



BIROn - Birkbeck Institutional Research Online

Laffere, A. and Dick, Frederic and Tierney, Adam (2020) Effects of auditory selective attention on neural phase: individual differences and short-term training. *NeuroImage* 213 (116717), ISSN 1053-8119.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/31159/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively



Effects of auditory selective attention on neural phase: individual differences and short-term training

Aeron Laffere^a, Fred Dick^{a,b}, Adam Tierney^{a,*}

^a Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

^b Division of Psychology & Language Sciences, UCL, Gower Street, London, WC1E 6BT, UK



ARTICLE INFO

Keywords:

Attention
Training
Auditory
EEG
Rhythm

ABSTRACT

How does the brain follow a sound that is mixed with others in a noisy environment? One possible strategy is to allocate attention to task-relevant time intervals. Prior work has linked auditory selective attention to alignment of neural modulations with stimulus temporal structure. However, since this prior research used relatively easy tasks and focused on analysis of main effects of attention across participants, relatively little is known about the neural foundations of individual differences in auditory selective attention. Here we investigated individual differences in auditory selective attention by asking participants to perform a 1-back task on a target auditory stream while ignoring a distractor auditory stream presented 180° out of phase. Neural entrainment to the attended auditory stream was strongly linked to individual differences in task performance. Some variability in performance was accounted for by degree of musical training, suggesting a link between long-term auditory experience and auditory selective attention. To investigate whether short-term improvements in auditory selective attention are possible, we gave participants 2 h of auditory selective attention training and found improvements in both task performance and enhancements of the effects of attention on neural phase angle. Our results suggest that although there exist large individual differences in auditory selective attention and attentional modulation of neural phase angle, this skill improves after a small amount of targeted training.

1. Introduction

Humans are constantly bombarded by multiple streams of information and must extract the most relevant signals while filtering out irrelevant information and noise. Unlike the visual system, the auditory system cannot rely on mechanical means to screen out unwanted information. Instead, the central nervous system must separate auditory streams (auditory object formation) and devote in-depth processing to one of those streams (auditory object selection (Shinn-Cunningham, 2008)).

When auditory objects have characteristic acoustic regularities, attention to acoustic dimensions such as time (Nobre and van Ede, 2018) and frequency (Dick et al., 2017) may help select target objects and potentially suppress distractors. For example, knowledge about auditory stream timing (Gordon-Salant and Fitzgibbons, 2004; Best et al., 2007) and pitch (Darwin et al., 2003) facilitates listening in cocktail party paradigms, while stimulus timing uncertainty makes detection of target tones among distractors more difficult (Bonino and Leibold, 2008). Over the last decade research on speech perception has uncovered a possible neural signature underlying temporally-selective attention: alignment of

neural activity with low-frequency fluctuations in the amplitude of stimulus streams, i.e. neural entrainment (Luo and Poeppel, 2007; Lakatos et al., 2008, 2019; Ding and Simon, 2012; Power et al., 2012; Zoefel et al., 2018; Obleser and Kayser, 2019). (By “temporally-selective attention”, we are referring to the use of temporal information to select one of multiple auditory streams. This concept, therefore, overlaps with but is distinct from the concept of temporal attention, i.e. direction of attention to a point in time (Nobre and van Ede, 2018), which could be performed on a single stream.) Selecting one auditory stream and ignoring another could therefore be facilitated by neural entrainment to predictable temporal patterns in the target stream (Schroeder et al., 2010). Indeed, when multiple speech streams are presented, phase-locking to the target speaker is enhanced relative to distractors (Kerlin et al., 2010; Ding and Simon, 2012; Horton et al., 2013; Zion Golumbic et al., 2013). Moreover, transcranial magnetic stimulation studies have shown that manipulating neural entrainment can facilitate perception of speech in competing speech (Riecke et al., 2018; Zoefel et al., 2018), demonstrating that neural entrainment may play a causal role in auditory object selection.

* Corresponding author. Birkbeck College, Malet Street, London, WC1E 7HX, UK.
E-mail address: a.tierney@bbk.ac.uk (A. Tierney).

<https://doi.org/10.1016/j.neuroimage.2020.116717>

Received 18 November 2019; Received in revised form 2 March 2020; Accepted 4 March 2020

Available online 10 March 2020

1053-8119/© 2020 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

There is reason to believe that neural entrainment could facilitate selective attention in complex non-verbal stimuli as well, such as music. In response to task instructions, listeners can modify the phase and tempo of a perceived musical beat (Iversen et al., 2009; Nozaradan et al., 2011). Musical beat perception has been tied to increased neural entrainment at the frequency of the perceived beat, as measured via increases in inter-trial phase coherence (Doelling and Poeppel, 2015) and spectral power (Nozaradan et al., 2011, 2012, 2016; Tierney and Kraus, 2014; Cirelli et al., 2016). One possible mechanism facilitating auditory selective attention to musical melodies, therefore, is that listeners could control the phase and frequency of the musical beat they perceive via modulation of the entrainment of slow neural activity, so that the perceived beat aligns with the temporal structure of the attended melody.

Prior work, therefore, has shown that temporally-selective and spectrally-selective attention can facilitate auditory object selection and has established neural correlates of these abilities. Individual differences in auditory selective attention, however, are poorly understood, as previous studies of the neural correlates of selective attention have focused on main effects of attention. In a preliminary study we found large individual differences in auditory selective attention performance when tone streams were separated in frequency and time (Holt et al., 2018), suggesting that individuals vary in their ability to control the direction of their attention within auditory dimensions. Here we investigated the neural foundations of individual differences in auditory selective attention performance by asking participants to respond to targets in an attended sequence of tones while ignoring a competing sequence in a different frequency band. The tone streams were close together in frequency but completely non-overlapping in time. This complete temporal separation made auditory scene analysis relatively trivial, given the importance of temporal coherence in the formation of auditory streams (Shamma et al., 2011), ensuring that this paradigm specifically measured individual differences in sustained auditory object selection. Moreover, the temporal separation caused temporally-selective attention to be a particularly useful strategy for object selection (but spectrally-selective attention a potentially useful strategy as well). We predicted that participants who demonstrated proficient selective attention would show greater neural entrainment to the temporal structure of the attended sequence.

