NeuroImage 244 (2021) 118544

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Dimension-selective attention and dimensional salience modulate cortical tracking of acoustic dimensions

Ashley E. Symons[a,*], Fred Dick[a,b], Adam T. Tierney[a]

[a] Department of Psychological Sciences, Birkbeck College, University of London UK
[b] Division of Psychology & Language Sciences, University College London UK

ARTICLE INFO

Keywords:
Auditory
Attention
Salience
EEG
Cortical tracking

ABSTRACT

Some theories of auditory categorization suggest that auditory dimensions that are strongly diagnostic for particular categories - for instance voice onset time or fundamental frequency in the case of some spoken consonants - attract attention. However, prior cognitive neuroscience research on auditory selective attention has largely focused on attention to simple auditory objects or streams, and so little is known about the neural mechanisms that underpin dimension-selective attention, or how the relative salience of variations along these dimensions might modulate neural signatures of attention. Here we investigate whether dimensional salience and dimension-selective attention modulate the cortical tracking of acoustic dimensions. In two experiments, participants listened to tone sequences varying in pitch and spectral peak frequency; these two dimensions changed at different rates. Inter-trial phase coherence (ITPC) and amplitude of the EEG signal at the frequencies tagged to pitch and spectral changes provided a measure of cortical tracking of these dimensions. In Experiment 1, tone sequences varied in the size of the pitch intervals, while the size of spectral peak intervals remained constant. Cortical tracking of pitch changes was greater for sequences with larger compared to smaller pitch intervals, with no difference in cortical tracking of spectral peak changes. In Experiment 2, participants selectively attended to either pitch or spectral peak. Cortical tracking was stronger in response to the attended compared to unattended dimension for both pitch and spectral peak. These findings suggest that attention can enhance the cortical tracking of specific acoustic dimensions rather than simply enhancing tracking of the auditory object as a whole.

1. Introduction

Auditory categorization requires the mapping of continuous acoustic dimensions onto discrete categories. A central issue in theoretical accounts of auditory categorization is how listeners dynamically integrate and weight information from different acoustic dimensions. Much of the previous work on auditory categorization has focused on speech perception (Francis et al., 2000; Francis and Nusbaum, 2002; Gordon et al., 1993; Heald and Nusbaum, 2014; Idemaru et al., 2012; Idemaru and Holt, 2011; Jasmin et al., 2019; Jasmin et al., 2021; Kim et al., 2018; Kong and Edwards, 2016). Prior work in this domain has shown that listeners weight acoustic dimensions according to the reliability with which each dimension distinguishes between categories (Holt et al., 2018; Toscano and McMurray, 2010). When the reliability of an acoustic dimension changes due to noise (Winn et al., 2013) or short-term changes in cue distribution (Idemaru and Holt, 2011, 2014), listeners dynamically reweight acoustic dimensions accordingly. Listeners show stable individual differences in dimensional weighting strategies (Idemaru et al., 2012; Kim et al., 2018; Kong and Edwards, 2016), which may reflect differences in auditory perceptual ability (Jasmin et al., 2019, 2020) and prior language experience (Jasmin et al., 2021). Despite the central role of dimensional weighting in speech perception, and auditory category learning more generally (Holt and Lotto, 2006), surprisingly little is known about the neural mechanisms underlying this process.

One process that may contribute to the flexibility of dimensional weighting strategies across different contexts as well as the variability between individuals is dimension-selective attention. According to some theoretical accounts of speech perception, for example, listeners dynamically allocate attentional resources towards dimensions that are informative, and away from those that are less informative (Francis et al., 2000; Francis and Nusbaum, 2002; Gordon et al., 1993; Heald and Nusbaum, 2014; Holt et al., 2018). Thus, dimension-selective attention has been suggested to play a potential role in dimensional weighting in speech perception. Long-term prior experience with speech and language may change the salience of different acoustic dimensions, potentially accounting for the differences in dimensional weighting strategies across speakers of different languages (Jasmin et al., 2021). Al-

* Corresponding author.
E-mail address: a.symons@bbk.ac.uk (A.E. Symons).

https://doi.org/10.1016/j.neuroimage.2021.118544.
Received 5 May 2021; Received in revised form 19 August 2021; Accepted 31 August 2021
Available online 4 September 2021.
1053-8119/

though there is little direct evidence for the involvement of attention in auditory category learning, eye-tracking studies have provided extensive evidence that the salience of dimensions changes during visual category learning, with increased salience for dimensions that distinguish between categories (Blair et al., 2009, 2009; Carvalho and Goldstone, 2017; Chen et al., 2013; Kim and Rehder, 2011; Rehder and Hoffman, 2005a, 2005b; Zaki and Salmi, 2019). Thus, there is robust evidence that dimension-selective attention plays a crucial role in visual category learning. However, despite the central role that dimension-selective attention plays in many theories of speech perception, there is little direct evidence that auditory dimension-selective attention exists, let alone what neural mechanisms might subserve it.

