Slow phase-locked modulations support selective attention to sound

Magdalena Kachlicka [a], Aeron Laffere [a], Fred Dick [a,b], Adam Tierney [a,*]

[a] Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, England
[b] Division of Psychology & Language Sciences, UCL, Gower Street, London WC1E 6BT, England

ARTICLE INFO

ABSTRACT

To make sense of complex soundscapes, listeners must select and attend to task-relevant streams while ignoring uninformative sounds. One possible neural mechanism underlying this process is alignment of endogenous oscillations with the temporal structure of the target sound stream. Such a mechanism has been suggested to mediate attentional modulation of neural phase-locking to the rhythms of attended sounds. However, such modulations are compatible with an alternate framework, where attention acts as a filter that enhances exogenously-driven neural auditory responses. Here we attempted to test several predictions arising from the oscillatory account by playing two tone streams varying across conditions in tone duration and presentation rate; participants attended to one stream or listened passively. Attentional modulation of the evoked waveform was roughly sinusoidal and scaled with rate, while the passive response did not. However, there was only limited evidence for continuation of modulations through the silence between sequences. These results suggest that attentionally-driven changes in phase alignment reflect synchronization of slow endogenous activity with the temporal structure of attended stimuli.

Introduction

As we navigate the world, we are bombarded by more sensory information than we can process and respond to. One of the brain's most important tasks, therefore, is to focus on the most relevant information while ignoring the rest. Whereas the eye can be directed to point at and focus on regions of visual space, this mechanical strategy is generally not available to the human auditory system. Instead, it must use neural mechanisms to segregate acoustic information into sources or 'streams', select one stream, and then actively attend to that stream to extract task-relevant information (Shinn-Cunningham, 2008; Holt et al., 2018).

What neural strategies might the listener use to focus in on and track a task-relevant stream? One possibility is to filter auditory streams based on acoustic dimensions along which they differ. For example, if one stream lies in a higher frequency band than the other, listeners could selectively attend to that stream by directing attention to a particular frequency band (spectrally-selective attention; Paltoglou et al., 2009; Fritz et al., 2012; Da Costa et al., 2013; Dick et al., 2017; Riecke et al., 2017). Another possible strategy is to take advantage of timing or rhythmic differences between streams, so that attention can be directed to time points likely to contain the to-be-attended stream and less likely to contain the to-be-ignored stream (Nobre and van Ede, 2018).

That listeners make use of temporally-selective attention to select sound streams for further processing is supported by findings that temporal differences between a target stream and a distractor stream help boost performance on attention tasks. When target stimuli are presented among distractors, prior knowledge about target timing onsets has been shown to boost tone detection (Bonino and Leibold, 2008), birdsong identification (Best et al., 2007), word recognition (Gatehouse and Akeroyd, 2008), and speech comprehension (Kitterick et al., 2010). Moreover, when speaking, talkers modulate the temporal characteristics of speech to minimize overlap with background speech, thus allowing listeners to use such a temporally-selective strategy to understand what the speaker is saying (Cooke and Lu, 2010).

How might the brain carry out such temporally-selective attention? One possibility is that brain activity "entrains" to the attended stream, such that endogenous neural oscillations phase-lock to its temporal structure (Schroeder and Lakatos, 2009).[1] Indeed, when non-human primates attend to one stream in an inter-modal attention task, slow neural activity aligns with the temporal structure of the attended stream (Lakatos et al., 2008). Similarly, in humans, attention to alternating auditory versus visual stimuli is linked to a shift in phase of 180° in a slow modulation of neural activity at the within-modality presentation rate (Besle et al., 2011).

This endogenous oscillation hypothesis has influenced a large body of cognitive neuroscience research on selective attention. However, most of this research has not directly examined the exact form of the

---

[1] In recent years the term "neural entrainment" has been sometimes used more broadly to refer to any regular temporal relationship between the rhythmic structure of a stimulus and a neural response, regardless of whether this relationship reflects the influence of endogenous oscillations; see below for a discussion.
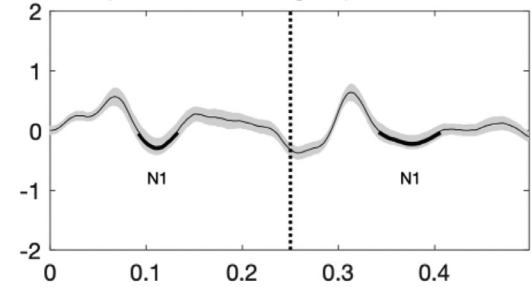
attentional modulation of neural activity. Instead, most studies on both non-human animals and human participants rely on less direct measures of neural entrainment, such as inter-trial phase-locking. For example, Lakatos et al. (2013) and Lakatos et al. (2016) showed that when rhesus macaques attend to one of two tone streams presented at different rates, phase-locking at the attended rate increases. Similarly, when human participants attend to one of two temporally interdigitated tone streams, there is a roughly 180-degree shift in the phase of the neural response at the tone stream presentation rate (Laffere et al., 2020a, b). Studies of selective attention to speech have also shown that low-frequency neural activity more closely mirrors the slow amplitude modulation patterns of the attended stimulus stream, compared to the ignored stream (Kerlin et al., 2010; Ding and Simon, 2013; Zion Golumbic et al., 2013; O'Sullivan et al., 2015; Ghinst et al., 2016). This phenomenon is sometimes interpreted according to the endogenous oscillation framework, i.e. as reflecting entrainment of cortical oscillations to the temporal structure of attended speech (Horton et al., 2013; Riecke et al., 2018; Fuglsang et al., 2020; for a discussion, see Obleser and Kayser, 2019).
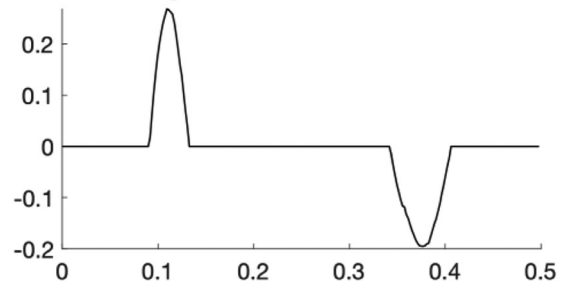
Such research has shown that attention modulates the strength of the relationship between neural activity and the temporal structure of the attended signal. However, this entrainment in a broad sense—the presence of stimulus-brain temporal alignment—is not strong evidence for entrainment in a narrow sense—phase-locking of ongoing oscillations. As defined by Obleser and Kayser 2019, a more stringent test of entrainment in a narrow sense is to show that the phase of endogenous, ongoing oscillators is adjusted, such that peaks align with timepoints containing acoustic edges in attended stimuli. With the exception of Lakatos et al., 2008 and Besle et al., 2011, both of whom showed evidence for the existence of slow neural rhythms aligned with the temporal structure of attended stimulus streams, the remainder of these findings are consistent with an alternate explanation – that attention acts as a filter attenuating and amplifying exogenous responses to sound. Attention-driven increases in phase-locking or shifts in neural phase at the rate of stimulus presentation could potentially be generated by enhancement of the magnitude of ERPs to the onsets of sounds in attended streams, with responses to sounds in unattended streams either unaltered or attenuated. Indeed, a large body of research has shown that selective attention to a sound stream can increase the amplitude of exogenously-evoked potentials to sound onsets within the target stream (Hillyard et al., 1973; Chait et al., 2010; Choi et al., 2013; Dai et al., 2018), including the P50 and N100 components (Woldorff et al., 1993). These enhanced responses would then lead to the increased phase-locking commonly reported as a neural correlate of auditory selective attention.

In sum, much of the existing research on auditory selective attention is consistent with two competing explanations: one positing selective attention as a filter that amplifies or attenuates exogenous responses to stimuli, and the other suggesting that endogenous oscillators synchronize to the temporal structure of attended stimuli. In the study reported below, we designed an EEG experiment to potentially adjudicate between these two explanations. Two groups of participants were asked to attend to one of two streams composed of temporally interleaved tone sequences. Across conditions, we manipulated tone presentation rate; across groups, we varied tone duration. By examining the shape of the attentional modulation of the ERP waveform, we can test predictions linked to the neural entrainment versus attentional filter accounts. First, if the 'attention as neural filter' account holds, and attention modulates the gain of event-related responses to sound onset, then the width of the attentional modulation should be limited to the time window of the N100 – or, if it extends to time points containing adjacent positive components, the modulation's polarity should flip (changing from negative to positive). On the other hand, the neural entrainment account suggests that the attentional modulation should extend over a greater length of time, especially at slower rates. The contrasting predictions of the attentional filter and neural entrainment accounts are illustrated in Fig. 1, which shows how the response waveform in the 2 Hz condition would be modulated by 1) attentional enhancement of the N100,
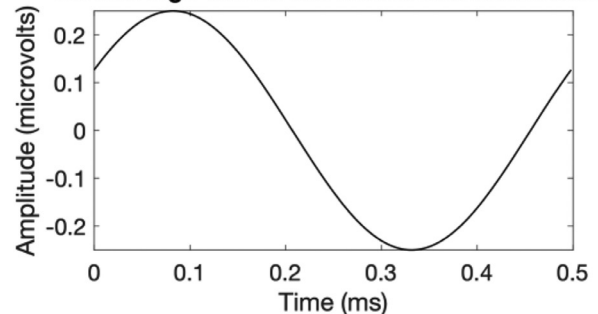

Fig. 1. Contrasting predictions of the attentional filter and neural entrainment accounts of auditory selective attention for the shape of the attentional modulation waveform. Top: Passive response to low/high tone pairs presented at a rate of 2 Hz. N1 responses are indicated via text label and bold outline. Middle: expected modulation waveform (attend high versus attend low conditions) if selective attention to a frequency band enhances N1 responses in the attended band. Bottom: expected modulation waveform if selective attention is carried out via alignment of endogenous neural rhythms with the attended tones. Note that the neural entrainment account predicts that attentional waveform modulation will begin before the onset of the N1.

as posited by the attentional filter account, and 2) alignment of an oscillation with the attended stream, as posited by the neural entrainment account.