If individual differences in auditory selective attention exist, they could either reflect intrinsic predispositions or experience. We examined the relationship between experience and the behavioral and neural foundations of auditory selective attention in two ways. First, we investigated the relationship between long-term musical experience and auditory selective attention. Prior studies comparing musicians and non-musicians have found that musicians perform better on selective attention tasks; this advantage is not limited to the auditory modality, but extends to the visual modality as well (Oxenham et al., 2003; Parbery-Clark et al., 2009; Rodrigues et al., 2013; but see Ruggles et al., 2014). We predicted, therefore, that musicians would show greater neural entrainment to the attended sound stream. Second, we examined changes in auditory selective attention and neural entrainment after short-term computerized training. Prior work has shown that short-term training can lead to enhancements in spatially-selective attention (Stevens et al., 2008, 2013; Isbell et al., 2017) and comprehension of speech in background noise (Whitton et al., 2014, 2017), but it is unknown whether short-term enhancements of attentional modulation of neural entrainment are possible. We tested participants before and after a 2 h online attention training session, and predicted that task performance and neural entrainment to the attended stream would be improved after training.

2. Materials and methods

2.1. Participants

The participants in this experiment were recruited over the course of

several months from a population of adults without a diagnosed neurological disorder or hearing impairment. Informed consent was obtained from all participants. Study procedures were approved by the Birkbeck Department of Psychological Sciences ethics committee. 43 participants (ages 18–43 years, $M = 28.9$, $SD = 7.5$, 23 female) in total were recruited for the initial testing session through a university database. Hearing thresholds were assessed prior to study procedures by presenting tones at octaves from 500 to 4000 Hz, with the intention that any subject who failed to hear tones presented at 20 dB SPL in a pretest assessment would be excluded from analyses. However, it was not necessary to exclude any participant on the basis of this hearing assessment. The sample size of 43 subjects was powered to detect a medium effect of $d = 0.5$ between conditions with parameters $\alpha = 0.05$ and $\beta = 0.8$. After the participants completed the first session, we invited them to complete a two-hour online training and return for a second EEG recording session. Because the participants had not already committed to a subsequent training and testing session, not all of them elected to return; however, approximately half ($n = 24$, ages 18–43 years, $M = 28.6$, $SD = 7.5$, 11 female) were willing and able to complete both the training and second EEG recording session. All analyses except the analyses of the effects of training were conducted on the data from the first testing session across all 43 participants, including the participants who did not return for the second testing session. (see Fig. 1)

2.2. Stimulus design

Stimuli were 125 ms cosine-ramped sine tones (sampling rate 48 KHz, generated in Matlab (MathWorks, Inc)). Tones were grouped in lower (185, 207.7, and 233.1 Hz) and higher (370, 415.3, and 466.2 Hz) frequency bands, with tones corresponding to musical notes F#3, G#3, and A#3 for the low band and F#4, G#4, and A#4 for the high band. For each trial, a sequence of 3 tones was pseudorandomly selected from each band and then concatenated in an interleaved manner to form a repeating pattern of six notes which alternated between low and high frequencies, followed by a 250 ms pause (Fig. 1). Thus, a single 1-s epoch contained two three-note sequences, one in each frequency band, followed by a pause. Within-band sequence repetitions occurred between three and six times in each thirty-trial block. After a repetition occurred, there was always at least one intervening non-repeating sequence before another repetition could begin. Trials were concatenated together into blocks of 30 trials; there was one run of 35 blocks per attention condition (Attend



Fig. 1. Schematic of stimulus design. Stimuli consisted of tones in a low-frequency band and a high-frequency band, with one octave of frequency separation between bands. Tone presentation was interleaved between bands, such that when a tone was presented in one band the other band was silent. Each trial consisted of three tone presentations in each band, followed by a silence equal in length to two tones. In the attend high condition (middle), participants detected repeats in the high band (orange notes) and ignored the low band. In the attend low condition (bottom), participants detected repeats in the low band (orange notes). (Note that repeats also occur in the passive condition and the unattended band, but are not depicted here). For ease of representation on one staff, the depicted note values are one octave higher than those used in the experiment.

High Band, Attend Low Band, Passive Listening) per experimental session (see below). The amplitude of the tones in the high band was set at 40% of the amplitude of the tones in the low band. This amplitude ratio was chosen to balance the approximate loudness of the stimuli, given the tendency for higher-frequency stimuli to sound louder.

2.3. Task

All trials for a particular condition were presented in a single set of 35 blocks, in order to minimize task-switching requirements; condition order was counterbalanced across participants. In the “Attend High” and “Attend Low” conditions, participants were instructed to listen to the attended band and respond to occasional sequence repeats by clicking the mouse, while ignoring repeats in the other band. In other words, this was a “1-back” task, which required participants to continually compare the current sequence to the sequence immediately prior, and press a button if they were identical; participants were not asked to compare non-adjacent sequences. In the “Passive Listening” condition (EEG only), participants sat quietly and listened to the tones but were not asked to actively attend to either band. The latency window for a response to be recorded was from 250 ms before to 1250 ms after the end of the last tone in a sequence. (Note that response recording began before the end of a stimulus because, in theory, participants could detect a repeat as soon as the final tone of a sequence began.) Text feedback was presented on a plain background computer display to notify subjects of correct responses, missed targets, and incorrect responses. Performance was measured as d-prime. This task was designed to require participants to direct attention to a specific sound stream, while also requiring the distractor sound stream to be ignored (since it also contained targets). The length of the sequences was set at three notes to minimize the difficulty of the 1-back task, in an attempt to ensure that performance largely reflected the ability to direct attention to and integrate over the target sequence (by taking advantage of the spectral and temporal separation of the two tone streams), while ignoring the distracting tone sequence.

2.4. Experimental design

The experiment took place over three days. On the first day, participants completed a pre-training EEG session; on the second day, participants completed a two-hour online training session, while on the third and final day, participants returned for a post-training EEG session. Visual feedback was displayed on all days for correct responses, missed targets, and incorrect responses.

EEG recording sessions were conducted in a sound-attenuated room at the Birkbeck Department of Psychological Sciences. Participants completed thirty-five blocks of thirty trials, for a total of 1050 trials, in each condition (Attend High Band, Attend Low Band, and Passive Listening). ER-3A insert earphones (Etymotic Research, Elk Grove Village, IL) were used for sound presentation at 80 dB SPL during EEG sessions.

Online training took place at the participant’s home, with sounds presented through the participant’s own headphones at a comfortable amplitude of their choosing. The training session began by training participants to detect repeated sequence targets in a single frequency band in isolation (5 blocks of 30 trials). On the sixth block of trials, sequences from the other (distracting) frequency band were introduced at an attenuated level (−30 dB) and then increased in amplitude by 5 dB every subsequent 3 blocks until reaching their original amplitude presentation. Participants then completed 70 blocks of 30 trials at the typical amplitude presentation before repeating the above procedure while targeting the other frequency band. Condition order (i.e. whether attend high or attend low was trained first) was counterbalanced across participants.

Opportunities for rest were made available to subjects at the end of each set of trials. The full study procedure lasted about 2 h on each day.