How might auditory dimension-selective attention be carried out in the brain? One possibility is that neural activity tracks variations in different acoustic dimensions, with attention enhancing representations of attended dimensions and potentially suppressing representations of unattended dimensions. Attention-driven enhancement of cortical tracking has been demonstrated in studies of object-based auditory selective attention. For instance, prior work has shown that neural activity synchronizes with low-frequency fluctuations in the amplitude (Ding et al., 2016; Horton et al., 2013; Luo and Poeppel, 2007; Obleser and Kayser, 2019; Zoefel et al., 2018) and pitch contour (Teoh et al., 2019) of continuous speech. Attention has been shown to enhance this cortical tracking for attended versus ignored speech streams (Ding et al., 2016; Kerlin et al., 2010; Reetzke et al., 2021; Zion Golumbic et al., 2013). Moreover, studies using non-verbal stimuli have shown that attention strengthens cortical tracking to target relative to distractor tone sequences (Elhilali et al., 2005; Laffere et al., 2020; Laffere et al., 2021). Overall, these studies suggest that attention can modulate the extent to which the auditory system tracks variations in different sound streams. Here we suggest that a similar mechanism could underlie attention to acoustic dimensions within a single sound stream.

There is some initial evidence that attention may modulate the cortical tracking of different acoustic dimensions. For example, Costa-Faidella et al. (2017) had participants listen to single auditory streams that varied in duration and intensity at different rates; they either responded as to whether five consecutive tones were long or short (duration task) or silently counted the number of loud tones (intensity task). Cortical tracking was stronger for the attended compared to the ignored dimension. However, the dimension to which attention was directed was not independent from a difference in task across conditions, and so the condition effect on neural tracking could have partially reflected task demands. Moreover, in the absence of a passive listening condition, it is not possible to discern whether this finding reflects neural enhancement of the attended dimension, or suppression of the unattended dimension. Therefore, it is still unclear whether the listeners are increasing the gain on the attended dimension or actively inhibiting task-irrelevant information in unattended dimension. Disentangling the effects of enhancement versus suppression will provide crucial insight into the mechanisms underpinning auditory dimension-selective attention.

In addition to being modulated by top-down attention, cortical tracking of acoustic dimensions may also be enhanced by bottom-up attentional salience. Multiple acoustic dimensions contribute to the perceptual salience of natural sounds (Huang and Elhilali, 2017; Zhao et al., 2019). Salient changes in one or more of these acoustic dimensions can modulate physiological measures of attentional orienting such as skin conductance response (Siddle et al., 1984) and pupil dilation response (Bala and Takahashi, 2000; Liao et al., 2016; Marois et al., 2018; Wetzel et al., 2016; but see Zhao et al., 2019), with the magnitude of the response varying in proportion to the size of the change (Marois et al., 2018; Wetzel et al., 2016). Prior EEG work has also shown that both the mismatch negativity (MMN) and P3 responses, associated with detection of acoustic change and orientation of attention respectively, are sensitive to the magnitude of the change along multiple acoustic dimensions (Berti et al., 2004; Escera et al., 1998; Rinne et al., 2006; Schröger, 1996). Salient acoustic changes can also influence the degree of cortical tracking of acoustic streams, with reduced tracking of attended streams following salient background sounds (Huang and Elhilali, 2020) and increased tracking of acoustic melodies following deviations in pitch, timbre and intensity (Kaya et al., 2020). These studies suggest that salient changes in a sound stream along a number of different dimensions can attract attention to the sound stream. However, it remains unclear whether dimensional salience can attract attention to specific acoustic dimensions within a sound stream, as revealed by enhanced cortical tracking of acoustic dimensions with high versus low salience.