The endogenous entrainment account also predicts that attentional modulations could continue through the silence between tone sequences ("forward entrainment"; Saberi and Hickok, 2021), as some oscillatory systems are self-sustaining. By contrast, the attentional filter account predicts that attentional modulations should die out rapidly after the offset of the final tone of the sequence. However, sustained oscillation in silence is not a characteristic of all oscillatory systems, and so while finding a lack of endogenous entrainment would place constraints on the characteristics of endogenous models of temporally-selective attention, it would not be strong evidence against a role for neural oscillations (Doelling & Assaneo, 2021).

To ensure that any increases in the duration of the modulated portions of the waveform at slower rates are not simply driven by increases in tone duration, we included a between-subjects manipulation, such that for half of the participants the tones scaled in duration with rate,

while for the remaining participants, the tones were always 40 ms in duration. Endogenous entrainment accounts posit that neural rhythms align with sudden stimulus changes or acoustic "edges" (Doelling et al., 2014), and so these accounts would predict that attentional modulations should align with the onset of the tones and so be relatively unaffected by manipulations of tone duration.

A second goal of this project was to examine the relationship between presentation rate and the effects of attention on neural phase-locking to stimulus temporal structure. The motor system has been proposed to generate temporal predictions that increase neural sensitivity to upcoming onsets in attended sound streams (Morillon and Schroeder, 2015). That the motor system is involved in temporally-selective attention is supported by the finding that individuals with more consistent self-paced tapping show more robust effects of attention on neural phase-locking to attended sound streams (Laffere et al. 2020b). If temporally-selective attention does rely on implicit motor planning, then phase-locking to attended sound streams may be more robust for slower rates, as the ability to align movements with the temporal structure of sound is rate-limited (Repp 2003). To investigate this possibility, we compared degree of phase-locking to the attended band and the difference in phase between the attend high and attend low conditions across rates, predicting that the phase of neural activity would be better aligned with attended streams at slower rates.

Materials and methods

Participants

Participants were postgraduate students enrolled in courses related to auditory processing (audiology, auditory neuroscience and acoustics) or professional musicians or audio engineers living in London. Participants were selected for higher levels of auditory experience and expertise because the experiment was quite demanding of auditory attention, and in previous studies, non-expert participants required somewhat more training to achieve good performance in the task (Laffere et al., 2020a). 15 participants (aged 19–48, $M = 29.47$, SD=7.41; 9 females) took part in the first experiment (with variable tone durations across presentation rates). 14 different participants (aged 22–36, $M = 28.36$, SD=4.89; 9 females) took part in the second experiment (with a single fixed short tone duration across presentation rates). All participants reported no prior diagnosis of hearing impairment or neurological disorders affecting hearing. The experimental paradigm was approved by the Research Ethics Committee of the Department of Psychological Sciences at Birkbeck, University of London. Processed data are available at https://osf.io/mds9q/.

Experimental and stimulus design and presentation

In both experiments, participants listened to series of interleaved tone sequences segregated into high and low frequency bands (see Fig. 2). Both experiments used a fully within-subject design, crossing attention condition (attend-high band sequences, attend-low band sequences, listen passively) with within-band tone presentation rate (2 Hz, 3 Hz, 4 Hz, 5 Hz). (Our use of 2 Hz as the slowest condition was motivated by time considerations; the experiment lasted approximately two hours, and slower rates (such as 1 Hz) would have required an infeasible amount of extra time, relative to the other conditions.) Each of the 12 combinations of attention condition by tone presentation rate was completed by each participant, with condition order randomized across participants.

The basic stimulus units were cosine-ramped pure tones constructed at a 48 kHz sampling rate using MATLAB (The MathWorks, Inc., Natick, MA). These tones were arranged into sequences of six, followed by a silent period; tone and silence duration varied with experiment and condition, as explained below. Stimuli were presented diotically through Etymotic 3A insert earphones (Etymotic, Elk Grove Village, IL) at 80 dB

sound pressure level. The ramp duration was equal to 1/5 of the total tone duration.

The tones were presented in two frequency bands. For each band, a set of three possible fundamental frequencies was randomly sampled to create mini-sequences of three tones (185, 207.7, and 233.1 Hz for the low band and 370, 415.3, and 466.2 Hz for the high band). These tones were presented 180° out of phase with a within-band rate of presentation varying between blocks of 2 Hz, 3 Hz, 4 Hz, or 5 Hz (i.e. with 500 ms, 333 ms, 250 ms or 200 ms between tone onsets respectively). Tones in low (A) and high (B) frequency bands alternated, forming a repeating ABABAB pattern followed by a silence equal in duration to one cycle of the within-band rate (i.e. 500 ms, 333 ms, etc.). The ABABAB pattern plus silence composed one trial. (Note that the first tone of each sequence was always a low tone; this ensured that time of onset of sequences in each frequency band was always predictable, facilitating stream segregation and selection). Each block consisted of 30 trials; 3–6 sequence repetitions were spaced semi-randomly within each frequency band, with the constraint that there was at least one non-repeating sequence between the repetitions. Finally, for each of the two experiments, there was a single run for each condition, with seven 30-trial blocks per run and a total of 210 trials per condition.

The Variable Tone Length and Fixed Tone Length Experiments differed in the length of the tones used to construct the sequences. In the Variable Tone Length Experiment, tone length differed across the presentation rate conditions, such that, within each condition, tone duration was always equal to half of one cycle of the within-band presentation rate. This equaled 250 ms for 2 Hz, 166.66 ms for 3 Hz, 125 ms for 4 Hz, and 100 ms for 5 Hz. In the Fixed Tone Length Experiment, tones were always 40 ms in length, and this duration did not differ across the presentation rate conditions.

Behavioral task

In the two 'Active' conditions, participants were asked to attend to tone sequences in the low- or high-frequency band, while ignoring the tones presented in the competing frequency band. The participants' task was to identify the repetition of a mini-sequence in the attended frequency band and respond by clicking the mouse. In other words, they were asked to respond whenever the mini-sequence they were currently hearing was identical to the previous mini-sequence (a one-back memory task). In a third 'Passive' condition, participants were asked to sit quietly and listen to the tones without attending to either band, pressing a button at the end of each block to advance to the next block.

For the Active conditions, the latency window for recording behavioral responses to the target began at the onset of the third tone of the second presentation of the repeated sequence (the first time point at which the participant could theoretically detect the repetition). The target window extended for 1.5 s across conditions. Text feedback was displayed for correct answers, missed targets, and false alarms. Performance was measured as D-prime, with the false alarm rate calculated using the total number of non-target sequences as the highest possible number of false alarms. A repeated measures ANOVA with one within-subjects factor (Rate: 2, 3, 4, and 5 Hz) was used to investigate whether D-prime differed across rates, with the Greenhouse-Geisser correction applied due to violation of sphericity.

EEG data acquisition

Electrophysiological data were recorded using a BioSemi ActiveTwo 32-channel EEG system at a 16,384 Hz sample rate and with open filters in Acti-View (BioSemi) acquisition software. A standard 10/20 montage of active electrodes positioned in a fitted head cap with a sintered Ag-AgCl pallet was used. Two additional electrodes were placed on the left and right earlobes as external reference points. Contact impedance was kept below 20 kΩ throughout the testing session.
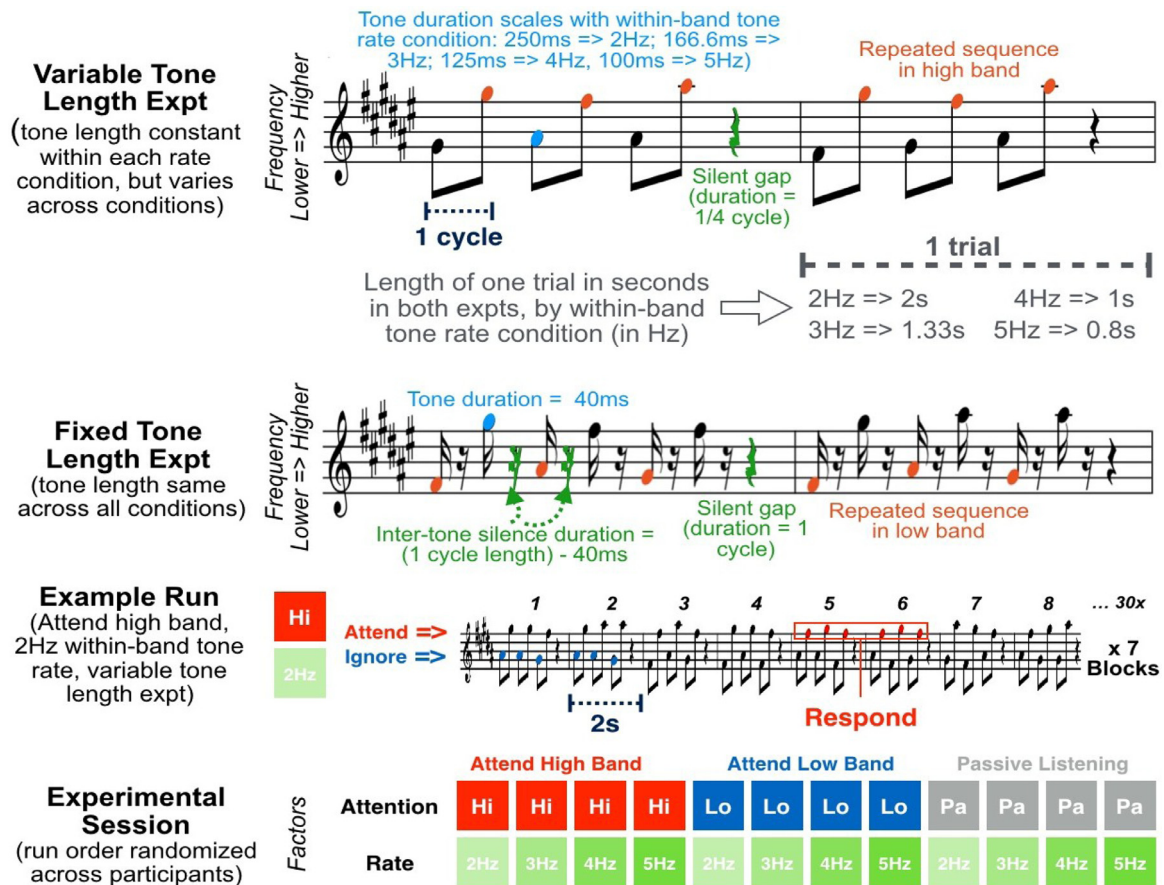
Fig. 2. Schematics showing experimental design. The top row shows two trials of the Variable Tone Length Experiment, and the second row two trials of the Fixed Tone Length Experiment, in quasi-musical notation. For ease of depiction on one stave, fundamental frequencies depicted by musical notes are one octave higher than those used in the experiment. For each tone rate condition in the variable tone length experiment, all tones are set to a condition-specific duration (see Figure and Methods); in the fixed tone length experiment, all tones are 40 ms long regardless of rate condition. The third row shows a segment of an example run of the 'attend high band, 2 Hz within-band tone presentation rate' condition from the variable tone length experiment. Each run is made up of 7 blocks of 30 trials, where a trial contains two interleaved 3-tone sequences, one in the higher frequency band, and one in the lower band. There are three low/high-tone 'cycles' in a trial, followed by a silence equal in duration to an additional cycle. The bottom row shows all twelve runs for a single experiment, where each run is a combination of attention condition and within-band tone presentation rate. Note that run order is randomized across participants. See supplementary material for audio of example blocks.