2.5. EEG recording procedure

All data were recorded from a BioSemi™ ActiveTwo 32-channel EEG system and digitized with 24-bit resolution. All channels were linked to Ag–AgCl active electrodes recording at a sampling rate of 16,384 Hz and positioned in a fitted headcap according to the standard 10/20 montage. External reference electrodes were placed at the earlobes to record unlinked data for offline re-referencing. Contact impedance was maintained beneath 20 kΩ.

2.6. EEG data processing

Markers for the beginning of each block of 30 trials were recorded from trigger pulses sent to the neural data collection computer. EEG data were downsampled to 500 Hz and then segmented into 1 s epochs aligned with trial onsets. Any epoch containing a button press was not analyzed so as to minimize contamination of the EEG with motor-related potentials. Sources of artefact such as eye blinks and eye movements were identified by independent component analysis of the epoched recording and removed after visual inspection of component topographies and time courses. An average of 1.78 (± 0.70) components was removed per recording. A low-pass zero-phase sixth-order Butterworth filter with a cutoff of 30 Hz was applied (to minimize rejection of epochs due to myogenic artefact) and a high-pass zero-phase fourth-order Butterworth filter with a cutoff of 0.3 Hz was applied (to minimize rejection of epochs due to slow drift which would not affect our time-frequency analyses). Any epochs containing signal intensity in any individual channel exceeding $\pm 100 \mu\text{V}$ were rejected. All preprocessing steps were carried out using a blend of custom and premade scripts from the FieldTrip M/EEG analysis toolbox (Oostenveld et al., 2011) in MATLAB (Mathworks, Inc.).

EEG data were processed to extract measurements of average neural phase and inter-trial phase coherence (ITPC; Tallon-Baudry et al., 1996). Inter-trial phase coherence (ITPC) is a measure of the consistency of alignment of neural phases at a particular frequency, while average neural phase indicates the timing of the neural response with respect to the stimulus at a particular frequency. To calculate ITPC, a Hann-windowed fast Fourier transform was conducted on each trial. The amplitude of the resulting complex vectors for each frequency was then set to one by dividing by the vector’s length. The vectors were averaged, and the length of the resulting vector was calculated as ITPC, while the phase of the resulting vector equaled the average phase across trials. This measure varies from 0 (no phase consistency) to 1 (perfect phase alignment). In other words, following Delorme and Makeig (2004), ITPC was calculated as:

$$ITPC(f, t) = \frac{1}{n} \sum_{k=1}^n \frac{F_k(f, t)}{|F_k(f, t)|}$$

where t is time, f is frequency, F is the Fourier transform, and n is the number of trials. This procedure calculates ITPC at 1 Hz steps from 1 Hz up to the Nyquist frequency, but we only analyzed ITPC at 4 and 8 Hz and average neural phase only at 4 Hz. 4 Hz and 8 Hz were analyzed due to the presence of these rates in the stimuli (4 Hz is the within-band presentation rate, while 8 Hz is the across-band presentation rate).

We used ITPC and average neural phase angle to test different predictions of the neural entrainment account of auditory selective attention. According to this account, attention to sound streams is linked to alignment of slow neural rhythms to the temporal structure of the attended streams. Given that both the high and low streams contained temporal regularities at the same rate (4 Hz), attention to either stream (relative to the passive condition) should be linked to an increase in the degree of neural synchronization across trials at this rate. Accordingly, we predicted that ITPC would be greater at 4 Hz in the active relative to the passive conditions. However, although the high and low streams contained regularities at the same rate, they were separated in time by

half a phase cycle (180°) at 4 Hz, since when a tone in one band was presented, the other band was silent. As a result, if neural activity aligns with the attended stream, the direction of attention should be linked to a shift in average neural phase angle. Accordingly, we predicted that average neural phase angle at 4 Hz would differ when participants directed their attention to the high band versus the low band.

To limit the analysis to channels where ITPC was most strongly present, any EEG channel in which ITPC was less than 0.1 at 4 Hz averaged across all conditions was excluded from further analyses (of 32 original channels, 14 channels remained: Fp1, AF3, F3, FC1, FC5, C3, CP1, C4, FC6, FC2, F4, AF4, Fz, and Cz). (Note that, because the data was collapsed across conditions prior to channel selection, this procedure did not bias our analysis of differences in ITPC and average neural phase angle across conditions. Also, to confirm that our results are robust to channel montage, we reran all analyses with a montage including all 32 channels, and replicated all of the significant effects described below.) See Fig. 2A for a topographic plot of ITPC at 4 and 8 Hz in active and passive conditions.

2.7. Analysis

Descriptive and inferential analyses of neural phase data were carried out using the Matlab circular statistics toolbox (Berens, 2009). Given that the main neural outcome measures were phase-based, resulting in distributions which are bounded and therefore not expected to be normally distributed, non-parametric tests were used throughout. Processed data are available at osf.io/kwhdz/.

3. Results

3.1. Effects of direction of auditory selective attention on average neural phase angle and ITPC

To examine whether actively directing attention to one of the two frequency bands affected the strength of neural phase-locking, we calculated inter-trial phase coherence (ITPC) at 4 Hz (the within-band presentation rate). ITPC was higher for the active conditions (mean active ITPC: 0.111 ± 0.008 ; passive ITPC: 0.078 ± 0.007 ; Wilcoxon signed rank test, $z = 2.7652$, $p = 0.0057$). We also calculated ITPC at the cross-frequency-sequence rate of stimulus presentation (8 Hz) and found no significant difference between active and passive conditions (mean active ITPC: 0.198 ± 0.017 , passive ITPC: 0.195 ± 0.016 ; Wilcoxon signed rank test, $z = -0.2173$, $p = 0.83$). A follow-up Wilcoxon signed rank test showed that the active-passive difference was significantly greater at 4 Hz compared to 8 Hz ($z = 1.9924$, $p = 0.046$), indicating that the greater consistency of neural alignment across trials in the active condition was specific to the within-band repetition rate. ITPC in active versus passive conditions at 4 and 8 Hz is displayed in Fig. 2A.

Given that the tones within the high and low frequency bands were presented 180° out of phase, we expected attention to the high versus low band to be linked to a roughly 180° -degree shift in average neural phase angle at 4 Hz. To test this hypothesis, we investigated whether the direction of attention (i.e. attend high or attend low) led to a change in the average neural phase angle at 4 Hz across conditions. The distribution of participants' average neural phase angle at 4 Hz was significantly different between the two attention conditions (Hotelling paired sample test, $F(2,41) = 6.58$, $p = 0.0033$; Fig. 2B), suggesting that attention modulated the phase alignment between neural activity and the temporal structure of the stimuli.