Taken together, the studies reviewed above provide compelling evidence that salience and selective attention can enhance the neural representation of auditory objects. However, there has been less empirical work investigating whether different acoustic dimensions within a single auditory object can attract attention, even though dimensional salience and dimension-selective attention play important roles in theoretical accounts of auditory categorization. One possible reason for this may stem from the difficulty in finding a reliable measure of auditory dimensional salience and dimension-selective attention. Therefore, our aim was to establish a neural measure of auditory dimensional salience and dimension-selective attention.

In two experiments, we used a frequency-tagging approach – where changes in different dimensions are tagged to fixed rates (e.g., Ding et al., 2016; Nozaradan et al., 2011) – to investigate whether the cortical tracking of acoustic dimensions is modulated by dimensional salience (Experiment 1) and dimension-selective attention (Experiment 2). In each experiment, listeners heard sequences of synthesized complex tones that varied in pitch (fundamental frequency) and in spectral peak frequency, each at a given fixed rate. If attention can be directed to specific acoustic dimensions, we would expect attention to elicit an increase in cortical tracking at the frequency tagged to that dimension. In contrast, if attention operates solely at the object level, we would expect attention to enhance the cortical tracking of the entire auditory object instead (leading to increased tracking at the baseline stimulus presentation rate).

In Experiment 1, the salience of the pitch dimension was manipulated by altering the pitch step sizes (1 versus 2 semitones) between blocks, while the step sizes of the spectral peak frequency remained constant (2 semitones). Based on previous research showing that responses to pitch deviants vary as a function of step size (Berti et al., 2004; Marois et al., 2018; Wetzel et al., 2016), we hypothesized that cortical tracking would be modulated by increased salience along the pitch dimension. Thus, we predicted that stronger cortical tracking at the rate tagged to pitch changes would be observed in blocks with larger pitch step sizes (2 semitones) compared to smaller pitch step sizes (1 semitone).

In Experiment 2, the tone sequences were identical across conditions, while only the focus of attention varied. In two attention conditions, listeners either attended to variations in the pitch, or the spectral peak frequency dimension, while ignoring variations in the other dimension. In a third ‘neutral’ condition, listeners monitored for occasional quiet tones. This condition was designed to test whether the effects of attention were due to attentional enhancement of the attended dimension, or suppression of the unattended dimension. We predicted that dimension-selective attention would result in enhanced cortical tracking of a dimension when it was attended, and potentially suppressed cortical tracking of a dimension when it was unattended.

2. Experiment 1

2.1. Methods

2.1.1. Participants

We recruited and tested twenty-nine participants (14 female, 15 male) between the ages of 19–59 with no known hearing impairments