EEG data processing

Event markers for the beginning of each block were recorded from trigger pulses sent to the neural data collection computer. The resulting data were downsampled to 500 Hz and eye blinks and muscle contraction artifacts were identified by independent component analysis (Hyvärinen and Oja, 2000; Vigário et al., 2000) and removed after visual inspection of component topographies and time courses. Data were segmented into epochs aligned with trial onsets, with epoch duration equaling 2.00, 1.33, 1.00, and 0.80 s for the 2, 3, 4, and 5 Hz conditions, respectively. Then, individual segments were excluded if the signal intensity of the channel exceeded $\pm$ 100 µV. All preprocessing steps were conducted with the use of custom MATLAB scripts and the FieldTrip M/EEG analysis toolbox (Oostenveld et al., 2011).

Frequency domain analyses

Time-frequency analysis was conducted via a Hann-windowed FFT calculated over the entirety of each epoch. Inter-trial phase coherence (ITPC) was calculated via the following process. First, the amplitude and phase at the frequency of within-band tone presentation was extracted from the FFT. Second, the amplitude of this vector was set to be equal to 1 by dividing the vector by its length. This resulted in a set of vectors with amplitude 1 but varying phases, with a single vector per trial. These

vectors were then phase-averaged; the resulting average vector is longer for vectors that have more consistent phases. The length of the average vector, therefore, was calculated as ITPC, which varies from 0 (no phase consistency across trials) to 1 (perfect phase alignment). Data analysis (both frequency domain and time domain) was conducted across the five channels with the highest inter-trial phase coherence at the within-band presentation rate across all conditions and subjects (FC6, F4, FC1, FC2, Fz). For frequency domain analyses, inter-trial phase coherence was calculated on a channel-by-channel basis across these five channels and then averaged.

ITPC in the attend-high and attend-low conditions was averaged together and compared to ITPC in the passive condition. This served to test the hypothesis that attention to one of the two bands would be linked to an increase in phase coherence at the attended rate, and that this attentional modulation would vary across frequencies. A 2 × 4 repeated measures ANOVA was run with two within-subjects factors—task (attention versus passive) and rate (2, 3, 4, and 5 Hz)—with log-transformed ITPC at the frequency of tone presentation as the dependent variable.

Time domain analyses

Effects of attention on the event-related waveform were investigated by computing the average waveform across trials for the attend-high,

attend-low, and passive conditions (separately for each presentation rate). Due to the continuous rhythmic stimulus presentation, there was not a meaningful pre-stimulus period for baseline amplitude calculation. Although there was a brief pause between sequences, one of our hypotheses was that effects of attention would continue through the pause. As a result, we could not assume that the period just before the onset of each trial was devoid of neural responses, and so epochs were baselined prior to averaging by subtracting the mean across the entire epoch.

Each sequence within each frequency band consisted of three tones. Thus, according to the neural entrainment hypothesis, each epoch should contain four cycles of a sinusoidal modulation of the ERP waveform due to attention: three cycles aligned with the three tones of the sequence, followed by a fourth continuing through the silence. To test this hypothesis, we calculated the average passive waveform and attentional modulation waveform (attend-high minus attend-low waveforms) underlying a single cycle of the hypothesized sinusoid. For example, for the 5 Hz condition, we averaged together the period between 0 and 0.2, 0.2 and 0.4, and 0.4 to 0.6 s to calculate a single average response with a duration of 0.2 s. This period consisted of the average response to a pair of tones, a low tone (since the sequence in the low band always began first) followed by a high tone. To investigate the extent to which passive responses and attentional modulation continued into the silence between sequences, we separately analyzed the remaining portion of the response (which, in the 5 Hz condition, would be the period between 0.6 and 0.8 s).

For the average response to each 'tone pair', paired t-tests were conducted to determine the time points (at 2 ms temporal resolution) where there was a significant difference in amplitude between a) attend-high and attend-low conditions, b) attend-high and passive conditions, and c) attend-low and passive conditions. We also used paired t-tests to investigate the difference in amplitude between the attend-high and attend-low conditions during the silence between sequences. In addition, unpaired t-tests were conducted to determine the time points at which there was a significant difference between the short-tone and long-tone experiments in the attentional modulation (attend-high minus attend-low) and passive response. Each of these analyses was separately corrected for multiple comparisons using the False Discovery Rate method (Benjamini and Hochberg, 1995).

Finally, we calculated the degree to which the attentional modulation (attend-high minus attend-low) was sinusoidal in shape; we did this by fitting a sinusoid to each participant's data for each presentation rate condition as follows. (To avoid overfitting, the phase of the sinusoid was determined using a leave-one-participant-out procedure). First, an average attentional modulation was calculated across all but one of the participants. Next, an FFT was used to extract the phase of the signal at the within-band presentation rate (2 Hz for the 2 Hz condition, for example). This phase was then used to construct a model sinusoid which was correlated with the attentional modulation waveform of the left-out participant. This procedure was conducted across all participants and all conditions. The resulting r-values were then converted to z-scores to allow comparison of the sinusoidal fit across presentation rate conditions.

Results

Behavioral results

Attention performance (see **Table 1**) differed across the four presentation rates (F(2.67, 74.68) = 2.94, $p < 0.05$), with better performance for slower rates. Nonetheless, average behavioral performance for each rate was well above chance levels (all d' > 2). Post-hoc Bonferroni-corrected t-tests were used to examine pairwise differences in attention performance between rates. No significant difference between conditions was found.

EEG time-frequency analyses

Because we did not have a prior hypothesis regarding effects of tone duration on attentional modulation of ITPC, EEG time-frequency analyses were collapsed over the Variable Tone Duration and Fixed Tone Duration experiments. When participants were asked to actively attend to one of the two bands and detect occasional repeated sequences, ITPC at the within-band presentation rate was higher than during passive listening (main effect of task F(1,28) = 37.65, $p < 0.001$; see **Table 1** and **Fig. 3**), thus replicating previous findings (Laffere et al., 2020a,b). Although there was no overall ITPC difference across rates (F(3,84) = 1.10, $p > 0.1$), the ITPC difference between active and passive conditions was smaller for the faster rates (task x rate interaction, F(3,84) = 7.00, $p < 0.001$). In the active conditions, post-hoc Bonferroni-corrected paired t-tests showed that ITPC was smaller in the 5 Hz condition relative to the 2 Hz (t(28) = 3.73, p(corrected) = 0.005), 3 Hz (t(28) = 4.26, p(corrected) = 0.001), and 4 Hz (t(28) = 3.01, p(corrected) = 0.033) conditions. No other differences between conditions were significant. In the passive conditions, no significant differences in ITPC between rate conditions were found (all $p > 0.1$).

EEG time-domain analyses: effect of tone duration

To investigate the effects of tone length on the passive response and attentional waveform modulation for the tone pairs, we first compared event-related waveforms for the Variable Tone Length and Fixed Tone Length Experiments, using False Discovery Rate to control for multiple comparisons (**Fig. 4**). Only for the 2 Hz rate were there significant differences between conditions for the attentional modulations. Importantly, there was no obvious trend for the *width* of the attentional modulation to be shorter in the Fixed Tone Length Experiment (in which the tones were 40 ms in length) as compared to the Variable Tone Length Experiment (in which the tone length varied between 100 and 250 ms). Indeed, any trend was in the opposite direction, and the significant difference in the 2 Hz condition showed longer modulations for the Fixed (40 ms) Tone Length Experiment (particularly in the 2 Hz and 5 Hz conditions). Overall, this analysis showed that, if sinusoidal attentional modulations scale with rate (as predicted by the neural entrainment hypothesis), this effect cannot be accounted for by a confounding effect of tone length. Subsequent analyses collapsed across the Fixed and Variable Tone Length Experiments.