3.2. Individual differences in auditory selective attention and average neural phase angle

Before training, there were large individual differences in behavioral performance: d-prime averaged over both attention conditions ranged between 0.05 and 2.98 (mean d-prime collapsed across conditions: 1.38

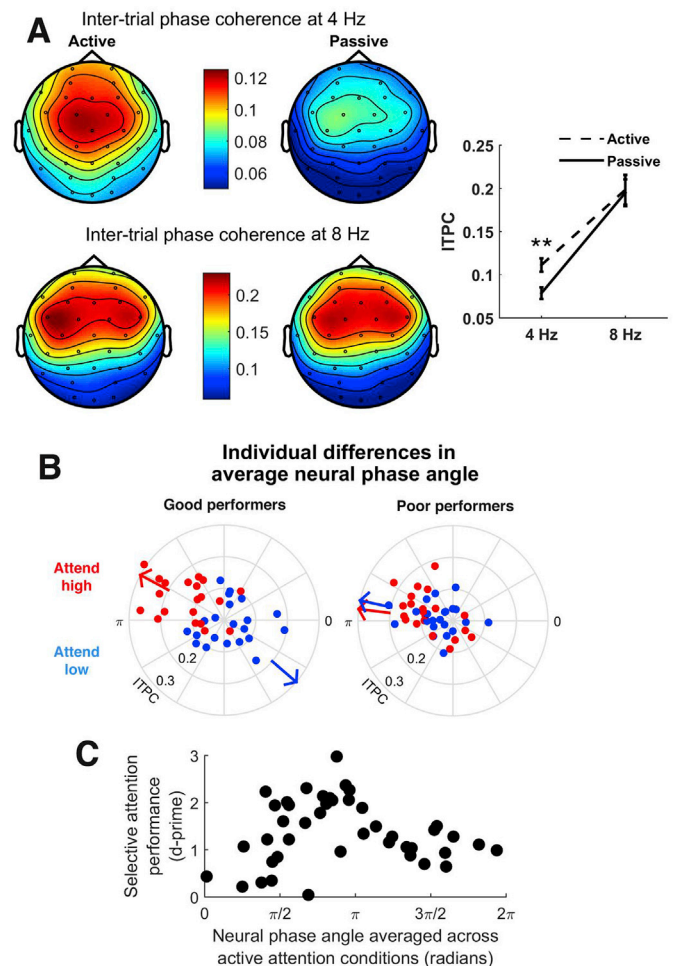


Fig. 2. (A) Left: topographic distribution of inter-trial phase coherence (ITPC) at 4 Hz (top) and 8 Hz (bottom) in the active (left) and passive (right) conditions. (ITPC was computed separately for attend high and attend low conditions and then averaged). ITPC was significantly greater in the active conditions than the passive conditions at 4 Hz (Wilcoxon signed rank test, $z = 2.7652$, $p = 0.0057$) but not 8 Hz ($z = -0.2173$, $p = 0.82$). Right: average ITPC at 4 and 8 Hz in active (dashed line) and passive (solid line) conditions. Error bars indicate standard error. (B) Average neural phase angle for each participant at 4 Hz in attend high (red) and attend low (blue) conditions. The greater the distance from the center, the higher the ITPC. Arrows indicate the average neural phase angle across all participants. Average neural phase angle significantly differed between attend high and attend low conditions (Hotelling paired sample test, $F(2,41) = 6.58$, $p = 0.0033$). Participants are divided into good performance and poor performance groups via a median split performed on hit rates across conditions to make clear the greater separation in average neural phase angle between attention conditions in good performers. (C) Scatterplot displaying relationship between participants' average neural phase angle at 4 Hz (averaged over active conditions) and selective attention performance ($R^2 = 0.33$, $p = 3.8 \times 10^{-4}$). The average neural phase angle from the attend low condition was reversed by 180° prior to averaging with the average neural phase angle from the attend high condition. Each data point corresponds to average neural phase angle across trials calculated from a single participant.

± 0.67). Given this high degree of performance variability, we investigated the relationship between behavioral performance and several neural metrics. First, we calculated the correlation between performance and ITPC at the within-band presentation rate (4 Hz) during the active conditions, and found a trending positive relationship (Spearman's $\rho = 0.29$, $p = 0.061$). Next, we asked whether certain average neural phase angles at the attended stimulus rate of 4 Hz were linked to better

performance. We averaged neural phase angle across conditions (after flipping average neural phase angle in the attend low condition by 180°), and correlated average neural phase angle with task performance by calculating the circular-linear correlation coefficient using `circ_corrcl` in the Matlab circular statistics toolbox (Berens, 2009). There was a significant relationship between performance and average neural phase angle ($R^2 = 0.33$, $p = 3.8 \times 10^{-4}$). To illustrate the relationship between attention performance and average neural phase angle, Fig. 2B displays average neural phase angle for the attend high and attend low conditions for good versus poor performers, defined via median split on d-prime scores. However, note that the analyses of individual differences reported in the main text were conducted continuously across all participants. Fig. 2C displays the continuous relationship between average neural phase angle and performance.

3.3. Relationship between musical experience and auditory selective attention

Degree of musical training was assessed by asking participants how many years of formal musical training they had received on a musical instrument, including singing. Participants with more years of musical training demonstrated better performance (Fig. 3; $\rho = 0.55$, $p = 1.2 \times 10^{-4}$) but did not have stronger ITPC at the within-band presentation rate (4 Hz) in the active conditions ($\rho = 0.08$, $p = 0.62$). However, musicians and nonmusicians differed in the magnitude of the effect of attention condition on average neural phase angle at 4 Hz. The difference in average neural phase angle between the attend high and attention low conditions was greater in participants with more musical training ($\rho = 0.57$, $p = 6.4 \times 10^{-5}$; Fig. 3).

3.4. Changes in auditory selective attention after computerized training

After training, participants' performance improved (mean d-prime before training, 1.56 ± 0.59 ; after training, 2.23 ± 0.59 ; Wilcoxon signed rank test, $z = 4.2857$, $p = 1.8 \times 10^{-8}$; Fig. 4A). However, there was no significant difference between ITPC for the active conditions at 4 Hz at day 1 and day 3 (before training: 0.124 ± 0.011 ; after training: 0.11 ± 0.009 ; Wilcoxon signed rank test, $z = -1.6857$, $p = 0.0919$; Fig. 5A). Nevertheless, ITPC and performance were correlated both before training ($\rho = 0.42$, $p = 0.042$) and after training ($\rho = 0.46$, $p = 0.024$; Fig. 5B). The gains in performance were accompanied by an enhancement of the effect of attention on average neural phase angle at 4 Hz, as the difference in average neural phase angle between the two attention conditions was significantly greater on day 3 than on day 1 (before

training: 1.30 ± 1.02 radians; after training: 1.87 ± 0.86 radians; Wilcoxon signed rank test, $z = 2.51$, $p = 0.012$; Fig. 4B and C).

