, 2018, for a review; Lin et al., 2021, for an empirical challenge). One critical issue concerns the degree of variability of entrainment effects across participants and studies. For instance, in revisiting the paradigm used in Jones et al. (2002), Bauer et al. (2015) failed to replicate the entrainment effect at the group level that was observed in the original study. Crucially, after assessing individual data, the authors found that only 40 of 140 tested participants exhibited the original effect. The authors suggested that the specific pitch comparison task used in the original paradigm may contribute to this variability, as the task can induce mixed listening strategies among participants regarding their treatment of the entraining signal. Behavioural differences also exist among participants who have exhibited a positive entrainment effect. These differences concern the temporal alignment between the phase of the entraining signal and the effect of perceptual facilitation

---

[1]Note that Farahbod et al. (2020) argued for an alternative attentional mechanism to account for the effect.

on acoustic targets, which has been mostly demonstrated in studies using rhythmic carriers that undergo continuous acoustic modulations (Farahbod et al., 2020; Forseth et al., 2020; Henry et al., 2014; Henry & Obleser, 2012; Hickok et al., 2015). For instance, Henry et al. (2014) and Henry and Obleser (2012) observed that while an ongoing frequency modulation of a background noise impacted most participants' detection of a short acoustic gap embedded at different locations of the noise, there was little convergence across participants regarding the exact phase of the modulation that induced facilitation of target detection. In the case of perceptual facilitations from post stimulation entrainment, Hickok et al. (2015) found an anti-phase effect that was consistent across the tested participants. However, in a subsequent study using the same paradigm (Forseth et al., 2020), the authors observed that the perceptual facilitation for target detection seemed to occur in-phase with the preceding modulation, with an increase of target detectability during the rising portion of the amplitude modulation.[2] The conflicting findings from the two studies cast doubt on the robustness with which perceptual facilitation is mechanistically associated with entrainment to a specific phase of amplitude modulation in the acoustic signal.

Here, we aim to address the issue of individual variability in the effect of neural entrainment on the processing of post-stimulation targets. Specifically, we revisit the study of Hickok et al. (2015), as it provides an opportunity to examine both aspects of the variability mentioned above. One particular aspect of that study that is of significance is the small sample size ($N = 5$), which by itself is not necessarily problematic (the authors argue that the small sample size is justified by the large number of trials each participant performed for each experimental condition). Moreover, the individual results showed substantial convergence across the participants in terms of both the existence and the phase alignment of the entrainment effect. However, several studies incorporating entrainment phenomena indeed reveal population-level differences in the existence of effects at the individual level (Assaneo et al., 2019; Assaneo, Rimmele et al., 2021; Bauer et al., 2015). The findings from these studies motivate a reassessment of the nature of the entrainment effect from Hickok et al. (2015) across a larger number of participants. A second unique feature of Hickok et al. (2015) is the way the authors presented target stimuli during the experiment and the way in which they selected trials for analysis. In fact, one common design across previous studies is to

present the task-related property of the target stimuli at a near-threshold level, which decreases the likelihood of ceiling performance and enhances the sensitivity of the paradigm to reveal the effect of the experimental manipulation. In order to achieve this goal, a common practice has been to measure the level of the acoustic property that corresponds to threshold-level performance for each participant and to present the target stimuli at this level during the main experiment (Henry et al., 2014; Henry & Obleser, 2012; Ng et al., 2012). In contrast, Hickok et al. (2015) used a different approach: they presented the tone target at five levels of intensity during the main experiment, orthogonal to the manipulation of the target's temporal position. They observed that targets at a specific intensity level yielded—across all the five participants—a response accuracy that was closest to their detection threshold and thus included trials only from this level in the analysis on the impact of temporal position on target detectability. One downside of this design choice is the substantial number of unanalyzed trials, corresponding to ~80% of the total trials. The authors argue, building on findings from a later study, that the inclusion of intensity variation during the experiment is crucial for the observation of the entrainment effect (Farahbod et al., 2020). In the current study, we will also examine whether the original entrainment effect can be reproduced with a shortened experimental procedure. Altogether, the current study assesses the robustness of the entrainment effect from Hickok et al. (2015) using both a larger sample size and a procedural modification, on the basis of two experiments: one is a shortened experimental procedure (i.e., *conceptual replication*) and the other is the same procedure as the original study (i.e., *exact replication*).

2 | MATERIALS AND METHODS

2.1 | Participants

A total of 47 participants (28 females; average age: 22.04, range 18 to 27) provided written informed consent to take part in the study and received monetary compensation for their participation. Twenty-four took part in Experiment 1 (conceptual replication) and 23 in Experiment 2 (exact replication). All participants reported normal hearing. The experimental procedure was approved by the Ethics Council of the Max-Planck Society (no. 2017_12).

2.2 | Stimuli

In both experiments, we used the same stimulus design as Hickok et al. (2015). Each trial consisted of a

---

[2]Note that the experimental design in Forseth et al. (2020) was not suitable to reveal a bi-cyclic pattern in target detectability as in Hickok et al. (2015). However, results of Forseth et al. (2020) did not conform to the prediction that would have been generated from an anti-phase relation between amplitude modulation and target detectability.
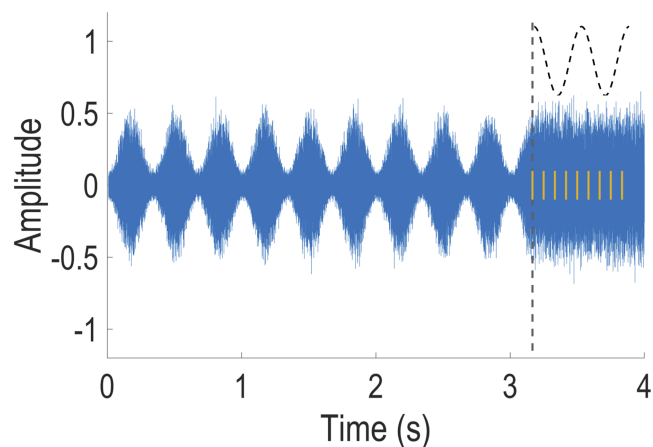
FIGURE 1 Stimulus design for the two experiments (adapted from Hickok et al., 2015). The Gaussian noise (blue) was amplitude modulated at 3 Hz for the first 3.167 s, then flat for 0.833 s. The grey dashed line marks the end of amplitude modulation. The black dashed curve represents the expected amplitude envelope of the noise had the modulation carried on. The yellow solid vertical lines indicate the nine temporal positions at which a tonal target could occur

broadband Gaussian noise that lasted for 4 s (Figure 1). The amplitude of the first 3.167 s of noise was sinusoidally modulated at 3 Hz with 80% modulation depth. This portion of the noise corresponds to 9.5 cycles of modulation, starting with the lowest amplitude and ending at maximum amplitude. The final 0.833 s of the noise were unmodulated, with the amplitude remaining at the maximum level. In half of the trials, a 50-ms duration 1-kHz pure tone target (with 5-ms rise-and-decay time) was presented at one of nine temporal positions. These positions started at the offset of the amplitude modulation (3.167 s from the onset of the noise) and were successively spaced 83.3 ms apart. The interval between two successive target positions corresponded to one-quarter of a modulation cycle, such that the nine temporal positions covered the time period equivalent to two full cycles of the amplitude modulation if the modulation had continued. In Hickok et al. (2015), the authors presented the tonal targets at five different levels of signal-to-noise ratio (SNR) with respect to the unmodulated part of the broadband noise. The two experiments in the current study differ in this feature of the stimulus design.

2.2.1 | Experiment 1

In Experiment 1, we presented the tonal target at two SNR levels only. The two levels corresponded to (a) the *near-threshold* level, which was determined for each participant using a staircase procedure prior to the

experiment and (b) the *above-threshold* level, which was fixed at 6 dB above the threshold level for each participant. Note that this design preserves the presence of intensity uncertainty in auditory targets, which was argued to be critical for the entrainment effect to take place (Farahbod et al., 2020).

2.2.2 | Experiment 2

In Experiment 2, we adopted the same design as the original Hickok et al. (2015) study, which presented the tonal targets at five SNR levels covering a range of 12 dB. Specifically, the five levels were composed of one below-threshold level (Level 1), one near-threshold level (Level 2) and three above-threshold levels (Levels 3–5). Note that the same SNR levels were used for all participants, such that the near-threshold level (Level 2) in the original study was not determined for each participant. In the current experiment, we aimed to recreate the same correspondence between SNR levels and overall detection accuracy as in the original study. The key was to assure that targets at Level 2 resulted in near-threshold performance at the group level. In order to achieve this goal, we estimated the group level detection threshold by averaging individual threshold SNRs across all participants from Experiment 1 and used this level as SNR Level 2 in Experiment 2. Finally, the SNR of the other four levels were spaced such that the five levels covered 12 dB with equal distance.