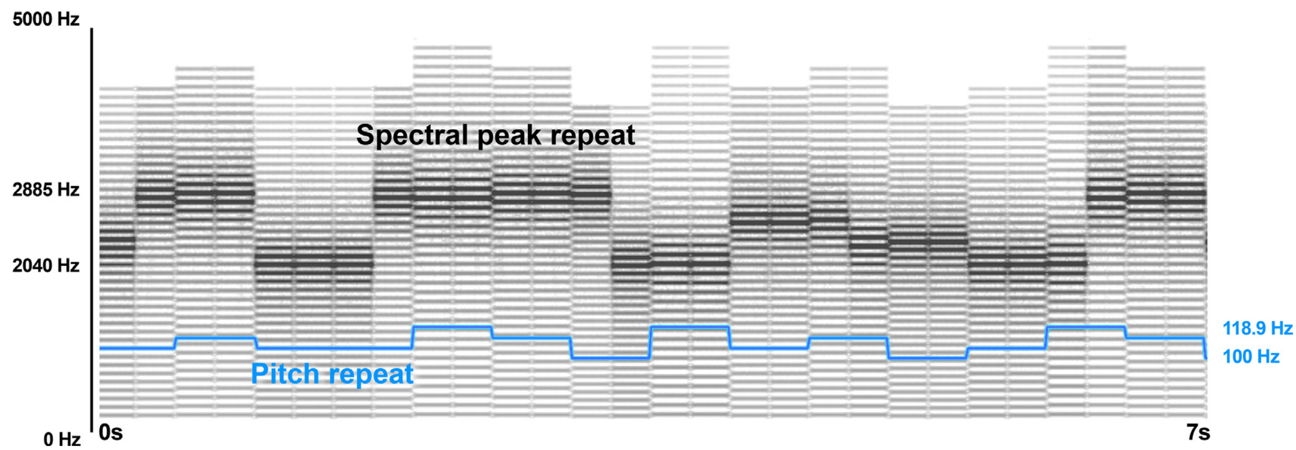

Fig. 1. Spectrogram showing a cropped portion of a stimulus varying in pitch (blue line) and spectral peak. In this example, pitch step sizes are separated by 1 semitone (min = 100 Hz, max = 118.9 Hz) and spectral peak step sizes are separated by 2 semitones (min = 2040 Hz, max = 2885 Hz).

and no diagnosis of a language or learning disorder. Since our hypothesis for this experiment was directional (in that we predicted stronger cortical tracking of pitch in the 2-semitone compared to 1-semitone condition), we conducted a post hoc power analysis using the *pwr* package in R (Champely et al., 2020) for a one-tailed test of whether cortical tracking was greater in the 2-semitone versus 1-semitone step size condition (one-tailed) assuming a medium effect size ($d = 0.5$, $\alpha = 0.05$, $\beta = 0.8$). A sample size of 27 provided sufficient power for this test.

Three participants were subsequently excluded. Two of the three were excluded as a result of insufficient EEG data, one due to excessive EEG artefacts resulting in the loss of over 50% of trials in multiple conditions and another due to a technical error that resulted in the loss of data from one condition. The third participant was excluded on the basis of poor behavioral performance (< 10% hit rate and >100 false alarms).

The final sample consisted of 26 participants (12 female, 14 male) with a mean age of 31.65 years (standard deviation = 10.06 years). Native languages spoken by participants included English (19), Polish (2), Greek (1), Japanese (1), Mandarin (1), Romanian (1), and Russian/Uzbek (1). Twelve participants reported receiving some form of musical training, ranging from 3 to 20 years (mean = 10.96, standard deviation = 5.60).

The Ethics Committee in the Department of Psychological Sciences at Birkbeck, University of London approved all experimental procedures. Informed consent was obtained from all participants. Participants were compensated for their participation in the form of course credits, or payment at a standard rate.

2.1.2. Design

In this experiment, participants listened to isochronous sequences of complex tones that varied in pitch and spectral peak frequency at different rates (see Fig. 1). The rates at which the pitch and spectral peak changed were counterbalanced within participants to ensure that any effect of dimensional salience was not due to differences in cortical tracking of different presentation rates. Pitch salience was manipulated by varying the pitch step sizes. This resulted in an experimental design with two pitch step size conditions (1 versus 2 semitones) and two varying acoustic dimensions (pitch versus spectral peak).

2.1.3. Stimuli

The stimuli consisted of 250 ms complex tones (40 harmonics and a 15-ms linear on/off ramp) with a single spectral peak (Smith, 2007). Two 2-dimensional stimulus spaces were created. In one, complex tones with one of four fundamental frequencies (each separated by two semitones, 91.13 Hz, 102.29 Hz, 114.82 Hz, 128.87 Hz) were modulated by one of four spectral peaks (again separated by two semitones, 2040.00 Hz, 2289.82 Hz, 2570.24 Hz, 2884.99 Hz). The second stimulus space only differed from the first in that the fundamental frequencies were separated by a single semitone (100 Hz, 105.94 Hz, 112.25 Hz, 118.92 Hz). Importantly, although the pitch step sizes varied, the mean fundamental frequency of the two spaces was the same (109.28 Hz) .

These two sets of tones were concatenated without pause to form 4 Hz tone sequences (480 tones, 120 s). Crucially, pitch and spectral peak changed at different rates (1.33 Hz and 2 Hz). When pitch varied at 2 Hz, spectral peak varied at 1.33 Hz. When pitch varied at 1.33 Hz, spectral peak varied at 2 Hz (Fig. 1). These rates were chosen because they fall within the range of frequencies observed in naturalistic speech. While the 4-Hz tone presentation rate is consistent with the syllable rate, the slower rates of dimensional change are consistent with acoustic modulations at the phrasal and sentential levels (Ding et al., 2016).