4. Discussion

We investigated selective attention to sound by asking participants to detect targets in one of two sound streams that were non-overlapping in time and frequency. We found large individual differences in this task that were related to the extent to which the average neural phase angle at the within-band presentation rate (4 Hz) was modulated by the direction of attention. This suggests that alignment of neural activity with moments in time at which a target is likely to appear may help listeners select a sound stream for further processing. This finding is consistent with previous reports of electrophysiological studies in non-human animals and electrocortical studies of human epilepsy patients indicating that switching attention from one stimulus stream to another is linked to a shift in neural phase angle (Lakatos et al., 2008, 2009; 2013, 2016; Besle et al., 2011). However, the large individual differences in attentional modulation of neural entrainment we find suggest that there exists widespread variability in the extent to which listeners are able to use spectrotemporal regularities to focus on a target sound stream.

Participants with greater amounts of musical training showed greater effects of attention condition on average neural phase angle at 4 Hz, as well as a sizeable advantage in performance, suggesting that musical experience can enhance auditory selective attention to sound and top-down modulation of the timing of neural activity. These results are consistent with prior reports that musicians display an advantage for selective attention in visual (Rodrigues et al., 2013) and non-verbal auditory (Oxenham et al., 2003) stimuli, as well as enhanced segregation of auditory streams (Zendel and Alain, 2013). Indeed, given the usefulness of temporally-selective attention for speech perception in complex environments, our finding that musicians are better able to modulate the timing of their neural responses in response to task demands provides one possible explanation for prior reports that musicians display an advantage in perceiving speech in background noise (Parbery-Clark et al., 2009; Swaminathan et al., 2015; Clayton et al., 2016; Slater and Kraus, 2016; Deroche et al., 2017; Meha-Bettison et al., 2017; Morse-Fortier et al., 2017; Tierney et al., 2020; but see Ruggles et al., 2014; Madsen et al., 2017). Future work could test this explanation by examining whether attention-driven alignment of neural activity to the speech envelope is enhanced in musicians as well.

By contrast, we found that inter-trial phase coherence did not relate to amount of musical training in any condition, whether at the within-band or the across-band presentation rate. This suggests that the musician

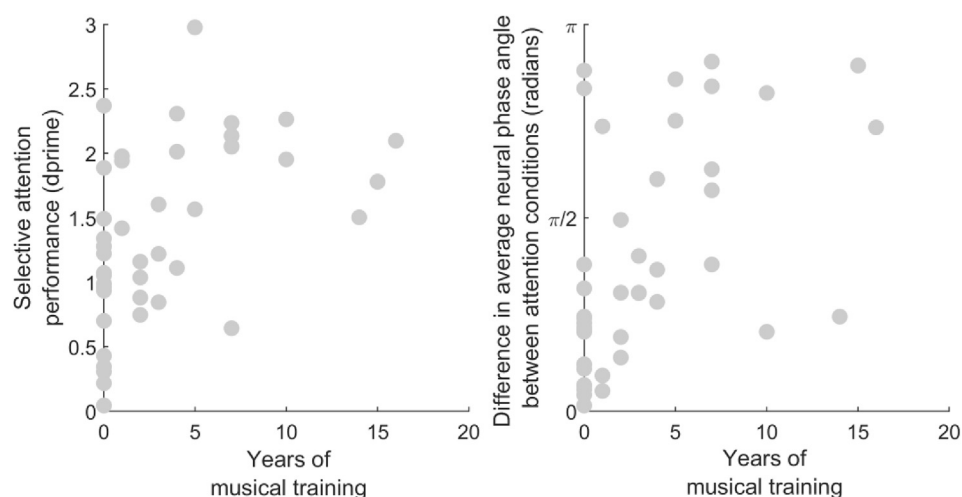


Fig. 3. Left: Relationship between years of musical training and selective attention performance (Spearman's $\rho = 0.55$, $p = 1.2 \times 10^{-4}$). Right: relationship between years of musical training and the effect of attention on average neural phase angle at 4 Hz ($\rho = 0.57$, $p = 6.4 \times 10^{-5}$).

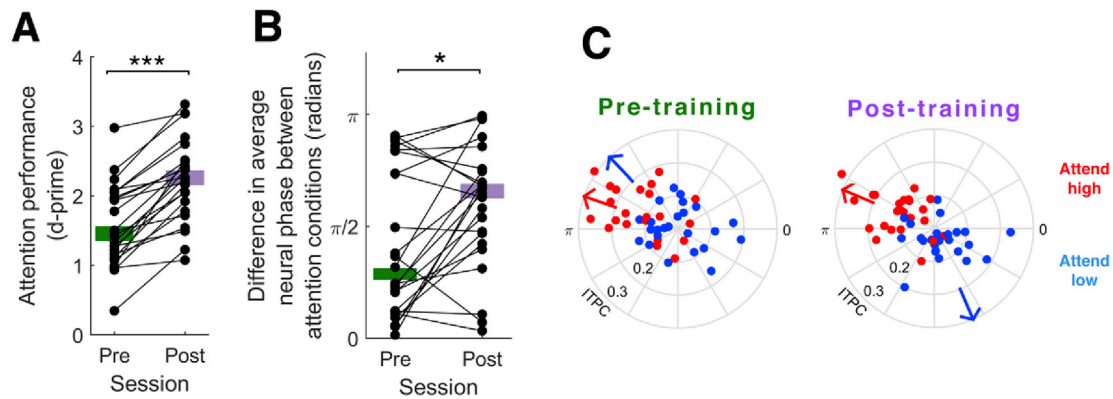


Fig. 4. (A) Change in selective attention performance before and after 2 h of online training. Selective attention performance (d-prime) was higher after training compared to before training (Wilcoxon signed rank test, $z = 4.2857$, $p = 1.8 \times 10^{-8}$). Thick horizontal lines indicate median performance. (B) Change in the difference in average neural phase angle at 4 Hz between attend high and attend low conditions from pre-training to post-training. Thick horizontal lines indicate median phase difference. The effect of attention on average neural phase angle was greater after training compared to before training (Wilcoxon signed rank test, $z = 2.51$, $p = 0.012$). (C) Average neural phase angle at 4 Hz in attend high (red) and attend low (blue) conditions before (left) and after (right) training. Distance from the center indicates ITPC. Arrows indicate the average neural phase angle across all participants.