2.3 | Procedure

For both experiments, participants were seated in a sound-proof booth in front of a LCD monitor to receive instructions and feedback during the experiment. Auditory stimuli were generated using MATLAB (The MathWorks, Natick, MA, USA) at 44.1 kHz/16 bits, output by a high-quality interface (RME Fireface UCX) and presented to participants binaurally via electrodynamic headphones (Beyerdynamic DT770 PRO). The output intensity of the stimuli was calibrated at 70 dB (A-weighted) for the unmodulated part of the noise. Participants' responses were collected with a Cedrus response box (RB-844, Cedrus Corporation, San Pedro, CA, USA). The experiment was run using MATLAB Psychophysics Toolbox extensions (Brainard, 1997) on a Fujitsu Celsius M730 computer running Windows 7 (64 bit).

On a given trial, participants were instructed to listen to the auditory stimulus and to respond whether a tone signal was present during the unmodulated portion of

the broadband noise by pressing one of the two buttons assigned to 'present' and 'absent' responses. For each trial, participants were given a response time window of 1.5 s, measured from the offset of the noise stimulus. If no button press was registered before the elapse of the response time window, a message would be displayed on the monitor reminding the participant to provide a response. After participants' responses, an inter-trial interval randomly distributed between 1 and 2 s occurred before the beginning of the next trial.

2.3.1 | Experiment 1

We first determined the SNR level of near-threshold targets for each participant, using a staircase procedure. During this procedure, each trial was composed of the same elements as those from the main experiment, including 3.167 s of amplitude-modulated noise followed by 0.833 s of target time period with flat amplitude noise. The tonal target, when it was presented, occurred with equal probabilities at each of the nine temporal positions used in the main experiment. Note that we used the exact same trial structure as in the main experiment in order to assure that the near-threshold SNR level determined by the staircase procedure would result in a near-threshold detection rate of the tonal target in the main experiment. Put differently, the goal of this procedure is not to find each participant's 'real' detection threshold of a 1-kHz tone embedded in a broadband noise. Instead, our goal was to determine the SNR level at which the average detection rate of the near-threshold tonal target in the main experiment would reach a level that is comparable with that observed in the original study.[3] The staircase procedure was conducted in two runs, each of which contained 54 trials (27 target trials and 27 non-target trials). During each run, participants were asked to perform the same target detection task as in the main experiment. The first run started with a high SNR value for the tonal target, which progressively decreased upon successful detection of the target by the participant. Specifically, we used the Psi method (Kontsevich & Tyler, 1999)

implemented in the Palamedes toolbox for MATLAB (Prins & Kingdom, 2009). This method provides an estimation of the threshold and slope of a psychometric function of the participant's target detection performance after each trial and uses the estimated parameters to determine the SNR level of the tonal target that will be associated with the following trial. After the first run, we averaged the estimated threshold level SNR over the last 10 trials and used this level as the initial SNR value of the second run. After the second run, we averaged the estimated threshold level SNR over the last 10 trials and used it as the SNR level for the near-threshold targets in the main experiment. A stabilization of the threshold level SNR towards the end of the second run of the staircase procedure was observed for all participants (see Figure S7 for the evolution of SNR values in the two runs of the staircase procedure for all participants).

The main experiment was composed of 720 trials presented across 10 experimental blocks. Each block contained 72 pseudo-randomly ordered trials, with half of them containing a tonal target (i.e., target trials) while the other half not (i.e., non-target trials). When a tonal target was presented, its timing was randomly selected from one of nine temporal positions and its intensity was randomly selected from one of the two SNR levels. Note that the tonal target for each unique combination of temporal position and SNR-level occurred twice in a given block, such that participants' exposure to the two different SNR levels was kept balanced as the experiment progressed. This measure was particularly crucial to prevent participants from developing response biases for certain temporal positions due to overexposure or underexposure during a certain phase of the experiment. Overall, target trials from each combination between temporal positions and SNR levels were repeated 20 times over the 10 experimental blocks. The amount of trial repetitions per condition is comparable to the one from the original study (i.e., 22 on average). Each block lasted about 6 min, and participants were given a short break after each block. The experiment started with a practice phase composed of nine target trials and nine non-target trials. The tone signal in target trials were presented once in each of the nine temporal positions at the above-threshold SNR level. Trials from this phase were excluded from the analysis. Altogether, the experiment lasted for about 1.5 h.

2.3.2 | Experiment 2

Experiment 2 was composed of two sessions of 1.5 h that took place on two consecutive days. Each session had 10 blocks of 90 trials, including 45 targets trials and 45 non-target trials presented in random order. When the

---

[3] In a pilot study, we conducted the staircase procedure with simpler trial structures (e.g., without the preceding amplitude modulated noise or temporal uncertainty for the occurrence of the tonal target). We noticed that the near-threshold SNR level determined by these procedures would consistently result in below-threshold detection rate in following main experiment, in which the trial design was more complex (preceding amplitude modulated noise, uncertainty in the temporal position of the tonal target). These observations led us to use the same trial structure as in the main experiment during the staircase procedure. The SNR level determined by this procedure resulted in near-threshold detection rate in the main experiment across participants.

tonal target was presented, its temporal position was selected from one of the nine positions, and its intensity was randomly selected from one of the five SNR levels. The tonal target with each unique combination between temporal position and SNR level occurred once in a given block, such that participants received balanced exposure to different experimental conditions. Each block lasted around 7.5 min, and participants were given a short break after each block. In total, every participant completed 20 blocks (1800 trials) across the two sessions, which resulted in 20 repetitions of target trials for each combination of temporal position and SNR level, similarly to Experiment 1. This experiment also started with a practice session composed of nine target trials and nine non-target trials. The tone signal in the practice target trials was presented at the highest SNR level and its temporal position was uniformly distributed across the target trials (i.e., each position was presented once). Results from this phase were excluded from the analysis.

2.4 | Data analyses

Similar to the original study, we focused on the impact of temporal position on participants' detection of near-threshold targets, corresponding to targets at near-threshold SNR in Experiment 1 and those at SNR Level 2 in Experiment 2. In order to examine the entrainment phenomenon, we introduced three analytic modifications compared to the original study.

2.4.1 | Trial selection

The first modification consisted of the exclusion of non-target trials from the analysis. In Hickok et al. (2015), the authors measured the average response accuracy at each temporal position for all trials associated with the position, including both target and non-target trials. We argue that the inclusion of non-target trials is problematic. Although each target trial can be objectively associated with a specific experimental condition based on the physical properties of the tonal signal that was presented during the trial, the association between non-target trials and experimental conditions can only be made arbitrarily, due to the absence of tonal targets in these trials. Consequently, only participants' performance on target trials (i.e., hits and misses) can be attributed to manipulations of the tonal target, while their performance on non-target trials (i.e., false alarms and correct rejections) cannot.

To further motivate our rationale regarding non-target trials, consider a scenario wherein a participant gives a false alarm response to a non-target trial that was labelled, for example, as 'Position 4'. The association between the participant's false alarm response and 'Position 4' is solely due to the fact that this non-target trial happened to be *labelled* as 'Position 4' during the random assignment of temporal positions to non-target trials. Thus, there is no basis to link any perceptual processes that drove the participant's false alarm response in this trial to the temporal Position 4. This is due to the randomness in the association between this non-target trial and the label of Position 4 in the first place. Given that there is no objective way to attribute participants' performance on non-target trials to any specific temporal positions, we therefore excluded non-target trials from our analysis; in other words, we examined the variation of hit rate as a function of temporal position of the tone target.

One might argue that only considering hit rates cannot rule out the impact of potential response biases in performance. We agree with this assumption, but we submit that the inclusion of non-target trials in the analysis would actually not solve the issue of response biases. This is, again, due to the fact that the association between false alarm rates and temporal positions arises solely from the random labelling of non-target trials with temporal positions. Thus, there is no basis to assume that the false alarm rate that ends up being associated with a given temporal position reflects the participant's bias to respond 'present' at that temporal position when the target was not presented.

We did, however, test whether participants in our experiments on average presented similar levels of response biases (reflected in the overall false alarm rate across all non-target trials) as participants from the Hickok et al. (2015) study. We also conducted a restricted analysis, which excluded participants whose overall false alarm rate exceeded a cutoff level that was defined by the distribution of overall false alarm rates across the participants from Hickok et al. (2015). These measures assured strong comparability in terms of participants' overall response biases between our experiments and the original study.

2.4.2 | Quantification of the strength of 3-Hz modulation for each participant

The second methodological change consisted of a different quantification of the strength of modulation of target detectability by temporal position for each individual participant. In Hickok et al. (2015), the authors calculated the Fourier transforms of the detection curve from each participant, in order to examine the presence of cyclic

pattern in participants' target detection as a function of temporal position. They observed the peak of the power spectrum for all five (equal to the total number of participants) curves at 3 Hz, which in turn corresponded to the frequency of the amplitude modulation of the preceding noise carrier. However, they did not provide quantifications for the observed 3-Hz power. We argue that such quantifications are important not only for the assessment of cross-participant variabilities in the size of the entrainment effect but also for the estimation of the overall effect size of the entrainment phenomenon at the group level, which can be referred to by follow-up studies to examine the robustness of the entrainment effect against variations in experimental procedure and/or investigated populations.