EEG time-domain analyses: effect of attention

Next, we investigated the time course of the effects by computing the difference in the responses to low-high tone pairs between attend-low and attend-high conditions, collapsed across tone length condition. We first determined which portions of the response showed a significant difference across low versus high attention conditions after correction for multiple comparisons, and then compared the alignment between these attention effects and the shape of the passive response (**Fig. 5**). Across rates, the attentional modulation followed a roughly sinusoidal shape, and the length of the modulated portions of the responses scaled with rate, such that slower rates were linked to longer modulations. For example, while in the 4 Hz condition the initial significant positive attentional modulation extended from 70 to 180 ms, in the 2 Hz condition the modulation extended from 4 to 193 ms. Across rates, the peaks of the attentional modulation aligned with the N1 of the two tones in the passive condition, such that the positive modulation peak was aligned with the N1 of the low tone response and the negative modulation peak was aligned with the N1 of the high tone response. This suggests that the attentional modulation could be partially accounted for via enhancement of the N1 response to the low tone in the attend-low condition and enhancement of the N1 response to the high tone in the attend-high condition.

Table 1

Performance and neural metrics (means and standard deviations) from time-frequency analyses across presentation rate conditions.

	2 Hz	3 Hz	4 Hz	5 Hz
Attention performance (dprime)	2.70 (0.88)	2.73 (0.78)	2.58 (0.84)	2.44 (0.91)
Active log(ITPC)	−1.77 (0.46)	−1.68 (0.44)	−1.83 (0.43)	−2.07 (0.37)
Passive log(ITPC)	−2.35 (0.49)	−2.39 (0.49)	−2.16 (0.51)	−2.19 (0.45)
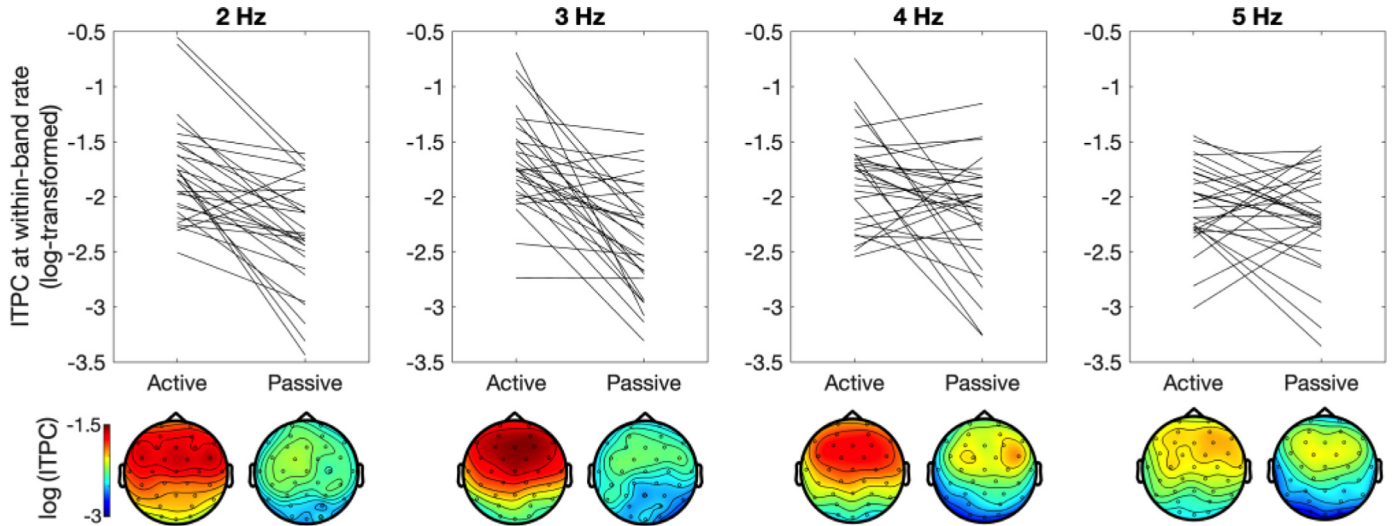
Fig. 3. ITPC at the within-band tone presentation rate in Active and Passive conditions across four different presentation rate conditions.
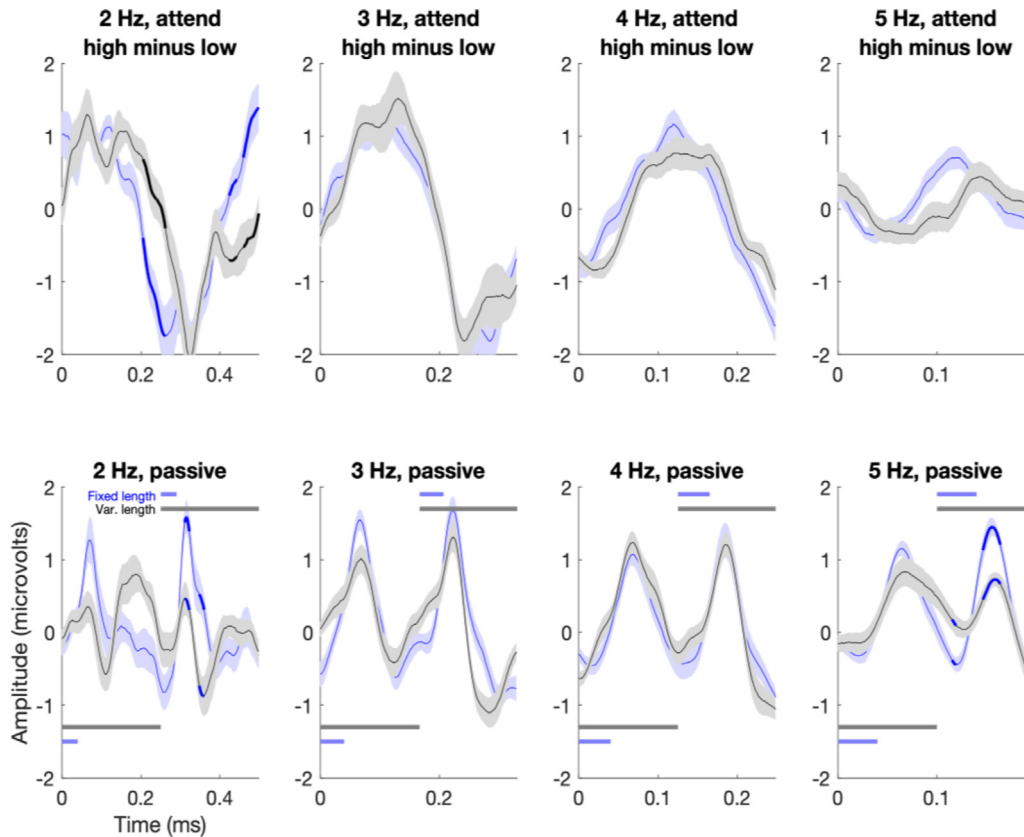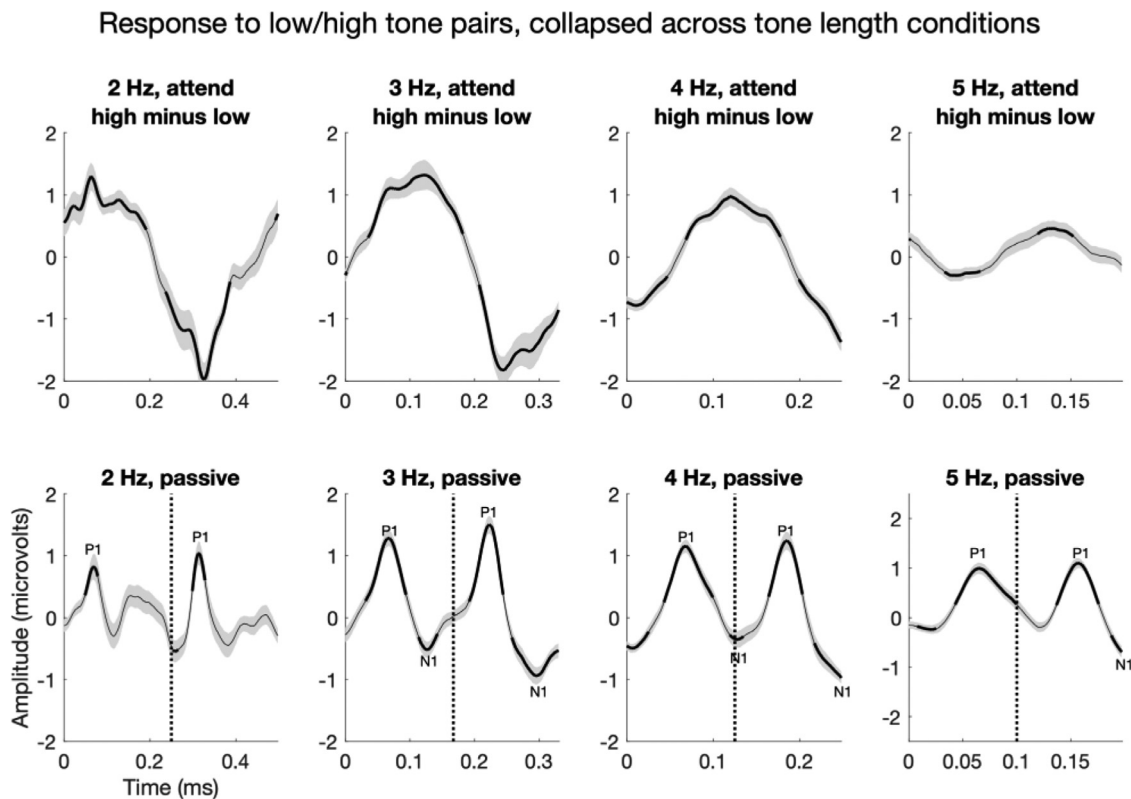

Fig. 4. Comparison of short tone (blue) and long tone (black) waveforms across four different rate conditions. Waveforms display the average response to low-high tone pairs. Horizontal blue and gray lines display the timing of tone presentation in the fixed length and variable length conditions, respectively, with the timing of the low tone displayed at the bottom of the plot, and that of the high tone displayed at the top of the plot. The top plots display the difference in waveforms between the attend high and attend low conditions, while the bottom plots display the responses in the passive conditions. Shaded regions indicate standard error of the mean. Thicker lines indicate time points in which difference between conditions survived correction for multiple comparisons.