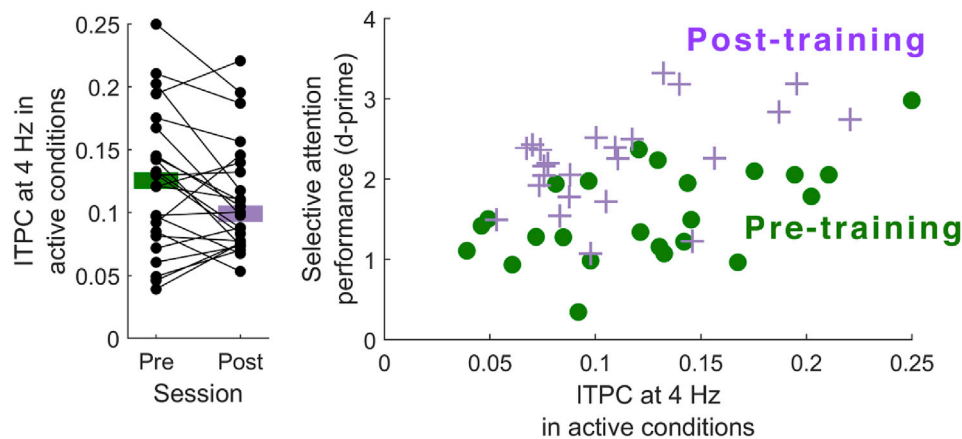


Fig. 5. (A) Change in ITPC at 4 Hz from pre-training to post-training. ITPC did not significantly change between pre-test and post-test (Wilcoxon signed rank test, $z = -1.6857$, $p = 0.092$). (B) Scatterplot displaying the relationship between ITPC at 4 Hz and selective attention performance before and after training. ITPC and performance were correlated both before training ($\rho = 0.42$, $p = 0.042$) and after training ($\rho = 0.46$, $p = 0.024$).

advantage was specifically for attentional modulation of average neural phase angle, rather than a more general enhancement of auditory responses regardless of task. These results were unanticipated, and are somewhat inconsistent with previous reports that participants with musical training show enhanced cortical responses in passive listening paradigms (Pantev et al., 1998; Shahin et al., 2003; Schneider et al., 2005; Seither-Preisler et al., 2014; Tierney et al., 2015; Habibi et al., 2016) and increased phase locking to music (Doelling and Poeppel, 2015; Harding et al., 2019) and to unattended speech (Puschmann et al., 2019). Our finding of a lack of a musician advantage for inter-trial phase coherence may reflect the simplistic nature of the stimuli: it is possible that musicians may demonstrate enhanced phase coherence only for stimuli which are timbrally, melodically, or rhythmically complex.

We found that after a few hours of training there was a considerable increase in selective attention performance, with an average change in d-prime from 1.56 to 2.23 between the pre-training and post-training testing sessions. Moreover, post-training we found an enhanced effect of attentional focus on average neural phase angle at 4 Hz, suggesting that individuals can rapidly improve their top-down control over the timing of neural activity. The behavioral and neural changes demonstrate the possibility of rapid short-term plasticity in the mechanisms of auditory selective attention. However, whether these improvements were

truly due to the specific training applied here—as opposed to simply being due to exposure to the task across the two in-lab testing sessions—cannot be concluded from our results, and will require additional studies with control treatment arms. Future work could examine whether short-term training programs could be a successful remediation strategy in populations who struggle to control attention. It also remains to be seen whether this enhanced auditory selective attention extends to perception of competing streams of speech as well.

Our findings are consistent with theories of auditory selective attention which suggest that auditory object selection is facilitated by the alignment of endogenous neural activity with the temporal structure of attended stimuli (Lakatos et al., 2008; Schroeder et al., 2010). According to these theories, the phase of ongoing neural oscillations is reset by acoustic edges in the attended signal, leading to a reliable phase relationship between the stimulus and neural activity (Lakatos et al., 2009; Mercier et al., 2015). However, our findings are also consistent with an interpretation based on increased gain of exogenous neural responses to sound (Chait et al., 2010; Choi et al., 2013; Lange et al., 2003; Lange, 2009; Rimmele et al., 2011; Woldorff et al., 1993). This ambiguity is partly due to the fact that inter-trial phase coherence, one of our main neural metrics, can reflect either temporal alignment (i.e. degree of jitter) or signal-to-noise ratio (because the phase of stronger signals will be less

affected by noise, leading to greater consistency in phase across trials). Our current dataset cannot distinguish between these two explanations, since the rapid presentation rate that was necessitated by our interest in perception of streams of sound prevents us from isolating individual ERP components. (As a result, our use of the term “neural entrainment” would fall under the “broad sense” introduced by [Obleser and Kayser, 2019](#).) Indeed, even for experimental designs which enable isolation of ERP components, endogenous versus exogenous interpretations of neural modulations of EEG signals have been debated for decades ([Hillyard et al., 1973](#); [Näätänen et al., 1978](#)), as it can be very difficult to distinguish between modulation of an ERP component such as the N1 and an endogenous attention-driven component which happens to overlap in time with that same ERP component. However, this auditory selective attention paradigm could be modulated in ways that could enable arbitration between these two potential theoretical explanations in the future. For example, collecting ERPs to individual tones in a separate paradigm could enable comparison of the topography of ERPs to the topography of the within-band-rate phase-locking signal ([Henry et al., 2017](#)).

Given our use of relatively low-density EEG, our results do not provide information about the specific network of brain areas contributing to auditory object selection. Prior work has shown that frontal and motor signals can modulate the phase of auditory cortex activity ([Park et al., 2015](#); [Morillon and Schroeder, 2015](#); [Morillon and Baillet, 2017](#)). Moreover, rhythmic movements can enhance temporally-selective attention ([Morillon and Baillet, 2017](#)). One possible explanation for our findings, therefore, is that participants who are better able to use temporal information when selecting auditory objects have stronger auditory-motor neural connectivity. According to this explanation short-term attention training and long-term musical training enhance the motor system’s control over auditory neural timing, thereby enhancing temporally-selective attention. Future work using our nonverbal auditory selective attention paradigm in combination with cognitive neuroscience techniques with more spatial precision (such as MEG and fMRI) could test this hypothesis. Another way to test this hypothesis would be to investigate whether auditory selective attention training enhances the ability to align motor movements with temporal sound patterns.