Here, we quantified the strength of the entrainment phenomenon using the following methodology. First, we performed the spectral decomposition of the curves of target detectability (hit rates) using the standard Fast Fourier Transform (FFT) algorithm, as implemented in MATLAB function fft(). Given that each curve contained $N = 9$ data points with a sampling frequency $Fs = 12$ Hz, the FFT provided a frequency resolution of $Fs/N = 1.33$ Hz and four frequency bins centered at 1.33, 2.67, 4 and 5.33 Hz. Among these four frequencies, we selected 2.67 Hz—which is closest to the modulation rate (3 Hz) in the experiment—to be the frequency of interest and measured its power for each participant. Although power at 2.67 Hz in the power spectrum of participants' hit rate curves is indicative of the presence of a cyclic phenomenon near the entrainment frequency (3 Hz) in participants' detection performance, it is necessary to assess whether and to what extent the observed 2.67-Hz power could be attributed to the effect of entrainment to rhythmicity presented in the preceding stimulus. This requires normalization of the observed 2.67-Hz power in each participant's hit rate curve by a baseline level, which indicates the level 2.67-Hz power that arises from random fluctuations of the hit rate of the tonal target across the nine temporal positions that is unrelated to the entrainment to the preceding stimulus.

In order to establish this baseline level, we created, for each participant, a distribution of 2.67-Hz power using a permutation approach. Specifically, in each round of permutation, we shuffled the label of temporal position of all trials that were included in the analysis. We then calculated the performance curve of the shuffled data and computed the power at 2.67 Hz. For each participant, we repeated this permutation procedure 1000 times, thereby generating a baseline distribution of 2.67-Hz power arising from a random association between the participant's target detection performance and temporal position of the target. We then calculated the mean of the 2.67-

power of the baseline distribution, which indicates the average level of 2.67-Hz power that could arise from random fluctuations of participants' detection performance across the nine temporal positions. Finally, we computed, for each participant, the normalized 2.67-Hz power by subtracting the mean of the baseline distribution from the 2.67-Hz power observed in the real performance curve of the participant. We used the normalized 2.67-Hz power as the dependent variable in statistical analyses at the group level. Our hypothesis was that if participants' detection of the near-threshold targets is influenced by the entrainment to the preceding rhythmic stimulus, then the 2.67-Hz power observed from participants' real performance curves should be significantly higher than the one computed from baseline distribution. That is, the normalized 2.67-Hz power should be significantly higher than 0 at the group level.

2.4.3 | Extraction of common behavioural pattern across participants

In addition to assessing the strength of a bi-cyclic pattern that is present in each participant's detection curve, we also examined whether and to what extent a certain behavioural pattern is shared across all participants. For this purpose, we conducted a principal component analysis (PCA) across all participants' performance curves. Given that temporal variations of hit rate manifest across nine positions, PCA provides a better description of potentially complex patterns (such as a bi-cyclic pattern) within time series data than a mean-based analysis (e.g., analysis of variance used in the original study). Moreover, PCA provides for each extracted component the amount of variance it can account for across participants' data. We specifically focused on the first principal component, which explains the most variance, to highlight the dominant behavioural pattern shared across participants. Accordingly, the amount of variance that can explain by the first principal component indicates the stability of the most common behavioural pattern induced by the experimental manipulation across participants.

We conducted the PCA analysis using the native function 'pca.m' from the Statistics and Machine Learning Toolbox of MATLAB (The MathWorks, Natick, MA, USA). The dataset fed to the function consisted of a two-dimensional matrix, for which the first dimension consisted of the number of observations in each hit rate curve and the second dimension consisted of the number of participants. The data and the script used to conduct the analysis can be found in an online repository (https://doi.org/10.17617/3.5c).

3 | RESULTS

3.1 | Reanalysis of data from the original study: Hickok et al. (2015)

We first analysed the data from the original study of Hickok et al. (2015), using the modified analysis pipeline. The objective of this reanalysis was twofold. First, we sought to provide a more fine-grained quantitative description of the strength of 3-Hz modulation observed in the accuracy curves of the original study. Second, we wanted to compare the strength of 3-Hz modulation between accuracy curves and hit rate curves in order to test to what extent the entrainment effect presented in the original study persisted after the removal of random response noise from non-target trials. Recall that, due to the arbitrary assignment of experimental conditions to non-target trials, participants' performance on these trials cannot be objectively associated with any temporal position (see Section 2.4.1 for a detailed explanation). Modulation strength from hit rate curves then serves as a reference to which results of the current study will be compared. We used the preprocessed data from the original study, which were made public by the authors (Saberi et al., 2020).

We first computed mean hit rates for targets at five SNR levels across different participants (Table 1). Our results confirmed that the detectability of targets at SNR Level 2 is closest to participants' detection threshold ($Mean = 0.61$; standard deviation [SD] = 0.07). We then examined the impact of temporal position on target detectability at SNR Level 2. Accuracy and hit rate curves from the five participants are presented in Figure 2. For accuracy curves, both PCA and the analysis of normalized 2.67-Hz power confirmed strong evidence for near-3-Hz modulation in all five participants' performance. Results from the PCA revealed a dominant bi-cyclic, M-shape, pattern that accounted for 80.97% of the variance across the five participants' performance curves (Figure 3a). Moreover, analysis of the normalized 2.67-Hz power in accuracy curves showed positive modulation strength in all the five participants ($Mean = 1.14$; $SD = 0.39$). For hit rate curves, our results also showed evidence for near-3-Hz modulation, although the

strength of the effect manifested at a lower level than in the case of accuracy. Specifically, PCA analysis showed that the dominant pattern, which had a similar bi-cyclic shape as the one for accuracy curves, provided less explanatory power for the variance cross participants' performance (70.55%; Figure 3b). In a similar vein, analysis of the normalized 2.67-Hz power showed a decreased average modulation strength ($Mean = 0.75$; $SD = 0.56$). In fact, decreases in the modulation strength were observed in four out of five participants (Figure 3c), which is in line with the less apparent cyclic pattern in hit rate curves of these participants compared to accuracy curves (For instance, Figure 2c and 2e).

In summary, reanalysis of the data from the original study using our quantitative methods confirmed the observation of a positive entrainment effect (normalized 2.67-Hz power) in all participants' hit rate curves. Meanwhile, our results also showed that the strength of the effect decreased in most participants, after the removal of their performance in non-target trials, which cannot be objectively associated with any temporal position (see Section 2.4.1 for a detailed explanation).

3.2 | Modulation of target detectability by position: Current study

Having established a new quantitative description of the entrainment effect observed in Hickok et al. (2015), we applied the same analyses to hit rate results from the two experiments of the current study. We first assessed the average hit rate for near-threshold targets in both new experiments. Our results showed comparable hit rates between our experiments (Experiment 1: $N = 24$, $Mean = 0.61$, $SD = 0.14$; Experiment 2: $N = 23$, $Mean = 0.59$, $SD = 0.16$) and the original study ($N = 5$; $Mean = 0.61$; $SD = 0.07$). In addition to the average hit rate for near-threshold targets, we also checked whether participants from our experiments exhibited similar performance in non-target trials as those from the original study. This assessment showed similar false alarm rate across all non-target trials between our experiments (Experiment 1: $N = 24$; $Mean = 0.047$; $SD = 0.055$; Experiment 2: $N = 23$; $Mean = 0.048$; $SD = 0.042$) and the original study ($N = 5$; $Mean = 0.050$; $SD = 0.011$).[4]

We then examined the impact of temporal position on detectability of the near-threshold targets. Interestingly, visual inspection revealed no sign of a bi-cyclic

TABLE 1 Average hit rate for target tones at the five SNR levels across the five participants in Hickok et al. (2015)

	1	2	3	4	5
Mean	0.22	0.61	0.95	0.99	0.99
SD	0.04	0.07	0.01	0.02	0.01

Abbreviations: SD, standard deviation; SNR, signal-to-noise ratio.