Response to low/high tone pairs, collapsed across tone length conditions

Fig. 5. Difference between waveforms in attend high and attend low conditions (top) and passive responses (bottom) across four different rate conditions (please note difference in x-axis timescale across conditions). Waveforms display the average response to low-high tone pairs collapsed across tone length condition; the high tone began at the time indicated by the dotted vertical line. Shaded regions indicate standard error of the mean. Thicker lines indicate time points in which either the comparison between attend high and attend low waveforms (top) or comparison with baseline (bottom) survived correction for multiple comparisons. Note that positive-going voltage values are plotted as positive on the y-axis.

However, at the slower rates, the modulation was not limited to the time points associated with the passive N1. Instead, it entirely overlapped with the time points for which the passive P1 significantly exceeded baseline. In the 3 Hz condition, for example, the initial significant positive attentional modulation extended from 37 to 184 ms, while P1 significantly exceeded baseline between 33 and 96 ms. Similarly, in the 2 Hz condition, the initial significant positive modulation extended from 4 to 194 ms, while P1 significantly exceeded baseline from 51 to 80 ms. In the 4 Hz condition, there was a very early negative modulation from 4 to 49 ms (likely reflecting carryover from a previous cycle), followed by a positive modulation from 70 to 180 ms, followed by a negative modulation from 201 to 248 ms. Finally, in the 5 Hz condition, there was a positive modulation from 35 to 68 ms, followed by a negative modulation from 121 to 154 ms.

Next, we computed the difference in the responses to low-high tone pairs between each separate attention condition (attend-low or attend-high) and the passive condition, determining which portions of the waveform showed a significant difference between conditions after FDR correction for multiple comparisons. Comparing each attention condition to the passive condition enabled us to test the hypothesis that attentional modulations would be biphasic (consisting of both positive and negative modulations), even when only one of the two tones in the tone pair was attended. We began by examining the attend-low minus passive listening difference wave (**Fig. 6, top**). The attentional modulation was roughly sinusoidal in shape, with significantly modulated portions that scaled in length with rate, and an initial negative modulation followed by a subsequent positive modulation (after the onset of the high tone). For the 2 Hz condition, there was a significant negative modulation from 16 to 162 ms, followed by a significant positive modulation from 244 to 373 ms. For the 3 Hz condition, there was a significant neg-

ative modulation from 33 to 170 ms, followed by a significant positive modulation from 219 to 330 ms. For the 4 Hz condition, there was a significant positive modulation from 4 to 29 ms (likely a carry-over from previous tone pair cycles), followed by a significant negative modulation from 70 to 154 ms, followed by a significant positive modulation from 211 to 248 ms. For the 5 Hz condition, there was a significant negative modulation from 100 to 156 ms.

Next, we computed the difference in the responses to low-high tone pairs in attend-high and passive conditions, again determining which portions of the waveform showed a significant difference between conditions after FDR correction for multiple comparisons (**Fig. 6, bottom**). The evidence regarding the shape of the attentional modulation was somewhat less clear for this comparison, with in some cases only brief portions of the waveform being significantly modulated; however, there was evidence for a biphasic modulation in two of the four conditions, with an initial positive modulation followed by a subsequent negative modulation. For example, for the 2 Hz condition, there were significant positive modulations from 4 to 10 ms, from 49 to 78 ms, and from 160 to 178 ms, followed by a significant negative modulation from 273 to 377 ms. For the 3 Hz condition, there was a significant positive modulation from 107 to 178 ms, followed by a significant negative modulation from 207 to 293 ms. For both the 4 Hz and 5 Hz conditions, there were no significant attentional modulations.

Next, we calculated the extent to which the attentional modulation (high minus low) of the response to low-high tone pairs for each participant/condition resembled a single cycle of a sinusoid with a frequency equal to the within-band presentation rate. **Fig. 7** displays the correlations between the attentional modulation waveform and the fitted sinusoid for each participant/condition, as well as data from one exemplary participant. Overall there was a high degree of resemblance between the
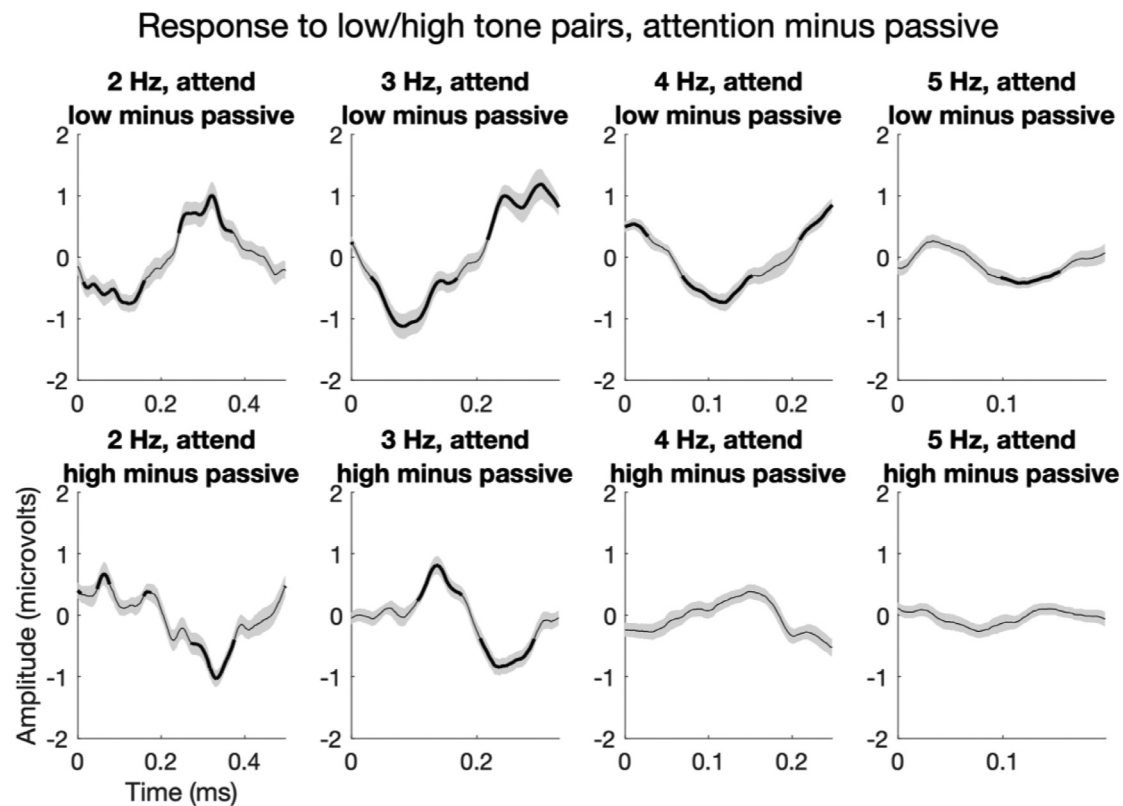
Fig. 6. Difference between waveforms in attend low versus passive conditions (top) and attend high versus passive conditions (bottom) across four different rate conditions (please note difference in x-axis timescale across conditions). Waveforms display the average response to low-high tone pairs. Shaded regions indicate standard error of the mean. Thicker lines indicate time points in which the comparison between waveforms survived correction for multiple comparisons. Note that positive-going voltage values are plotted as positive on the y-axis.

attentional modulation and the fitted sinusoid in all but the 5 Hz condition: median r-values in the 2, 3, 4, and 5 Hz conditions were 0.66, 0.83, 0.78, and 0.44, respectively. Most of these correlations were significant (FDR-corrected for multiple comparisons separately for each rate) and positive: 25, 28, 28, and 21 participants in the 2, 3, 4, and 5 Hz conditions, respectively. 3, 0, 0, and 2 participants showed non-significant correlations in the 2, 3, 4, and 5 Hz conditions. Significant negative correlations were found for one participant in the 2 Hz condition, one participant in the 3 Hz condition, one participant in the 4 Hz condition, and six participants in the 5 Hz condition (their modulations were inverted relative to the fitted sinusoid).