The order of the tones within each sequence was pseudorandomized such that tones had to change in pitch and spectral peak at the specified rate, apart from 20 repeated segments that occurred within each dimension. These repetitions were instances in which the dimension did not change at the expected rate. An example of a repetition would be if pitch was typically changing every 2 tones (2 Hz), but at the repetition did not change until after 4 tones. Repetitions were randomly inserted into the sequence, with the condition that two repetitions could not occur consecutively. The repetitions were included for comparability with Experiment 2 and are not directly relevant to Experiment 1; Experiment 1 participants were not alerted to their presence, and the repetitions were task-irrelevant.

The participants' explicit behavioral task was to respond to quiet 'oddball' tones, where the amplitude of 3–4 randomly selected tones was decreased by 25% (−12.0 dB, see below). Oddball timing was randomized in each sequence, with the exception that oddballs could not occur in the first or last 1.5 s of the sequence and could not occur consecutively. The same sequences were presented to all participants, but with the order of conditions counterbalanced across participants.

This resulted in four conditions, each consisting of four sequences that varied in a) pitch step size (1 semitone or 2 semitone difference between pitch steps) and b) dimension change rate (pitch at 1.33 Hz/spectral peak at 2 Hz, or pitch at 2 Hz/spectral peak at 1.33 Hz). Stimulus presentation was blocked by condition, but with the order of conditions counterbalanced across participants.

Stimuli were presented using PsychoPy3 (v 3.2.3) and the sound delivered via ER-3A insert earphones (Etymotic Research, Elk Grove Village, IL) at a level of 80 dB SPL.

2.1.4. Procedure

Prior to the EEG task, participants completed a short practice task outside of the EEG booth. Here, they listened to two short sequences

(48 s), each consisting of two oddball tones, and pressed the keyboard space bar when they detected those tones. Participants received visual feedback on their performance. If they failed to detect at least two of the four quiet oddball tones, or if they had more than four false alarms on the second sequence, the task was explained to them a second time and they were asked to complete a second practice block on a different set of sequences. All participants were able to pass the practice task.

In the main EEG task, participants listened to each 120-second sequence and responded via a keyboard press when they detected the oddball tones. This task was chosen to keep participants awake and alert throughout the task without focusing attention on one of the two dimensions of interest. In contrast to the training, participants did not receive visual feedback on their responses in this task.

2.1.5. EEG data acquisition and preprocessing

EEG data was recorded from 32 Ag-Cl active electrodes using a Biosemi™ ActiveTwo system with electrodes positioned according to the 10/20 montage. Data were recorded at a sampling rate of 16,384 Hz and digitized with 24-bit resolution. Two external reference electrodes were placed on the earlobes for off-line re-referencing. Impedance was kept below 20 kΩ. Triggers marking the beginning of each trial (every 6 tones, or 1.5 s) were recorded from trigger pulses sent to the data collection computer.

The data were down sampled to 512 Hz and re-referenced to the average of the earlobe reference electrodes. A low-pass zero-phase sixth-order Butterworth filter with a cutoff of 30 Hz was applied. A high-pass fourth-order zero-phase Butterworth filter with a cut-off of 0.5 Hz was then applied and the data epoched (1.5-seconds for analysis of phase and 30-seconds for analysis of signal amplitude) based on the recorded trigger pulses. Independent component analysis (ICA) was conducted to correct for eye blinks and horizontal eye movements. Components corresponding to eye blinks and movements were identified and removed based on visual inspection of the time courses and topographies.

Two measures of cortical tracking were computed: inter-trial phase coherence (ITPC) and signal amplitude. Both amplitude and ITPC provide complementary measures of cortical tracking. While analysis of signal amplitude over large time windows (30 s) provides fine-grained frequency resolution, analysis of ITPC over short time windows (1.5 s) allows for the exclusion of trials with excessive artifacts and behavioral responses which could not be done for the amplitude data. For these reasons, we report both measures.

Prior to calculation of ITPC, epochs containing artifacts exceeding +/- 100 μV were rejected. Additionally, trials in which participants made a response were excluded. A Hanning-windowed fast Fourier transform was then applied to each 1.5-second epoch. The complex vector at each frequency was converted to a unit vector and averaged across trials. The length of the average vector provides a measure of ITPC, which ranges from 0 (no phase consistency) to 1 (perfect phase consistency).