---

[4]Due to the fact that one cannot objectively label non-target trials with SNR levels, we could not calculate the average false alarm rate that is specific to the condition of near-threshold SNR. Instead, we reported participants' overall false alarm rate across all non-target trials.
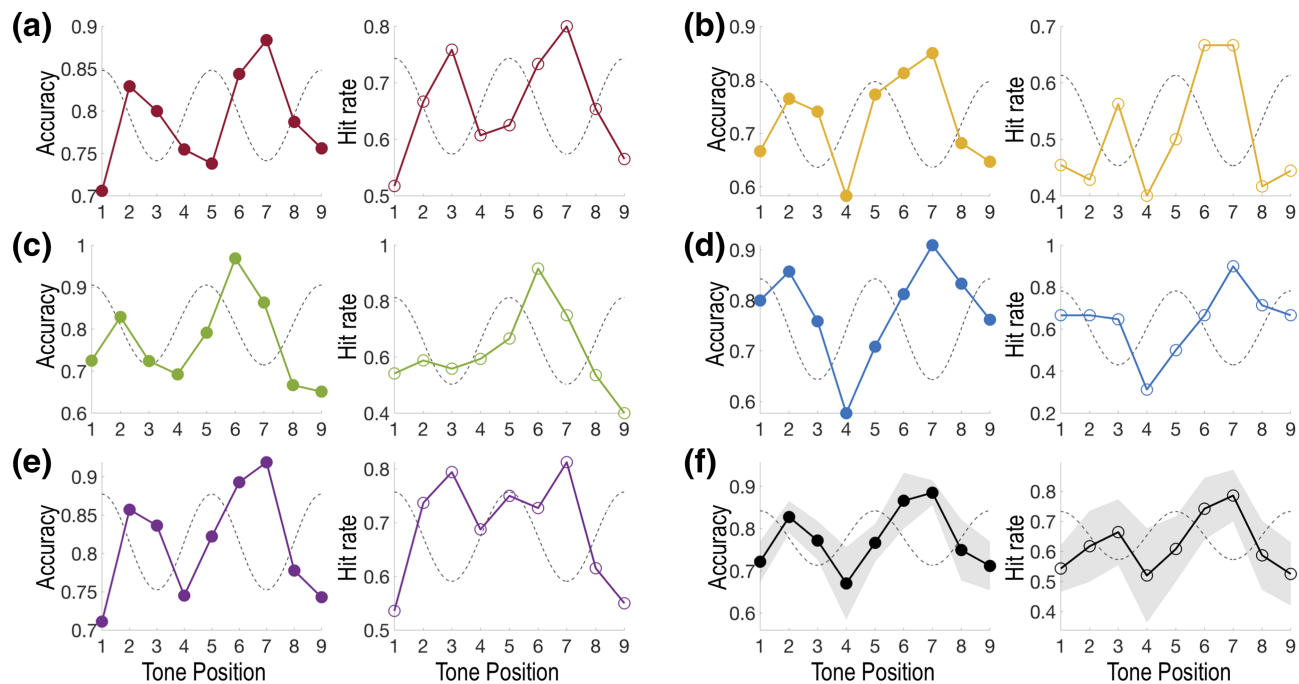
FIGURE 2 Accuracy (filled circles) and hit rate (open circles) of target tones as a function of temporal position from the five participants of Hickok et al. (2015). Data from each individual participant are presented in (a)–(e), and the grand average across participants is presented in (f). Error bars: standard deviation across participants
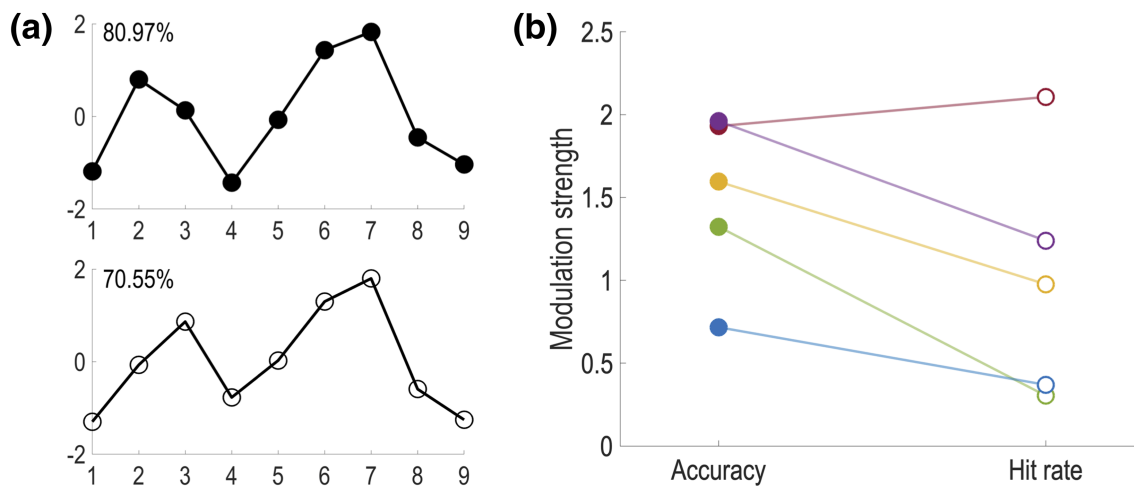

FIGURE 3 (a) The first principal component analysis (PCA) component in participants' accuracy curves (top) and hit rate curves (bottom) from Hickok et al. (2015). Filled circles represent results from accuracy, and open circles represent results from hit rate. The number in the top left corner of each figure indicates the percentage of the total variance explained by the component. (b) Strength of near-3-Hz modulation observed in accuracy and hit rate curves of each participant. Data for each participant follow the same colour code as in Figure 2

pattern in the average hit rate curves across participants for either experiment (Figure 4a; see Figures S1 and S2 for an inspection of individual hit rate curves of each experiment). Results from our analyses quantitatively confirmed this observation. For Experiment 1, the dominant PCA component presented a non-cyclic pattern across the nine temporal positions (Figure 4b). This component accounted for 41.62% of the variance in hit rate

curves across participants and indicated best target detectability at the Positions 5 and 6. Analysis of the normalized 2.67-Hz power revealed that the average modulation strength across participants was not significantly different from 0 (*Mean* $= -0.15$; *SD* $= 0.78$; $t[23] = -0.96$; $p > .1$). Assessment of individual results showed that only 9 out of 24 participants had a positive modulation strength (Figure 5a). For Experiment 2, the
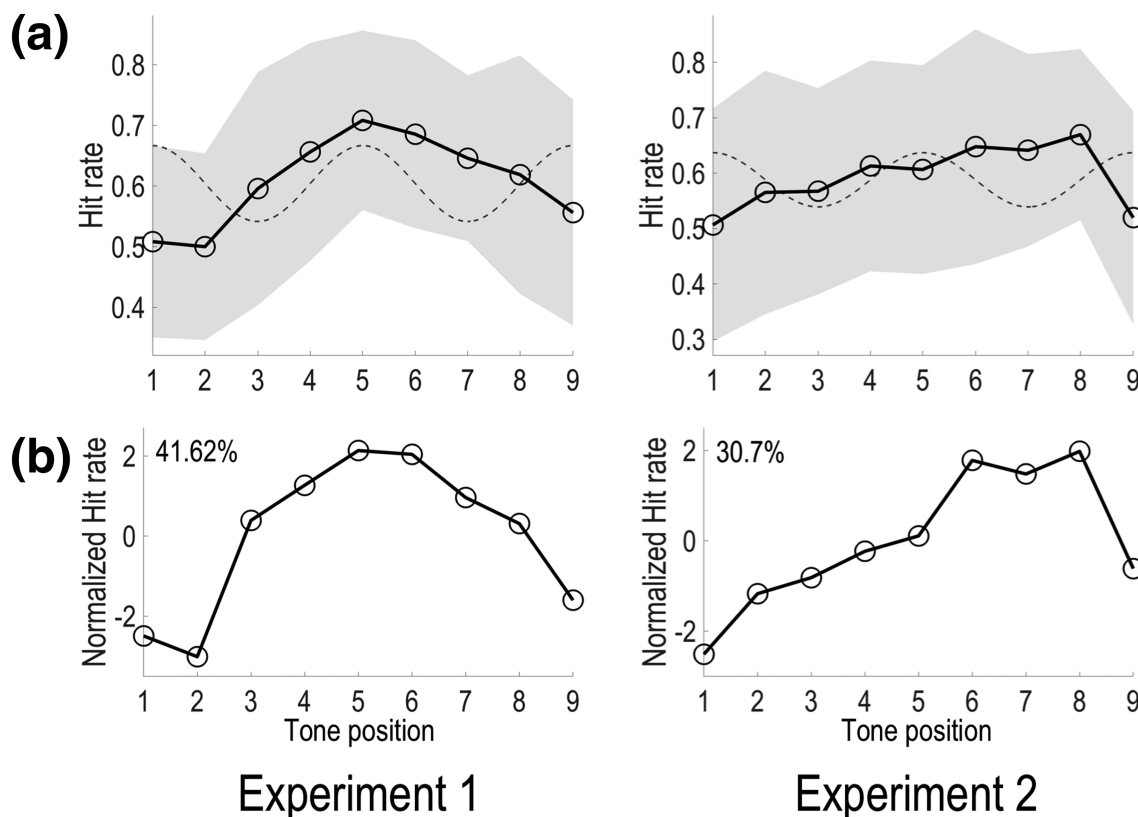
(a)

(b)

Experiment 1   Experiment 2

FIGURE 4 Target detectability as a function of temporal position from the two experiments. (a) Average hit rate across participants. (b) Dominant principal component analysis (PCA) component extracted from participants' hit rate curves. The number in each figure indicates the percentage of the total variance explained by the component

dominant PCA component did not show the bi-cyclic, M-shape, pattern either (Figure 4b). Instead, it also revealed a non-cyclic pattern, where target detectability was low at initial and final positions and peaked around Positions 6–8. In addition, analysis of the normalized 2.67-Hz power revealed that the average modulation strength was not significantly different from zero (*Mean* = −0.20, *SD* = 0.72, $t[22] = -1.32$, $p > .1$). Assessment of individual results showed 8 out of 23 participants with a positive modulation strength (Figure 5a). Finally, the comparison between the two experiments (Experiment 1 vs. Experiment 2) revealed no significant difference in average modulation strength ($t[45] = .21$, $p > .1$).