We then compared the degree to which the attentional modulation resembled the fitted sinusoid across rate conditions by first converting the p-values to r-values using Fisher's transform. The degree to which the attentional modulations matched the fitted sinusoid differed across rates ($F(2.27,63.51) = 12.64$, $p < 0.001$). Post-hoc Bonferroni-corrected paired t-tests showed that modulations in the 5 Hz condition matched the sinusoid less than in the 3 Hz ($t(28) = 4.55$, $p < 0.001$) and 4 Hz ($t(28) = 4.34$, $p < 0.001$) conditions. In addition, modulations in the 2 Hz condition matched the sinusoid less than in the 3 Hz condition ($t(28) = −5.16$, $p < 0.001$). All other comparisons were not significant.
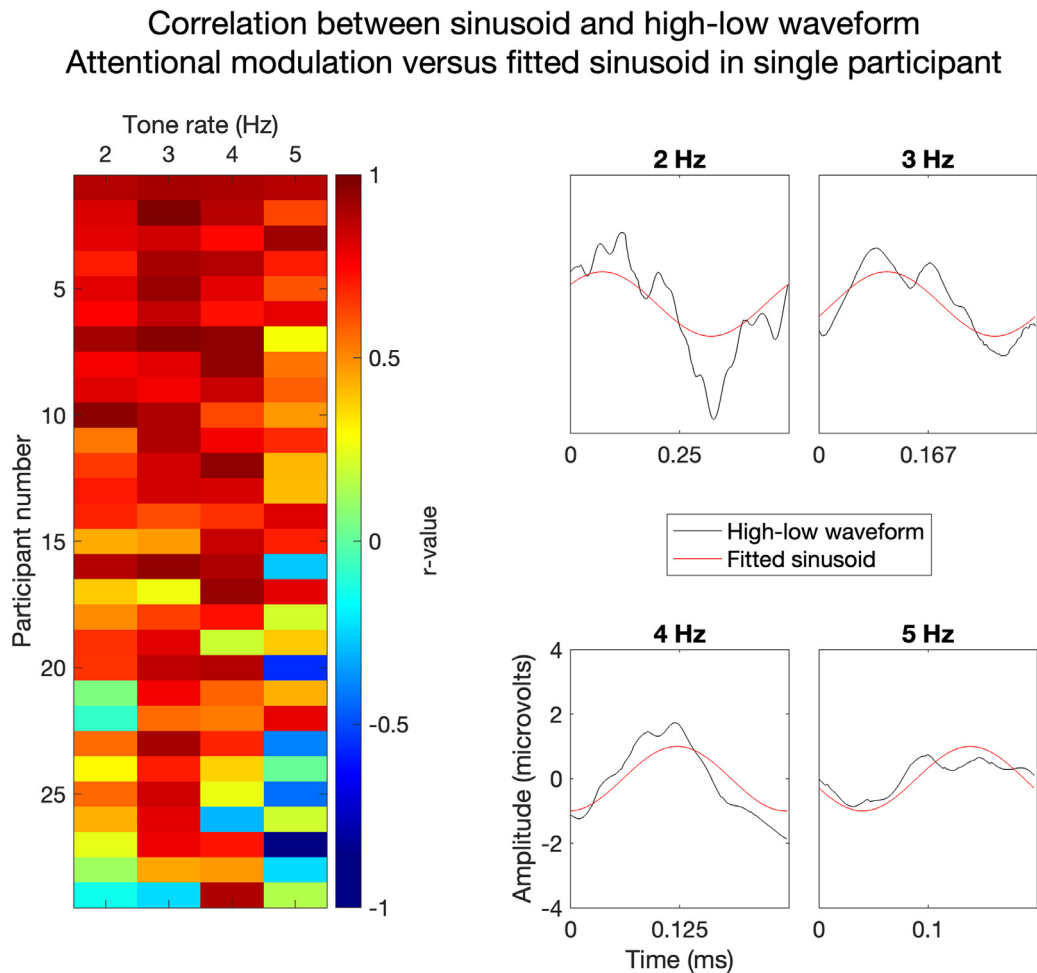
EEG time-domain analyses: forward entrainment

Finally, we investigated whether effects of attention extended into the silence between sequences (starting at the time point at which a seventh tone would have begun if the sequence were to continue through the silence). To do this, we compared the time course of these attentional modulations (attend-high minus attend-low) to the passive response, as shown in **Fig. 8**. Overall, we found limited evidence for a carry-over of

attention effects into the between-sequence period. At 2 Hz, for example, modulations were relatively brief compared to those identifiable during the sequence (see **Figs. 5 and 6**) and largely confined to the first half of the cycle. Specifically, the attend-high waveform was more positive than the attend-low waveform from 4 to 16 ms and from 29 to 88 ms, and more negative from 135 to 197 ms. Finally, there were two additional brief periods in which the attend-high waveform was more positive from 377 to 395 ms and from 430 to 449 ms. At 3 Hz, the attend-high waveform was more positive than the attend-low waveform between 80 and 180 ms but remained at baseline during the second half of the cycle. This modulation closely coincided with a significant negativity in the passive waveform between 74 and 178 ms. In the 4 Hz condition, a significant negative difference was found between 4 and 63 ms, and a significant positive difference was found between 131 and 227 ms. Only the negative difference could be aligned with a significant component in the passive response (between 4 and 74 ms). At 3 and 4 Hz, the time points showing significant attentional modulation overlapped with the time points showing modulations in the same direction (positive versus negative) in the response to tone pairs in **Fig. 5**. Finally, in the 5 Hz condition, a significant negative difference was found between 6 and 74 ms, and a significant positive difference was found between 150 and 197 ms. Only the negative difference could be aligned with a significant component in the passive response (between 4 and 115 ms).

To investigate the possible existence of late exogenous ERP components in our data, the modulation of which could account for the effects of attention, we examined the response to the final tone of each sequence in the 2 Hz and 3 Hz passive conditions. In these conditions participants were asked to sit quietly and were not given an explicit task, beyond advancing the stimulus presentation program at the conclusion of each block. This analysis revealed no clear evidence of any ERP components after around 200 ms (Fig. 9).

Correlation between sinusoid and high-low waveform
Attentional modulation versus fitted sinusoid in single participant

Fig. 7. (Left) Correlations (r-values) between the difference between the attend high and attend low waveforms for each participant and a sinusoid with a frequency at the tone presentation rate and a phase fitted to the average attend-high-minus-attend-low waveform across all other participants. Participants are sorted by the average correlation value across all four rates. (Right) Average responses to low-high tone pairs in the attend high versus attend low conditions (black) and the fitted sinusoid (red) in a single exemplary participant (corresponding to the top-most row of the plot on the left).
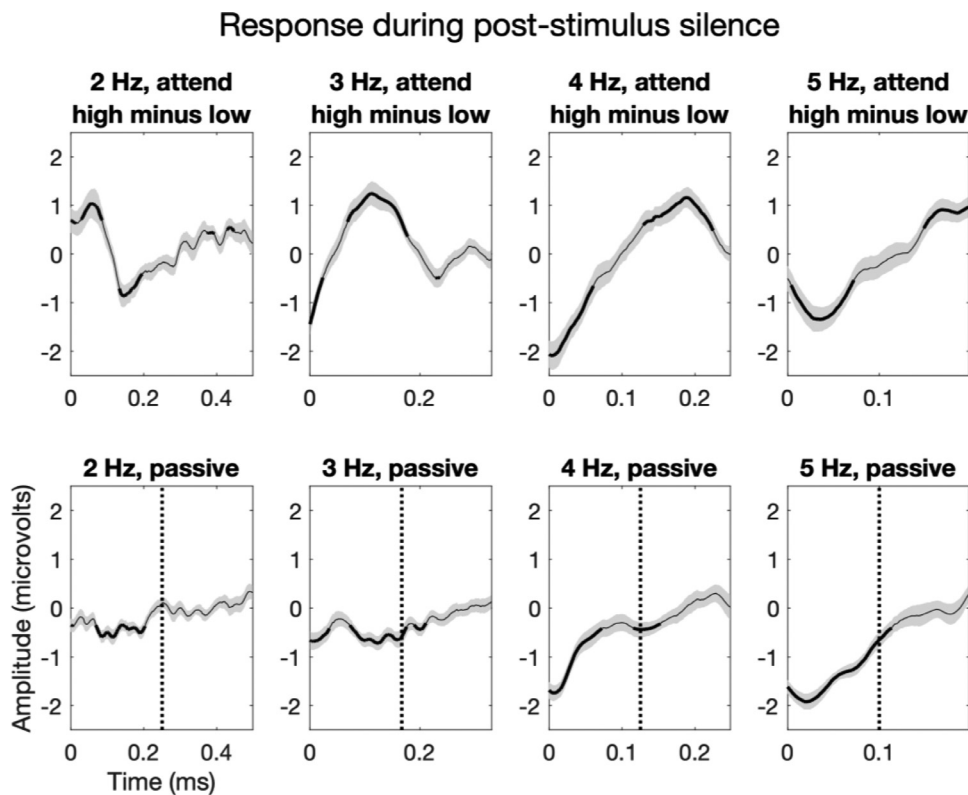
Summary and discussion

Predictions of neural entrainment and attentional filter accounts of selective attention
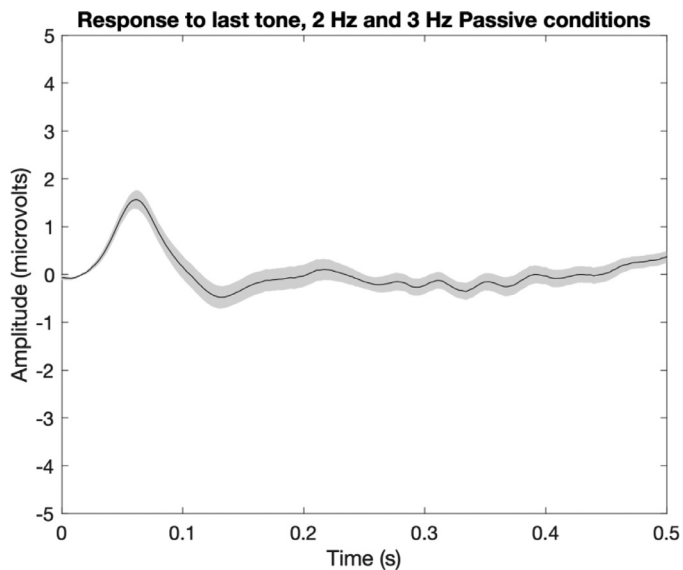
Here we used a sustained auditory selective attention task - where participants attended to one of two sound streams presented concurrently at the same rate but at opposite phases - to adjudicate between two theoretical accounts of auditory attentional mechanisms: an endogenous neural entrainment account versus an exogenous attentional filter account. Specifically, we asked whether attention-driven phase-locking reflects alignment of slow endogenous neural rhythms with the temporal structure of the stimuli, versus attentional modulation (or 'gain') of exogenous evoked waveforms. To disambiguate between these theoretical accounts, we manipulated tone stream rate across conditions. Under the endogenous attentional entrainment account, we would expect that the width of the phasic attentional modulation would scale inversely with increasing rate regardless of tone duration, with broader modulations for slower rates. We would also potentially expect that phasic attentional modulation would continue into the silent period at the end of each trial ("forward entrainment", Saberi and Hickok 2021). By contrast, on an attentional filter account, we would expect that the attentional modulation would remain limited to the time points associated with the N1, and that the phasic attentional EEG waveform would not continue into the silent period of each trial.