For analysis of signal amplitude, the 30 s epochs were averaged for each condition and transformed into the frequency domain using a fast Fourier transform. The resulting frequency spectrum represents the amplitude (in microvolts) at each frequency, with a frequency resolution of 1/30 (0.033 Hz). The frequency spectrum was then normalized by taking the difference between the amplitude at each frequency and the mean amplitude of the four neighboring frequencies (Nozaradan et al., 2011) to reduce the contribution of noise and other ongoing neural activity from the EEG signal.

All EEG data processing and analysis were carried out in Matlab (MathWorks, Inc) using the FieldTrip M/EEG analysis toolbox (Oostenveld et al., 2011) in combination with in-house scripts.

2.1.6. Data analysis

2.1.6.1. Behavioral data. The primary purpose of the behavioral task was to keep participants alert throughout the presentation of the stimuli, without having them explicitly attend to either of the two dimensions. Therefore, the task was designed to be easy, and performance was expected to be near ceiling. Nonetheless, the proportion of hits and false alarms was calculated for the 2-semitone and 1-semitone pitch step size conditions. The proportion of false alarms was defined as the number of responses that occurred outside of 1.25 s following an oddball, divided by the total number of non-oddball tones occurring outside of the target time windows. The proportion of hits and false alarms was converted to d-prime for analysis, with the loglinear approach used to avoid infinite scores (Hautus, 1995). Wilcoxon signed-rank tests (two-tailed) were used to test for an effect of pitch step size on performance. Effect size (r) was estimated by dividing the z-statistic by the square root of the sample size (Fritz et al., 2012).

2.1.6.2. EEG data. We extracted data from the 9 channels with the maximum signal when averaged across all 4 conditions, the 2 frequencies (1.33 Hz and 2 Hz) relevant for assessment of cortical tracking of stimulus dimensions, and all 26 participants. (Note that this choice of channels, which was based on collapsing across conditions, was orthogonal to our analysis, which was a comparison between conditions.) This resulted in the same set of frontocentral channels for both amplitude and ITPC (AF3, AF4, F3, F4, Fz, FC1, FC2, Cz, C3). The data were averaged across these channels prior to statistical analysis. We then collapsed the data across the two different rates of dimension change to determine the degree of cortical tracking of pitch and spectral peak dimensions. Wilcoxon signed-rank tests (two-tailed) were used to compare the effect of pitch step size on cortical tracking of pitch and spectral peak. False Discovery Rate was used to correct for multiple comparisons using the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995). Effect size (r) was estimated by dividing the z-statistic by the square root of the sample size (Fritz et al., 2012). Processed data and stimuli are available at: osf.io/c5s6u/.

2.2. Results

2.2.1. Behavioral

As expected, performance was near ceiling in both the small pitch step size (median d-prime = 5.106) and large pitch step size (median d-prime = 4.599) conditions. The difference between the two conditions was statistically significant ($W = 24$, $z = -3.186$ $p = 0.002$, $r = 0.625$). The difference in behavioral performance was unexpected but could suggest that larger pitch step sizes are more distracting, resulting in worse performance.

2.2.2. EEG

Fig. 2 displays cortical tracking across a range of frequencies (Fig. 2A) and averaged across frequencies for each pitch step size condition (Fig. 2B). Peaks in amplitude are observed at the two rates of dimensional change (1.33 Hz and 2 Hz), their harmonics (e.g., 2.67 Hz), and the tone presentation rate (4 Hz). Signal amplitude at the rate of pitch change was larger in the 2-semitone versus 1-semitone pitch step size condition ($W = 300$, $z = 3.162$, $p_{(corrected)} = 0.002$, $r = 0.620$). At the rate of spectral peak change, there was no significant difference in amplitude between 2-semitone and 1-semitone pitch step size conditions ($W = 148$, $z = -0.689$, $p_{(corrected)} = 0.499$, $r = 0.137$). In fact, there was no significant difference in amplitude at the rate of pitch compared to spectral peak change in the 1-semitone condition ($W = 199$, $z = 0.597$, $p_{(corrected)} = 0.565$, $r = 0.117$). In contrast, in the 2-semitone condition, amplitude was significantly larger at the rate of pitch compared to spectral peak change ($W = 326$, $z = 3.820$, $p_{(corrected)} < 0.001$, $r = 0.749$).