It is noteworthy that both of our experiments showed larger cross-participant variability in overall hit rate for near-threshold targets than the original study does. We next examined whether the absence of the entrainment effect at the group level in our experiments was due to this larger variability. One might argue that when participants' overall hit rate is too high (clearly above threshold) or too low (clearly below threshold), it reduces the likelihood of observing fluctuations across different temporal positions and hence decreases the sensitivity of the measurement in revealing potential modulation effects from the entraining signal. If there was any truth in this

conjecture, one would expect to observe a clearer presence of the entrainment effect across participants whose overall hit rate is confined within a narrower range, similar to the one reported in the original study. In order to test this hypothesis, we computed a range of interest, using the mean and SD of overall hit rates for near-threshold targets across participants of Hickok et al. (2015). Specifically, this range of interest corresponded to the 95% confidence interval of a normal distribution defined with that mean and SD ([0.47, 0.75]). We then selected, for each of our experiments, participants whose overall hit rate for near-threshold targets was enclosed in this range (Experiment 1: 15 out of 24 participants; Experiment 2: 16 out of 23 participants). In order to further increase the level of comparability in terms of overall performance between our experiments and the original study, we also excluded participants whose overall false alarm rate is higher than 95th percentile of the distribution estimated from the false alarm rate of the five participants of the original study (0.069). This procedure additionally excluded six participants from Experiment 1 and three participants from Experiment 2. We then compared the average modulation strength across the selected participants of each experiment against 0 using t tests. Results from this more restricted

FIGURE 5 Strength and phase of near-3-Hz modulation in hit rate from Hickok et al. (2015) and the two experiments of the current study. For modulation strength (a), filled circles show individual subjects' modulation strength. Blue areas indicate the distribution density of the data. Open circles indicate average across participants. Error bars: standard deviation. For modulation phase (b), filled circles indicate starting phase of 2.67-Hz modulation of participants who presented positive modulation strength. Blue line indicates the average phase across participants, and the blue area on each side of the average phase indicates standard deviation
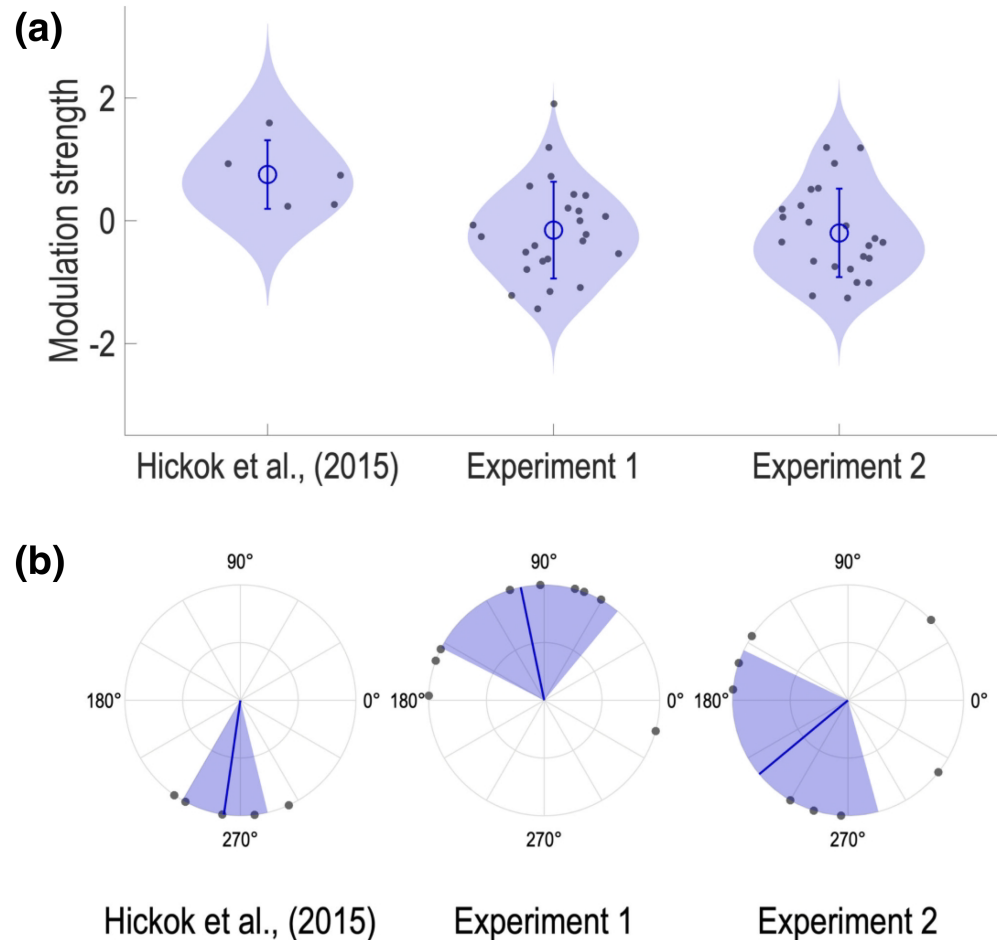
analysis showed that the average modulation strength of selected data is not significantly above 0 (Experiment 1: $Mean = -0.29$; $SD = 0.52$; $t(8) = -1.70$, $p > .1$; Experiment 2: $Mean = -0.21$; $SD = 0.72$; $t(12) = -1.04$, $p > .1$; Figure 6).

Another important finding from Hickok et al. (2015) was the anti-phase relation between target detectability and the pattern of the preceding amplitude modulation of the background noise. Specifically, detection accuracy often reached a local maximum at temporal positions where the phase of preceding amplitude modulation would have approached the minimum. In the original study, this observation was confirmed by a narrow distribution of starting phase of 3 Hz power near $-\pi/2$ ($-90°$ or $270°$), which is the opposite phase of noise modulation ($90°$) at the same time instant (see fig. 4 from Hickok et al., 2015, for details). Here, we first re-examined this phenomenon using the hit rate data from Hickok et al. (2015). For each participant, we computed the phase of 2.67 Hz from the FFT of the hit rate curve. Mean and SD of the 2.67-Hz phase were calculated using the CircStat toolbox in MATLAB (Berens, 2009). Our results confirmed the original observation, with the starting phase of 2.67-Hz modulation being around $270°$
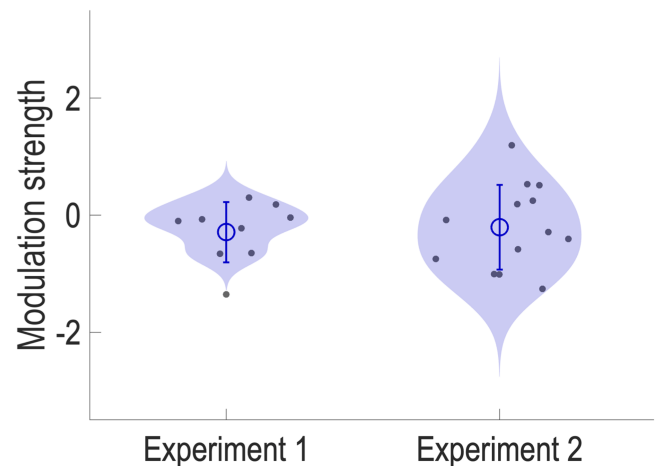

FIGURE 6 Strength of near-3-Hz modulation in hit rate curves of selected participants from the two experiments of the current study. Filled circles show individual subjects' modulation strength. Blue areas indicate the distribution density of individual data. Open circles indicate average across participants. Error bars: standard deviation

in all five participants ($Mean = 261.80°$; $SD = 21.88°$; Figure 5b). Although our own experiments did not provide any evidence for near-3-Hz modulation in target

detectability at the group level, some participants did exhibit positive modulation strength in their results. Thus, we examined the distribution of 2.67-Hz phase in hit rate curves across these participants (Experiment 1: $n = 9$; Experiment 2: $n = 8$). Overall, our results did not show an anti-phase relation between target detectability and noise modulation in either experiment (Experiment 1: *Mean* = 101.71°; *SD* = 51.23°; Experiment 2: *Mean* = 219.80°; *SD* = 65.50°; Figure 5b). In terms of individual results, for Experiment 1, no participant showed a starting phase within a range of ±30° from 270°. For Experiment 2, only three participants showed a starting phase within that range.