We have interpreted our finding of increased ITPC at 4 Hz in the attention conditions and a shift in average neural phase angle at 4 Hz between the attend high and attend low conditions as reflecting neural entrainment to the rhythmic structure of the attended tone sequence. However, given that our analysis windows included the pause between sequences, our results could reflect a combination of entrainment to the stimulus structure and maintenance of the entrainment through the silent period. Both of these processes are relevant when directing selective attention to ecologically valid sound sequences. Speech, for example, often includes pauses or breaks in the rhythm, followed by a somewhat temporally predictable sound onset. Listeners expect pauses to be present at the boundaries between linguistic phrases ([Hirsh-Pasek et al., 1987](#)), and so these pauses are an integral part of the rhythms listeners perceive when listening to speech, which can facilitate listening in cocktail party paradigms ([Gordon-Salant and Fitzgibbons, 2004](#); [Best et al., 2007](#)). Maintaining rhythms during speech pauses is also relevant to the ability to produce speech at the appropriate time after a conversational pause, a process which has been shown to be enhanced after rhythmic training ([Hidalgo et al., 2019](#)). Future work could try to disentangle these two processes to investigate how they individually relate to task performance and musical experience, as well as how they are altered by short-term training.

We have demonstrated a link between the direction of attention and modulations of average neural phase angle, but it remains to be seen whether neural entrainment plays a causal role in supporting attention to temporally distinct non-verbal sound streams. One way to test the causal nature of the link between neural entrainment and attention would be to experimentally manipulate neural phase using transcranial magnetic stimulation (TMS). Prior studies using a similar approach have found that

manipulating neural entrainment while participants listen to speech in competing speech can modulate speech recognition performance ([Riecke et al., 2018](#); [Zoefel et al., 2018](#)).

In our selective attention paradigm the sound streams were separated both in frequency and in time, with the result that the task could be completed either by directing attention to frequency, to time, or both. The medium-strength links we find between top-down modulation of average neural phase angle and selective attention performance suggest that direction of attention to task-relevant time points may be a useful strategy for auditory selective attention, even when spectral attention could also be used. Nevertheless, some listeners with poor temporal processing abilities may have relatively unaffected auditory selective attention, as long as they are able to direct attention to alternate task-relevant dimensions. For example, in ecologically valid listening situations such as a cocktail party there are many different cues available which can be used to select sound streams, including spectral, spatial, and visual information as well as temporal information ([Darwin et al., 2003](#); [Best et al., 2007](#); [Kitterick et al., 2010](#); [Lee and Humes, 2012](#)).

There are some limitations in our study design which could be clarified by future work. For example, our phase-based analysis cannot distinguish between contributions from neural enhancement of the attended stimulus stream and neural suppression of the ignored stimulus stream ([Chait et al., 2010](#)). Future work could distinguish between these two possibilities by presenting the two streams at different rates—for example, 4 Hz and 3 Hz—to determine whether phase locking to a given stream is enhanced when it is attended and suppressed when it is ignored relative to a passive condition. One possibility is that target enhancement is driven by neural entrainment within the theta range (3–8 Hz), while distractor suppression is driven by an alternate mechanism, linked to increased alpha power ([Wöstmann et al., 2019](#)). Also, because we did not include control training in our design, we cannot draw any conclusions about the efficacy of this particular selective attention training program relative to any alternate form of training. Future work, therefore, could use this auditory selective attention paradigm to examine the factors which modulate the efficacy of attention training programs. Moreover, future work could examine the extent to which selective attention enhancements are maintained long-term if training is discontinued.

In conclusion, we demonstrate that asking participants to attend to one of two simultaneously presented sound streams leads to a modulation of average neural phase angle and that this metric is linked to individual differences in selective attention performance. Our results further suggest that attentional control and top-down modulation of average neural phase angle are tied to long-term experience and are capable of rapid short-term improvement.

Declaration of competing interest

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

CRediT authorship contribution statement

Aeron Laffere: Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Fred Dick:** Conceptualization, Methodology, Writing - review & editing. **Adam Tierney:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing, Visualization, Supervision.

Acknowledgements

The authors are grateful to Maria Chait, Alex Billig, Sijia Zhao, and Martin Eimer for comments on earlier versions of the manuscript. The authors would also like to thank the participants for generously donating their time. This research was funded in part by the Waterloo Foundation, United Kingdom, grant #917/3367, and the National Institutes of Health United States, NIH NIDCD R01DC017734.