Attentional modulation is broader at slower rates

We found, when comparing the evoked waveforms elicited when attention was directed towards one versus the other stimulus stream, that direction of attention was linked to a roughly sinusoidal pattern of periodic positive and negative modulations. These negative/positive modulations were temporally centered on the N100 evoked responses to each of the tones in the attended/ignored stream, as recorded in the passive condition. This finding is broadly consistent with prior reports that the N100 can be modulated by selective attention (Hillyard et al., 1973; Sanders and Astheimer, 2008; Choi et al., 2013; Dai et al., 2018). However, the width of the attentional modulations scaled with rate, such that slower rates were linked to wider attentional modulations. This matches a prediction of the neural entrainment account that attentional modulation of the response to rhythmic sound streams would be sinusoidal in shape. In fact, the observed shape of the modulation of the waveform by attention could be closely modelled by fitting a sinusoid at the rate of tone presentation to the difference in response between the attend-high and attend-low conditions. With only a single free parameter (phase) and the use of a leave-one-participant-out method to avoid overfitting, the median correlation between the fitted sinusoid and the attentional modulation reached 0.66, 0.83, 0.78, and 0.44 for the 2, 3, 4, and 5 Hz conditions, respectively. Moreover, at slower rates the modulation extended into the time regions associated with other responses in the passive condition, including the P50 and (in the 2 Hz condition) the P200.

Response during post-stimulus silence

Fig. 9. ERP to the final tone in the sequence for the passive condition, averaged across the 2 Hz and 3 Hz rates. The shaded region indicates the standard error of the mean.

The only way this pattern of attentional modulations could be accounted for via modification of evoked exogenous responses would be to suppose that the P50 and the P200 were suppressed by attention. However, prior research suggests that the P50 and P200 are, rather, enhanced by attention (Woldorff and Hillyard, 1991; Woldorff et al., 1993).

Our finding that, at slower rates, the increased negativity associated with selective attention extends outside of the time region containing the N100 and into the time region containing the P50 suggests that modulation of exogenously-driven responses cannot be the primary mechanism

underlying the attention-driven changes in phase locking. However, this does not mean that the involvement of neural oscillations must necessarily be invoked to explain our findings. An alternate possibility is that the results reflect modulation of an endogenously generated potential that, at slower rates, exceeds the width of the N100. For example, several studies investigating the effects of selective attention on ERPs have reported the existence of an endogenous potential known as the prolonged negative shift (sometimes labelled Nd) that is temporally dissociable from the N100 (Näätänen et al., 1978; Michie et al., 1990; Woods and Alain, 1993, 2001; Degerman et al., 2008). Thus, our results could arguably be accounted for by invoking modulation of Nd. However, the Nd has not previously been reported to scale with rate in a manner that could generate the pattern of modulations across conditions which we find. For example, Woods and Alain (1993) report an Nd lasting around 200 msec in an experiment using inter-onset-intervals of only 50–210 milliseconds, whereas we found that in the 4 Hz condition (corresponding to 250 millisecond inter-onset intervals) the duration of the modulation was less than 100 ms. In addition, our results could potentially be explained by modulation of the P300. However, like the Nd, the P300 is an endogenous response (Donchin et al., 1978): it can be evoked in the absence of physical change in the stimulus (Cozzi et al., 2019), it is absent in non-REM sleep (Cote 2002), and it is completely abolished by anesthesia-induced unconsciousness (Plourde and Boylan 1991). The P300 was also not apparent in the passive conditions of the current dataset (see Fig. 9). Thus, even if our results could be explained via reference to the Nd or P300, these are endogenously generated potentials, and so this would not be consistent with an attentional filter account in which selective attention operates by controlling the gain of exogenous neural responses to sound.

Attentional modulation is rate-limited

Modulations of the evoked waveform appeared somewhat attenuated in the 5 Hz condition, suggesting that attentional modulation of neural entrainment differed across rates. To investigate this possibility,

we compared effects of attention on ITPC across rate conditions. The difference in ITPC between attention and passive conditions was significantly smaller in the 5 Hz condition compared to the 2 Hz, 3 Hz, and 4 Hz conditions. The attenuated attentional modulations in the 5 Hz condition could partially reflect task performance, which was lower at higher rates. However, performance was still well above chance in the 5 Hz condition, and this paradigm has been shown to be sensitive to the neural effects of selective attention even in participant populations who display relatively poor performance (such as children with ADHD; Laffere et al., 2020b). Moreover, we found no significant pairwise differences between performance at 5 Hz and the other three rates after correcting for multiple comparisons. The enhanced attentional modulations at slower rates could partially reflect the dominance of low frequencies in the EEG signal (Näpflin et al., 2007); however, there was no difference in ITPC between the 4 rate conditions for the passive task.

The relative lack of attentional modulation in the 5 Hz condition compared to the other rates could reflect a switch from a temporally-selective task strategy to a spectrally-selective strategy. Prior research has suggested involvement of the motor system in perception of an isochronous beat underlying a complex rhythmic stimulus, with activation of cortical and subcortical areas found during beat perception tasks even when participants were explicitly instructed not to move (Grahn and Brett, 2007; Grahn and Schuit, 2012; Nozaradan et al., 2017; Kotz et al., 2018; Cannon and Patel, 2021). One possibility, therefore, is that the motor system generates temporal predictions about the onset of an attended sound stream (Morillon and Schroeder, 2015). Temporal predictions which rely on the motor system may be rate-limited, as individuals' ability to align movements with the temporal structure of sound streams falls off rapidly as the stimulus presentation rate is increased (Repp, 2003). Supporting the involvement of implicit motor movement in temporal attention, Zalta et al. (2020) found that both motor tapping and auditory temporal attention showed similar dependence on rate, with optimal performance at close to 2 Hz and poorer performance at faster and slower rates.

Limited evidence for forward entrainment

If attentional modulation of neural entrainment reflects the alignment of endogenous neural oscillators with the temporal structure of attended stimuli, then attention-driven neural modulations could continue for a time even once stimuli have ceased ("forward entrainment", Saberi and Hickok 2021), given that some oscillators are self-sustaining (Doelling & Assaneo, 2021). To test this prediction, we designed the task such that participants attended to a series of short tone sequences so that we could investigate whether the sinusoidal attentional modulation continued through the silence between sequences. The evidence regarding continuance of modulations was not conclusive. Across all four rates, modulations were present in the silence between conditions, and there was temporal overlap between the modulated portions of the silence and the modulated portions of the phase cycle during stimulus presentation; however, in the 2 Hz and 3 Hz conditions these modulations were largely confined to the first half of the inter-sequence silence. These results place a constraint on an endogenous entrainment account of our findings, which would need to allow for rapid desynchronization/damping of neural oscillations after the cessation of stimulus presentation (Doelling & Assaneo, 2021). One explanation for this rapid damping could be that the silence between the sequences was both predictable and not task-relevant; future research could investigate whether the use of unpredictable silence timing leads to greater forward entrainment relative to predictable silences.

Conclusions

We find that attention to one of two sound streams, presented at the same rate but out of phase, is linked to an attentional modulation of the ERP waveform that has several interesting properties that help to constrain theories of temporally-selective attention. At slower rates, the attentional modulation of the evoked waveforms dissociates in time from passive evoked potentials, suggesting that the changes in phase alignment reflect synchronization of slow endogenous neural activity with the temporal structure of the attended stimulus rather than attenuation or enhancement of exogenous responses to stimuli. However, these modulations became smaller as the presentation rate increased, especially once the rate reached 5 Hz. This suggests that neural entrainment may only be a useful strategy for attentional selection at slower presentation rates, and that listeners may rely upon alternate mechanisms at higher rates.

Declaration of Competing Interest

The authors have no competing interests to declare, financial or otherwise

Credit authorship contribution statement

Magdalena Kachlicka: Methodology, Software, Investigation, Writing – original draft. **Aeron Laffere:** Methodology, Software, Investigation. **Fred Dick:** Conceptualization, Methodology, Writing – review & editing, Project administration. **Adam Tierney:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Project administration.

References

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. 57, 289–300.

Besle, J., Schevon, C., Mehta, A., Lakatos, P., Goodman, R., McKhann, G., Emerson, R., Schroeder, C., 2011. Tuning of the human neocortex to the temporal dynamics of attended events. J. Neurosci. 31, 3176–3185.

Best, V., Ozmeral, E., Shinn-Cunningham, S., 2007. Visually-guided attention enhances target identification in a complex auditory scene. JARO 8, 294–304.

Bonino, A., Leibold, L., 2008. The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex. JASA 124, EL321–EL327.

Cannon, J., Patel, A., 2021. How beat perception co-opts motor neurophysiology. Trends Cogn. Sci. 25, 137–150.

Chait, M., de Cheveigné, A., Poeppel, D., Simon, J., 2010. Neural dynamics of attending and ignoring in human auditory cortex. Neuropsychologia 48, 3262–3271.

Choi, I., Rajaram, S., Varghese, L., Shinn-Cunningham, B., 2013. Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. Front. Hum. Neurosci. 7, 115.

Cote, K., 2002. Probing awareness during sleep with the auditory odd-ball paradigm. Int. J. Psychophysiol. 46, 227–241.

Cooke, M., Lu, Y., 2010. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. JASA 128, 2059–2069.