3.3 | Additional analysis of participants' accuracy and d-prime data

Due to the arbitrary pairing of non-target trials with experimental conditions, participants' performances in these trials cannot be objectively assigned to any specific temporal position (see Section 2.4.1 for a detailed explanation). We thus did not include non-target trials in our previous analyses. Meanwhile, we deem it necessary to address potential criticism related to the exclusion of non-target trials. We therefore conducted the same analyses on the normalized 2.67-Hz power using two measurements that take into account participants' performance in non-target trials, namely, *accuracy* and *d-prime*.[5] If the labelling of non-target trials with temporal position is indeed fully arbitrary, then the distribution of participants' correction rejection and false alarm responses across the nine temporal positions should be random. Thus, the inclusion of participants' performance from the non-target trials in the analysis, although would introduce random noise in participants' performance across different temporal positions, should not affect the statistical outcome of the entrainment effect observed in participants' hit rate curves at the group level.

Results from these analyses are presented in Figure 7. For both accuracy (Figure 7a) and d-prime (Figure 7b), the average modulation strength across participants was not significantly different from zero in Experiment 1 (accuracy data: *Mean* = −0.17, *SD* = 0.77, $t(23) = −1.07$, $p > .1$; d-prime data: *Mean* = −0.04, *SD* = 0.78, $t(23) = −0.26$, $p > .1$) or Experiment 2 (accuracy data: *Mean* = −0.19, *SD* = 0.85, $t(22) = −1.06$, $p > .1$; d-prime data:

Mean = −0.23, *SD* = 0.60, $t(22) = −1.80$, $p = .09$). This finding supports the non-significance of the entrainment effect at the group level in both of our experiments, even when participants' performances in non-target trials were taken into account (see Figures S3–S6 for visual inspections of accuracy and d-prime curves of individual participants from the two experiments).

3.4 | Non-cyclic modulation of detectability across temporal positions

Although the main objective of the current study was to examine the presence of bi-cyclic modulation patterns in target detectability across temporal positions, results from our PCA analyses revealed the existence of a non-cyclic pattern in participants' performance from both experiments of the current study (Figure 4b). This pattern resembles an inverted U-shape, where target detectability is low at initial and final temporal positions and reaches the maximum around the positions in the middle. These observations invited an examination of the presence of a non-cyclic pattern in the individual data of our participants. For this purpose, we conducted an analysis similar to the one used to investigate the strength of bi-cyclic modulation. Specifically, we calculated the goodness of fit (i.e., R^2) of the hit rate curve of each participant to an inverted U-shape curve that was generated using a generic polynomial function.[6] This goodness of fit indicates the degree of presence of an inverted U-shape pattern in the hit rate curve of each participant. We then normalized the observed goodness of fit with a baseline level, which was computed from a permutated dataset created using the same permutation approach as for the analysis of 2.67-Hz power. Specifically, the baseline goodness of fit consisted of the averaged goodness of fit to the inverted U-shape function across hit rate curves from the permutated dataset, which indicates the average degree of presence of an inverted U-shape pattern that could arise from random fluctuations of participants' detection performance across the nine temporal positions. Finally, we computed the normalized goodness of fit by subtracting the baseline goodness of fit from the one observed in the real hit rate curve of each participant. Our hypothesis was that if participants' target detection across the nine temporal positions was resembling an inverted U-shape and this was not due to chance, then participants' real hit rate curves should show significantly stronger goodness of fit than the baseline level.

---

[5]D-prime, also referred to as *sensitivity index*, is a measure defined in signal detection theory (Stanislaw & Todorov, 1999). It reflects a participant's ability to distinguish between true presence and absence of a signal (or target) in a collection of trials. It is calculated as the difference between the Z score of hit rate and that of false alarm rate: d-prime = Z(hit rate) − Z(false alarm rate).

---

[6]The data and the script used to conduct the analysis can be found in an online repository (https://doi.org/10.17617/3.5c).
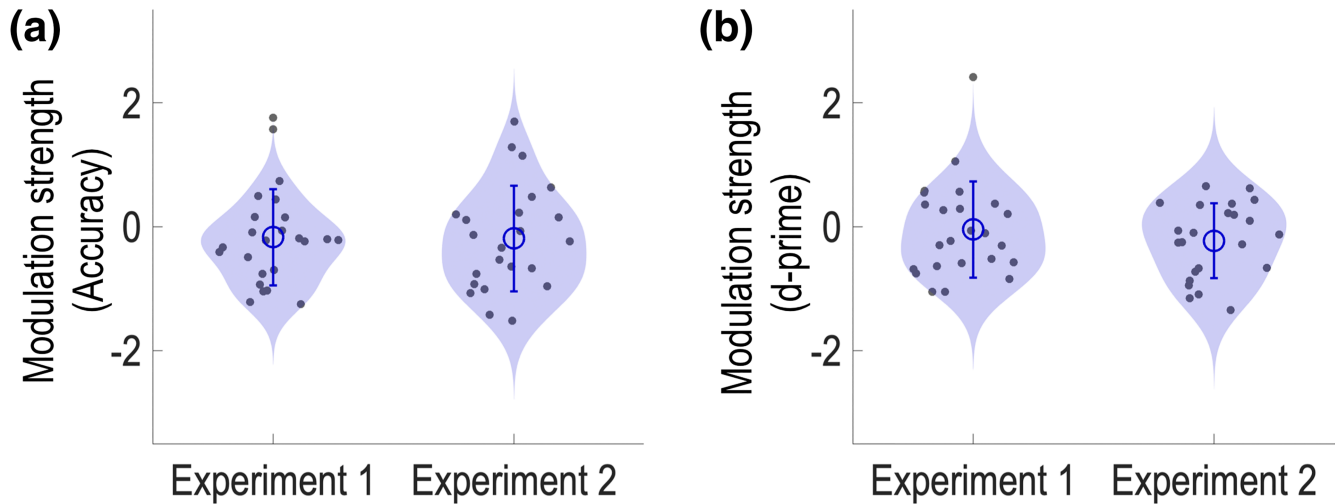
FIGURE 7 Strength of near-3-Hz modulation in accuracy (a) and d-prime (b) curves of participants from the two experiments of the current study. Filled circles show individual subjects' modulation strength. Blue areas indicate the distribution density of individual data. Open circles indicate average across participants. Error bars: standard deviation
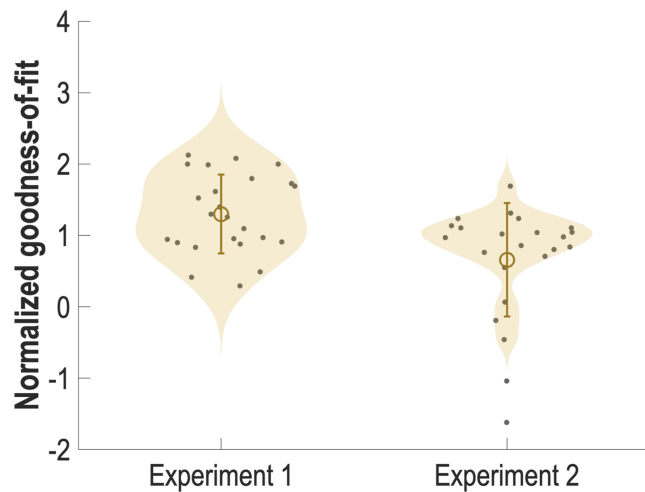

FIGURE 8 Normalized goodness-of-fit of hit rate curves to an inverted U-shape pattern from the two experiments. Filled circles indicate individual fit strength, and yellow areas indicate the distribution density of individual data. Open circles indicate average across participants. Error bars indicate standard deviation

Accordingly, the normalized goodness of fit should be significantly higher than 0 at the group level.

Results of our analysis showed a consistent presence of this non-cyclic pattern in participants' hit rate curves in both experiments (Figure 8). In Experiment 1, our results revealed that the normalized goodness of fit is significantly higher than 0 at the group level ($Mean = 1.30$; $SD = 0.55$; $t(23) = 11.51$; $p < .001$). Specifically, the normalized goodness of fit in every participant is higher than 0. In Experiment 2, our results showed that the normalized goodness of fit was also significantly higher than 0 at the group level ($Mean = 0.66$, $SD = 0.80$, $t(22) = 3.97$, $p < .001$). Assessment of individual results revealed

that the normalized goodness of fit in 19 of the 23 participants is positive. Meanwhile, comparison between the two experiments revealed that normalized goodness of fit in Experiment 2 is significantly lower than those in Experiment 1 ($t[45] = 3.22$; $p < .01$).