References

- Berens, Philipp, 2009. CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software* 31, 1–21.
- 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.
- 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.
- Cirelli, Laura, Spinelli, Crinstina, Nozaradan, Sylvie, Trainor, Laurel, 2016. Measuring neural entrainment to beat and meter in infants: effects of music background. *Frontiers in Neuroscience* 10, 229.
- Clayton, K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A., Kidd, G., 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLoS One* 11, e0157638.
- Darwin, C., Brungart, D., Simpson, B., 2003. Effects of fundamental frequency and vocal tract length changes on attention to one of two simultaneous talkers. *JASA* 114, 2913–2922.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.
- Deroche, M., Limb, C., Chatterjee, M., Gracco, V., 2017. Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. *JASA* 142, 1739–1755.
- 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., 2012. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89.
- Doelling, K., Poeppel, D., 2015. Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. Unit. States Am.* 112, E6233–E6242.
- Gordon-Salant, S., Fitzgibbons, P., 2004. Effects of stimulus and noise rate variability on speech perception by younger and older adults. *JASA* 115, 1808–1817.
- Habibi, A., Cahn, B., Damasio, A., Damasio, H., 2016. Neural correlates of accelerated auditory processing in children engaged in music training. *Developmental Cognitive Neuroscience* 21, 1–14.
- Harding, E., Sammler, D., Henry, M., Large, E., Kotz, S., 2019. Cortical tracking of rhythm in music and speech. *NeuroImage* 185, 96–101.
- Henry, M., Herrmann, B., Künke, D., Obleser, J., 2017. Aging affects the balance of neural entrainment and top-down neural modulation in the listening brain. *Nat. Commun.* 8, 15801.
- Hidalgo, C., Pesnot-Lerousseau, J., Parquis, P., Roman, S., Schön, D., 2019. Rhythmic training improves temporal anticipation and adaptation abilities in children with hearing loss during verbal interaction. *J. Speech Lang. Hear. Res.* 62, 3234–3247.
- Hillyard, S., Hink, R., Schwent, V., Picton, T., 1973. Electrical signs of selective attention in the human brain. *Science* 182, 177–180.
- Hirsh-Pasek, K., Nelson, D., Jusczyk, P., Cassidy, K., Druss, B., Kennedy, L., 1987. Clauses are perceptual units for young infants. *Cognition* 26, 269–286.
- 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.
- Isbell, E., Stevens, C., Pakulak, E., Wray, A., Bell, T., Neville, H., 2017. Neuroplasticity of selective attention: research foundations and preliminary evidence for a gene by intervention interaction. *Proc. Natl. Acad. Sci. Unit. States Am.* 114, 9247–9254.
- Iversen, J., Repp, B., Patel, A., 2009. Top-down control of rhythm perception modulates early auditory responses. *Ann. N. Y. Acad. Sci.* 1169, 58–73.
- Kerlin, J., Shahin, A., Miller, L., 2010. Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* 30, 620–628.
- Kitterick, P., Bailey, P., Summerfield, A., 2010. Benefits of knowing who, where, and when in multi-talker listening. *JASA* 127, 2498–2508.
- 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., O'Connell, M., Barczak, A., Mills, A., Javitt, D., Schroeder, C., 2009. The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430.
- Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., Schroeder, 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.
- Lakatos, P., Gross, J., Thut, G., 2019. A new unifying account of the roles of neuronal entrainment. *Curr. Biol.* 29, R890–R905.
- Lange, Kathrin, 2009. Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition* 69, 127–137.
- Lange, K., Rösler, F., Röder, 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.
- Lee, J., Humes, L., 2012. Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. *JASA* 132, 1700–1717.
- Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010.
- Madsen, S., Whiteford, K., Oxenham, A., 2017. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Sci. Rep.* 7, 12624.
- Meha-Bettison, K., Sharma, M., Ibrahim, R., Vasuki, P., 2017. Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians. *Int. J. Audiol.* 57, 40–52.
- Mercier, M., Molholm, S., Fiebelkorn, I., Butler, J., Schwartz, T., Foxe, J., 2015. Neuro-oscillatory phase alignment drives speeded multisensory response times: an electrocorticographic investigation. *J. Neurosci.* 35, 8546–8557.
- Morse-Fortier, C., Parrish, M., Baran, J., Freyman, R., 2017. The effects of musical training on speech detection in the presence of informational and energetic masking. *Trends in Hearing* 21, 1–12.
- Morillon, B., Schroeder, C., 2015. Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Ann. N. Y. Acad. Sci.* 1337, 26.
- Morillon, B., Baillet, S., 2017. Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci. Unit. States Am.* 114, E8913–E8921.
- Näätänen, R., Gaillard, A., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329.
- Nobre, A., van Ede, F., 2018. Anticipated moments: temporal structure in attention. *Nat. Rev. Neurosci.* 19, 34–48.
- Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240.
- Nozaradan, S., Peretz, I., Mouraux, A., 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32, 17572–17581.
- Nozaradan, S., Peretz, I., Keller, P., 2016. Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Sci. Rep.* 6, 20612.
- Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. *Trends Cognit. Sci.* 23, 913–926.
- 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.
- Oxenham, A., Fligor, B., Mason, C., Kidd, G., 2003. Informational masking and musical training. *JASA* 114, 1543–1549.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L., Hoke, M., 1998. Increased auditory cortical representation in musicians. *Nature* 392, 811–814.
- Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009. Musician enhancement for speech in noise. *Ear Hear.* 30, 653–661.
- Park, H., Ince, R., Schyns, P., Thut, G., Gross, J., 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25, 1649–1653.
- Power, A., Foxe, J., Forde, E., Reilly, R., Lalor, E., 2012. At what time is the cocktail party? A late locus of selective attention to natural speech. *Eur. J. Neurosci.* 35, 1497–1503.
- Puschmann, S., Baillet, S., Zatorre, R., 2019. Musicians at the cocktail party: neural substrates of musical training during selective listening in multispeaker situations. *Cerebr. Cortex* 29, 3253–3265.
- Riecke, L., Formisano, E., Sorger, B., Baskent, D., Gaudrain, E., 2018. Neural entrainment to speech modulates speech intelligibility. *Curr. Biol.* 28, 161–169.
- Rimmele, J., Jolsvai, H., Sussman, E., 2011. Auditory target detection is affected by implicit temporal and spatial expectations. *J. Cognit. Neurosci.* 23, 1136–1147.
- Rodrigues, A., Loureiro, M., Caramelli, P., 2013. Long-term musical training may improve different forms of visual attention ability. *Brain Cognit.* 82, 229–235.
- Ruggles, D., Freyman, R., Oxenham, A., 2014. Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One* 9, e86980.
- Schneider, P., Sluming, V., Roberts, N., Bleeck, S., Rupp, A., 2005. Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann. N. Y. Acad. Sci.* 1060, 387–394.
- Schroeder, C., Wilson, D., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176.
- Seither-Preisler, A., Parncutt, R., Schneider, P., 2014. Size and synchronization of auditory cortex promotes musical, literacy and attentional skills in children. *J. Neurosci.* 34, 10937–10949.
- Shahin, A., Bosnyak, D., Trainor, L., Roberts, L., 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552.
- Shamma, S., Elhilali, M., Micheyl, C., 2011. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123.
- Shinn-Cunningham, B., 2008. Object-based auditory and visual attention. *Trends Cognit. Sci.* 12, 182–186.
- Slater, J., Kraus, N., 2016. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognit. Process.* 17, 79–87.
- Stevens, C., Fanning, J., Coch, D., Sanders, L., Neville, H., 2008. Neural mechanisms of selective auditory attention are enhanced by computerized training: electrophysiological evidence from language-impaired and typically developing children. *Brain Res.* 1205, 55–69.
- Stevens, C., Harn, B., Chard, D., Currin, J., Parisi, D., Neville, H., 2013. Examining the role of attention and instruction in at-risk kindergarteners: electrophysiological measures

- of selective auditory attention before and after an early literacy intervention. *J. Learn. Disabil.* 46, 73–86.
- Swaminathan, J., Mason, C., Streeter, T., Best, V., Kidd, G., Patel, A., 2015. Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5, 11628.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1996. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J. Neurosci.* 16, 4240–4249.
- Tierney, A., Kraus, N., 2014. Neural entrainment to the rhythmic structure of music. *J. Cognit. Neurosci.* 27, 400–408.
- Tierney, A., Krizman, J., Kraus, N., 2015. Music training alters the course of adolescent auditory development. *Proc. Natl. Acad. Sci. Unit. States Am.* 112, 10062–10067.
- Tierney, Adam, Rosen, Stuart, Dick, Fred, et al., 2020. Speech-in-speech perception, nonverbal selective attention, and musical training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000767>. In press.
- Whitton, J., Hancock, K., Polley, D., 2014. Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proc. Natl. Acad. Sci. Unit. States Am.* 111, E2606–E2615.
- Whitton, J., Hancock, K., Shannon, J., Polley, D., 2017. Audiomotor perceptual training enhances speech intelligibility in background noise. *Curr. Biol.* 27, 3237–3247.
- 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. Unit. States Am.* 90, 8722–8726.
- Wöstmann, M., Alavash, M., Obleser, J., 2019. Alpha oscillations in the human brain implement distractor suppression independent of target selection. *J. Neurosci.* 39, 9797–9805.
- Zendel, B., Alain, C., 2013. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *J. Cognit. Neurosci.* 25, 503–516.
- 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.
- Zoefel, B., Archer-Boyd, A., Davis, M., 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Curr. Biol.* 28, 401–408.