Cozzi, J., Angel, R., Herdman, A., 2019. How can no change in an auditory stimulus generate an N2b-P3a? Brain Cogn 129, 9–15.

Da Costa, S., van der Zwaag, W., Miller, L., Clarke, S., Saenz, M., 2013. Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. J. Neurosci. 33, 1858–1863.

Dai, L., Best, V., Shinn-Cunningham, B., 2018. Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. Proc. Natl Acad. Sci. 115, E3286–E3295.

Degerman, A., Rinne, T., Särkkä, A., Salmi, J., Alho, K., 2008. Selective attention to sound location or pitch studied with event-related brain potentials and magnetic fields. Eur. J. Neurosci. 27, 3329–3341.

Dick, F., Lehet, M., Callaghan, M., Keller, T., Sereno, M., Holt, L., 2017. Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally directed attention and systematically related to cortical myeloarchitecture. J. Neurosci. 37, 12187–12201.

Ding, N., Simon, J., 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J. Neurosci. 33, 5728–5735.

Doelling, K., Arnal, L., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85, 761–768.

Fritz, J., David, S., Shamma, S., 2012. Attention and dynamic, task-related receptive field plasticity in adult auditory cortex. In: Neural Correlates of Auditory Cognition. Springer, New York, NY, pp. 251–291.

Fuglsang, S., Märcher-Rørsted, J., Dau, T., Hjortkjær, J., 2020. Effects of sensorineural hearing loss on cortical synchronization to competing speech during selective attention. J. Neurosci. 40, 2562–2572.

Gatehouse, S., Akeroyd, M., 2008. The effects of cueing temporal and spatial attention on word recognition in a complex listening task in hearing-impaired listeners. Trends Amplif. 12, 145–161.

Ghinst, M., Bourguignon, M., de Beeck, M., Wens, V., Marty, B., Hassid, S., Choufani, G., Jousmaki, V., Hari, R., Bogaert, P., Goldman, S., De Tiége, X., 2016. Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. J. Neurosci. 36, 1596–1606.

Grahn, J., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. J. Cogn. Neurosci. 19, 893–906.

Grahn, J., Schuit, D., 2012. Individual differences in rhythmic ability: behavioral and neuroimaging investigations. Psychomusicology 22, 105–121.

Hillyard, S., Hink, R., Schwent, V., Picton, T., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–180.

Holt, L., Tierney, A., Guerra, G., Laffere, A., Dick, F., 2018. Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. Hear. Res. 366, 50–64.

Horton, C., D'Zmura, M., Srinivasan, R., 2013. Suppression of competing speech through entrainment of cortical oscillations. Neurophysiology 109, 3082–3093.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13, 411–430.

Kerlin, J., Shahin, A., Miller, L., 2010. Attentional gain control of ongoing cortical speech representations in a "cocktail party. J. Neurosci. 30, 620–628.

Kotz, S., Ravignani, A., Fitch, W., 2018. The evolution of rhythm processing. Trends Cogn. Sci. 22, 896–910.

Kitterick, P., Bailey, P., Summerfield, A., 2010. Benefits of knowing who, where, and when in multi-talker listening. JASA 127, 2498–2508.

Laffere, A., Dick, F., Tierney, A., 2020a. Effects of auditory selective attention on neural phase: individual differences and short-term training. Neuroimage 213, 116717.

Laffere, A., Dick, F., Holt, L., Tierney, A., 2020b. Attentional modulation of neural entrainment to sound streams in children with and without ADHD. Neuroimage 224, 117396.

Lakatos, P., Karmos, G., Mehta, A., Ulbert, I., Schroeder, C., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320, 110–113.

Lakatos, P., Musacchia, G., O'Connel, M., Falchier, A., Javitt, D., Schoeder, C., 2013. The spectrotemporal filter mechanism of auditory selective attention. Neuron 77, 750–761.

Lakatos, P., Barczak, A., Neymotin, S., McGinnis, T., Ross, D., Javitt, D., O'Connell, M., 2016. Global dynamics of selective attention and its lapses in primary auditory cortex. Nat. Neurosci. 19, 1707–1719.

Michie, P., Bearpark, H., Crawford, J., Glue, L., 1990. The nature of selective attention effects on auditory event-related potentials. Biol. Psychol. 30, 219–250.

Morillon, B., Schroeder, C., 2015. Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. Ann. N. Y. Acad. Sci. 1337, 26–31.

Näätänen, R., Gaillard, A., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. 42, 313–329.

Näpflin, M., Wildi, M., Sarnthein, J., 2007. Test-retest reliability of resting EEG spectra validates a statistical signature of persons. Clin. Neurophysiol. 118, 2519–2524.

Nobre, A., van Ede, F., 2018. Anticipated moments: temporal structure in attention. Nat. Rev. Neurosci. 19, 34–48.

Nozaradan, S., Schwartze, M., Obermeier, C., Kotz, S., 2017. Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. Cortex 95, 156–168.

O'Sullivan, J., Power, A., Mesgarani, N., Rajaram, S., Foxe, J., Shinn-Cunningham, B., Slaney, M., Shamma, S., Lalor, E., 2015. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cereb. Cortex 25, 1697–1706.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci., 156869 2011.

Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. 23, 913–926.

Paltoglou, A., Sumner, C., Hall, D., 2009. Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. Hear. Res. 257, 106–118.

Plourde, G., Boylan, J., 1991. The long-latency auditory evoked potential as a measure of the level of consciousness during sufentanil anesthesia. J. Cardiothorac. Vasc. Anesth. 5, 577–583.

Repp, B., 2003. Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interval subdivision. J. Mot. Behav. 35, 355–370.

Riecke, L., Peters, J., Valente, G., Kemper, V., Formisano, E., Sorger, B., 2017. Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. Cereb. Cortex 27, 3002–3014.

Riecke, L., Formisano, E., Sorger, B., Baskent, D., Gaudrain, E., 2018. Neural entrainment to speech modulates speech intelligibility. Curr. Biol. 28, 161–169.

Saberi, K., Hickok, G., 2021. Forward entrainment: evidence, controversies, constraints, and mechanisms. bioRxiv doi:10.1101/2021.07.06.451373.

Sanders, L., Astheimer, L., 2008. Temporally selective attention modulates early perceptual processing: event-related potential evidence. Percept. Psychophys 70, 732–742.

Schroeder, C., Lakatos, P., 2009. Low-frequency neural oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18.

Shinn-Cunningham, B., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182–186.

Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., Oja, E., 2000. Independent component approach to the analysis of EEG and MEG recordings. IEEE Trans. Biomed. Eng. 47, 589–593.

Woldorff, M., Hillyard, S., 1991. Modulation of early auditory processing during selective listening to rapidly presented tones. Electroencephalogr. Clin. Neurophysiol. 79, 170–191.

Woldorff, M., Gallen, C., Hampson, S., Hillyard, S., Pantev, C., Sobel, D., Bloom, F., 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc. Natl Acad. Sci. 90, 8722–8726.

Woods, D., Alain, C., 1993. Frequency-related differences in the speed of human auditory processing. Hear. Res. 66, 46–52.

Woods, D., Alain, C., 2001. Conjoining three auditory features: an event-related brain potential study. J. Cogn. Neurosci. 13, 492–509.

Zalta, A., Petkoski, S., Morillon, B., 2020. Natural rhythms of periodic temporal attention. Nat. Commun. 11, 1051.

Zion Golumbic, E., Ding, N., Bickel, S., Lakatos, P., Schevon, C., McKhann, G., Goodman, R., Emerson, R., Mehta, A., Simon, J., Poeppel, D., Schroeder, C., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party. Neuron 77, 980–991.

Further reading

Alho, K., Töttölä, K., Reinikainen, K., Sams, M., Näätänen, R., 1987. Brain mechanism of selective listening reflected by event-related potentials. Electroencephalogr. Clin. Neurophysiol. 68, 458–470.

Alho, K., Woods, D., Algazi, A., 1994. Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. Psychophysiology 31, 469–479.

Bidet-Caulet, A., Mikyska, C., Knight, R., 2010. Load effects in auditory selective attention: evidence for distinct facilitation and inhibition mechanisms. Neuroimage 50, 277–284.

Choi, I., Wang, L., Bharadwaj, H., Shinn-Cunningham, B., 2014. Individual differences in attentional modulation of cortical responses correlate with selective attention performance. Hear. Res. 314, 10–19.

Donchin, E., Karis, D., Bashore, T., Coles, M., Gratton, G., 1968. Cognitive psychophysiology: systems, processes, and applications. In: Coles, M, Donchin, E, Porges, S (Eds.), Psychophysiology: Systems, Processes, and Applications. The Guilford Press, New York, pp. 244–267.

Huang, N., Elhilali, M., 2020. Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. Elife 9, e52984.

Itoh, K., Nejime, M., Konoike, N., Nakamura, K., Nakada, T., 2019. Evolutionary elongation of the time window of integration in auditory cortex: macaque vs. human comparison of the effects of sound duration on auditory evoked potentials. Front. Neurosci. 13, 630.

Lange, K., Rosler, F., Roder, B., 2003. Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: an event-related potential study. Psychophysiology 40, 806–817.

Michie, P., Solowij, N., Crawford, J., Glue, L., 1993. The effects of between-source discriminability on attended and unattended auditory ERPs. Psychophysiology 30, 205–220.

Schmidt-Kassow, M., Schubotz, R., Kotz, S., 2009. Attention and entrainment: p3b varies as a function of temporal predictability. Neuroreport 20, 31–36.

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., Kotz, S., 2011. Temporal regularity effects on pre-attentive and attentive processing of deviance. Biol. Psychol. 87, 146–151.

Schwartze, M., Farrugia, N., Kotz, S., 2013. Dissociation of formal and temporal predictability in early auditory evoked potentials. Neuropsychologia 51, 320–325.