4 | DISCUSSION

The study was designed to replicate and extend the findings of Hickok et al. (2015) in order to understand the robustness of the paradigm and investigate the impact of oscillatory entrainment to a rhythmic signal on subsequent auditory processing. Specifically, we examined to what extent the entrainment effect observed in the original study can be reproduced with an increased sample size and with simplifications in the experimental procedure. The results from two new experiments reveal that only a subset of participants (~36%) exhibited the entrainment effect in behavioural performance. Consequently, neither of our experiments showed a significant entrainment effect at the group level.[7] Importantly, for both experiments, the failure to observe a significant entrainment effect at the group level was revealed with

[7]In addition to the strength of 2.67-Hz modulation, we also examined the existence of a potential modulation at 5.33 Hz, which the first harmonic frequency of 2.67 Hz. Analysis of the normalized 5.33-Hz power showed that the modulation strength at 5.33 Hz is not significantly above 0 in either of our experiments (Experiment 1: $Mean = -0.15$, $SD = 0.79$, $t(23) = -0.95$, $p > .1$; Experiment 2: $Mean = -0.20$, $SD = 0.72$, $t(23) = -1.32$, $p > .1$) (Figure S8). Inspection of individual results showed positive modulation strength at 5.33 Hz in only a small subset of participants (7 of 23 in Experiments 1 and 4 of 24 in Experiment 2).

three different measures quantifying participants' performance across the nine temporal positions: hit rate, accuracy and d-prime. Second, further analysis focusing on subset of participants who did show the entrainment effect provided little evidence for a systematic anti-phase alignment between temporal fluctuations of target detectability and the pattern of the preceding entraining modulation. The current study does not demonstrate a consistent impact of a preceding rhythmically modulated background noise on the detection of upcoming near-threshold tone targets. It highlights substantial variability across subjects. The variable or absent entrainment effect held true in results from both the conceptual replication (Experiment 1) and the exact replication (Experiment 2).

In trying to understand the factors that underlie why our effects differ from Hickok et al. (2015), we first address methodological differences between the studies (which mostly concern modifications introduced in Experiment 1). First, in Experiment 1, we used individually measured detection thresholds to determine the SNR level of near-threshold tone targets for each participant, which contrasts with the use of a common near-threshold SNR level for all the participants in the original study. Despite this methodological difference, the average hit rate for near-threshold targets from Experiment 1 was at the same level as (i) that from the original study and (ii) our own exact replication (Experiment 2). Participants from both of our experiments also exhibited false alarm rates similar to the original study, which further confirmed a good level of comparability in terms of participants' overall performance between our experiments and the original study. Although the overall hit rates and false alarm rates from both of our experiments exhibited larger variability across participants than in the original study, our restricted analysis showed that this enlarged variability in participants' overall performance cannot account for the absence of the entrainment effect at the group level.

Second, in Experiment 1, we presented tone targets at two SNR levels instead of five in the original study. This modification not only shortened the duration of the experiment but also reduced the degree of uncertainty regarding the intensity of the targets. One might argue that this design difference potentially contributes to the disappearance of the entrainment effect in the experiment. However, that speculation is not supported empirically, as the results from Experiment 1 on modulation strength did not differ from those from Experiment 2, with the latter being an exact replication of the original study. In summary, the robust similarity between the results across our two experiments makes it unlikely that the methodological modifications in Experiment 1 played a major role to the absence of the entrainment effect at the group level.

Third, arguably the biggest difference between the current study and Hickok et al. (2015) concerns the sample size, with each of our experiments having data from four times more participants than the original study. Results from both of our experiments showed substantial cross-participant variability in the strength of near-3-Hz modulation compared to the level observed in the original study. Importantly, unlike the original study that showed a ubiquitous presence of positive modulation for all five participants, both of our experiments revealed a large proportion of participants who exhibited negative modulation strengths, which indicates the absence of a bi-cyclic pattern in hit rates. Consequently, our results reveal a decrease of average modulation strength from the level observed the original study to near 0. Although larger cross-participant variability could be expected with an increased sample size, the disappearance of the entrainment effect at the group level in both our experiments is surprising and could have multiple explanations. On the one hand, the overall higher modulation strength from Hickok et al. (2015) could result from certain contingent details of the experimental environment of the original study that implicitly facilitated the observation of entrainment effect across the tested participants. On this interpretation, one might expect stability of average modulation strength at a positive level, if more participants are tested in the same experimental environment despite a potential overall increase of cross-participant variability. On the other hand, it is also possible that the overall positive modulation strength observed in the study is due to an accidental underrepresentation of participants within the pool of tested participants who do not exhibit the entrainment effect. According to this interpretation, one should expect a decrease of the overall modulation strength at the group level, if more participants are tested in the same experimental environment.

Next, one characteristic of the original study is the consistent anti-phase alignment between the fluctuation of target detectability and the amplitude modulation in the preceding noise across participants. Our results did not reveal an anti-phase relation between participants' performance and noise modulation. In all participants who exhibited a positive entrainment effect, we observed sizeable variations in phase alignment within each experiment as well as between the two experiments.

Variable phase alignment between the rhythmic signal and perceptual facilitation could be due to multiple mechanisms. One possibility is that, given the continuous nature of acoustic modulation, the auditory system of different participants could phase lock to different parts of the entraining signal, which consequently induces perceptual facilitation at variable temporal

positions. Although this explanation may account for the widely distributed phase alignment between rhythmic stimulation and performance facilitation when participants were entrained to frequency-modulated noise (Henry et al., 2014; Henry & Obleser, 2012), it is most likely not the case for the entrainment to amplitude modulation. Indeed, existing neurobiological findings have demonstrated converging phase alignment of slow cortical oscillations across participants in response to the amplitude envelope of auditory input (Doelling et al., 2019; Forseth et al., 2020; Simon & Wallace, 2017). Moreover, in a recent study using electrocorticography (ECoG) (Oganian & Chang, 2019), the authors found that auditory neuronal populations in the superior temporal gyrus (STG) exhibit maximum firing rate during the rising portion of the amplitude envelope of the acoustic input.

As suggested by the above-mentioned neurobiological studies, consistent neural responses to specific amplitude changes in the acoustic input should reduce the chance for jittered phase alignment. However, it is not straightforward to assume that a converging phase correspondence between the input signal and auditory neural activity should definitely induce better target detection at the same temporal position for every participant. A crucial question, which these previous investigations did not address, is to what extent the temporally aligned neural excitability would specifically facilitate sensory processing of the target stimulus. For instance, Bauer et al. (2015) addressed this question in explaining the lack of consistent perceptual facilitation for in-phase targets in their study. Note that, in that study, participants were asked to perform a task on the pitch of a target tone that was preceded by a sequence regularly paced distractor tones with a different pitch. They argued that, although attending to the preceding rhythmic signal should induce stronger temporal expectation for target tones that occurred in-phase with the preceding rhythm, it could be the case that a strong association could be instantiated between in-phase positions and the pitch value of the distractor tones and consequently hinder the perception of target tones that occur at these positions. Therefore, the conflicting benefits of temporal and spectral information carried by the entraining signal could contribute to the lack of consistent observation of behavioural facilitation of in-phase targets. As for the current paradigm, note that there is no direct correspondence between the spectral content of the entraining signal (broadband noise) and the target stimulus (1-kHz pure tone). Therefore, it is not clear to what extent a specific part of amplitude envelope of a broadband noise would trigger neural responses that specifically facilitate the detection of a 1-kHz tone. Recent neurobiological studies demonstrated that neurons in primary auditory cortex phase lock to rhythmic acoustic stimulation in a frequency specific manner (Lakatos et al., 2013; O'Connell et al., 2011). That is, when being stimulated with a sequence of pure tones, A1 neurons align their high excitability phase to the input signal only if the frequency of pure tones correspond to the preferred frequency of these neurons. These findings suggest that one could increase the sensitivity of the paradigm used in Hickok et al. (2015) in order to observe an entrainment effect by applying a stronger spectral correspondence between the entrainment signal and the target stimulus.

Finally, despite the failure to find a consistent bicyclic pattern in participants' performance, our results based on the PCA analysis reveal a non-cyclic pattern that is present in both of our experiments. This pattern is characterized by an increase of target detectability from the initial temporal positions towards the ones in the middle, which is followed by a decrease of target detectability towards the final temporal positions. Although the extracted pattern of the two experiments differ with regard to the exact location where detectability reaches the maximum, both patterns exhibit lower detectability at positions near the borders of the target window. Our analysis of the normalized goodness of fit to an inverted U-shape function shows more consistent presence of this non-cyclic pattern across our participants than of a bicyclic pattern. Interestingly, the hit rate curves of two participants from Hickok et al. (2015) also resemble the form of this non-cyclic pattern. The presence of this pattern in all three datasets invites speculations regarding an alternative behavioural phenomenon induced by the paradigm. In fact, a previous study, in which target stimuli were uniformly distributed within a time range, also found better target detection towards the middle of the time range than that towards the borders (Ng et al., 2012).

One possible explanation for this phenomenon could be that participants, after noticing that the target could occur across a certain time range, preferentially focused their attention on the centre of the time range. This performance pattern could result from a form of statistical learning of the temporal distribution of the target stimulus based on local temporal cues, sometimes known as a foreperiod effect (see Hoehl et al., 2021, for a review). A foreperiod is referred to as the time interval between the target stimulus and the final stimulus of a preceding entraining sequence. In the current study, one could consider the beginning of the unmodulated noise as a reference stimulus. Then, each temporal position for the target stimulus presents a specific foreperiod with respect to the reference stimulus, with the shortest foreperiod for

the first position and the longest for the last position. Previous investigations have shown that when various foreperiods were mixed within an experimental block, target stimuli with different foreperiods would be processed with different efficiency (e.g., Ellis & Jones, 2010; Schirmer et al., 2021). In particular, it has been shown that participants were better at perceiving stimuli presented with the middle range foreperiod compared to those presented with foreperiods at the shortest and the longest ends (Ellis & Jones, 2010). Our finding is in line with these observations.

Given that the observed non-cyclic pattern manifests across the full time range for target occurrence, it should arise independently from the rhythmicity presented in the preceding stimulation. Meanwhile, studies have also shown that the shape of the performance pattern due to the foreperiod effect can be modulated by the presence of rhythmicity in a preceding entraining sequence (Ellis & Jones, 2010; Schirmer et al., 2021). Although it is uncontroversial to assume that the human brain integrates temporal information from different sources when predicting the timing of upcoming sensory stimuli (Hoehl et al., 2021), our findings suggests that the current paradigm allows more robust inferences about the impact of the foreperiod effect than that of the entrainment effect on the detection of tonal targets.

5 | CONCLUSION

The primary goal of the current study was to replicate the entrainment effect observed in Hickok et al. (2015). Our findings raise questions about the robustness of the paradigm used in the original study, in revealing the impact of oscillatory entrainment on subsequent auditory processing. Specifically, based on findings from two experiments with increased sample sizes, no significant bicyclic modulation in target detection was observed at the group level. We suggest that the real effect size is substantially lower than originally estimated. Interestingly, while existing neurobiological studies have provided converging evidence for brain's ability to entrain to external rhythmicity (e.g., Lakatos et al., 2013), the lack of consistency in the demonstration of an entrainment effect at the behavioural level raises an issue of generalizability of the entrainment phenomenon across different cognitive levels. Therefore, designing and testing behavioural paradigms that allow for robust evidence of the entrainment effect at the behavioural level would be an essential requirement for understanding the complicated pathway from regulated fluctuations of local neural activities to potential modulations of efficiency in perceiving auditory events.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS
Y. Sun and D. Poeppel designed research. Y. Sun collected experimental data. Y. Sun and G. Michalareas performed data analyses. Y. Sun wrote the initial draft. Y. Sun, G. Michalareas and D. Poeppel edited the final draft.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15367.

DATA AVAILABILITY STATEMENT
The raw data for all experiments and scripts for data analysis reported in this paper are available at the Open Access Data Repository of the Max Planck Society (https://doi.org/10.17617/3.5c).

ORCID
Yue Sun 🅾 https://orcid.org/0000-0001-6798-636X
Georgios Michalareas 🅾 https://orcid.org/0000-0002-8196-3049
David Poeppel 🅾 https://orcid.org/0000-0003-0184-163X

REFERENCES

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, *25*(9), 3077–3085. https://doi.org/10.1093/cercor/bhu103

Assaneo, M. F., Rimmele, J. M., Perl, Y. S., & Poeppel, D. (2021). Speaking rhythmically can shape hearing. *Nature Human Behaviour*, *5*, 71–82.

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, 22(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, 41(3), 254–311. https://doi.org/10.1006/cogp.2000.0738

Bauer, A. K. R., Jaeger, M., Thorne, J. D., Bendixen, A., & Debener, S. (2015). The auditory dynamic attending theory revisited: A closer look at the pitch comparison task. *Brain Research*, 1626, 198–210. https://doi.org/10.1016/j.brainres.2015.04.032

Berens, P. (2009). CircStat: A Matlab toolbox for circular statistics. *Journal of Statistical Software*, 31(10), 1–21. https://doi.org/10.18637/jss.v031.i10

Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, 33(8), 955–967. https://doi.org/10.1080/23273798.2018.1439179

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

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(June), 1926–1930. https://doi.org/10.1126/science.1099745

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences of the United States of America*, 116(20), 10113–10121. https://doi.org/10.1073/pnas.1816414116

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics*, 72(8), 2274–2288. https://doi.org/10.3758/BF03196701

Farahbod, H., Saberi, K., & Hickok, G. (2020). The rhythm of attention: Perceptual modulation via rhythmic entrainment is lowpass and attention mediated. *Attention, Perception, & Psychophysics*, 82(7), 3558–3570. https://doi.org/10.3758/s13414-020-02095-y

Forseth, K. J., Hickok, G., Rollo, P. S., & Tandon, N. (2020). Language prediction mechanisms in human auditory cortex. *Nature Communications*, 11(1), 1–14. https://doi.org/10.1038/s41467-020-19010-6

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. https://doi.org/10.1038/nn.3063

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, 86(July 2017), 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

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

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 109(49), 2009–2100. https://doi.org/10.1073/pnas.1213390109

Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science*, 26(7), 1006–1013. https://doi.org/10.1177/0956797615576533

Hoehl, S., Fairhurst, M., & Schirmer, A. (2021). Interactional synchrony: Signals, mechanisms and benefits. *Social Cognitive and Affective Neuroscience*, 16(1–2), 5–18. https://doi.org/10.1093/scan/nsaa024

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, 104(2007), 2500–2511. https://doi.org/10.1152/jn.00251.2010

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology*, 53(1), 59–96. https://doi.org/10.1016/j.cogpsych.2006.01.003

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

Kayser, C., Montemurro, M. A., Logothetis, N. K., & Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 61(4), 597–608. https://doi.org/10.1016/j.neuron.2009.01.008

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16), 2729–2737. https://doi.org/10.1016/S0042-6989(98)00285-5

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28(18), 2867, e3–2875. https://doi.org/10.1016/j.cub.2018.07.023

Lakatos, P., Gross, J., & Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. *Current Biology*, 29(18), R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

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

Lin, W. M., Oetringer, D. A., Bakker-Marshall, I., Emmerzaal, J., Wilsch, A., ElShafei, H. A., Rassi, E., & Haegens, S. (2021). No behavioural evidence for rhythmic facilitation of perceptual discrimination. *European Journal of Neuroscience*, 1–13. https://doi.org/10.1111/ejn.15208

Ng, B. S. W., Schroeder, T., & Kayser, C. (2012). A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *Journal of Neuroscience*, *32*(35), 12268–12276. https://doi.org/10.1523/JNEUROSCI.1877-12.2012

O'Connell, M. N., Falchier, A., McGinnis, T., Schroeder, C. E., & Lakatos, P. (2011). Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. *Neuron*, *69*(4), 805–817. https://doi.org/10.1016/j.neuron.2011.01.012

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5*(11), eaay6279.

Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what "speech" is: A hierarchical perspective. *Frontiers in Human Neuroscience*, *6*(NOVEMBER 2012), 1–4. https://doi.org/10.3389/fnhum.2012.00309

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Poeppel, D., & Teng, X. (2020). Entrainment in human auditory cortex: Mechanism and functions. In B. Fritzsch & B. Grothe (Eds.), *The senses: A comprehensive reference* (2nd ed., Vol. 2, pp. 63–76). Elsevier, Academic Press.

Prins, N., & Kingdom, F. (2009). Palamedes: Matlab routines for analyzing psychophysical data. Retrieved from http://www.palamedestoolbox.org

Saberi, K., Farahbod, H., & Hickok, G. (2020). Data from: The rhythm of attention: Perceptual modulation via rhythmic entrainment is lowpass and attention mediated, Dryad, Dataset. https://doi.org/10.7280/D1WD5D

Schirmer, A., Wijaya, M., Chiu, M. H., Maess, B., & Gunter, T. C. (2021). Musical rhythm effects on visual attention are non-rhythmical: Evidence against metrical entrainment. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 58–71. https://doi.org/10.1093/scan/nsaa077

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

Simon, D. M., & Wallace, M. T. (2017). Rhythmic modulation of entrained auditory oscillations by visual inputs. *Brain Topography*, *30*(5), 565–578. https://doi.org/10.1007/s10548-017-0560-4

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *31*(1), 137–149. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10495845

Thut, G., Schyns, P. G., & Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Frontiers in Psychology*, *2*(JUL), 1–10. https://doi.org/10.3389/fpsyg.2011.00170

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.