As shown in Fig. 3, similar effects were observed for ITPC. At the rate of pitch change, ITPC was larger for 2-semitone compared to 1-semitone pitch step size conditions ($W = 286$, $z = 2.806$, $p_{(corrected)} = 0.008$, $r = 0.550$). No significant difference in ITPC between 2-semitone and 1-semitone pitch step size conditions was observed at the rate of spec-

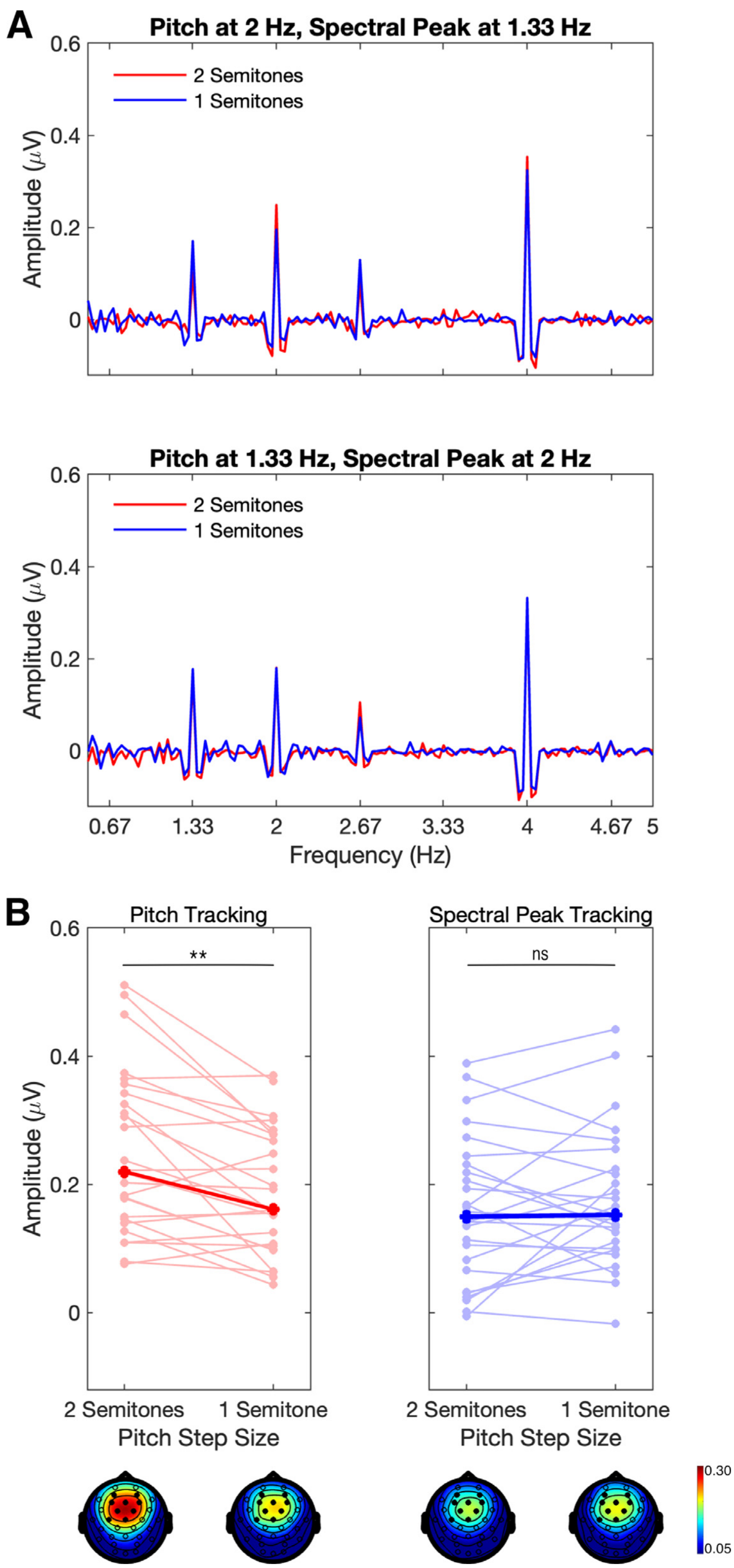

Fig. 2. (A) Median EEG signal amplitude across a range of frequencies (0.5 Hz – 5 Hz) for 2 semitone (red) and 1 semitone (blue) pitch step size conditions at each of the dimension change rates. (B) Amplitude at the frequencies corresponding to variations in the pitch (left) and spectral peak (right) dimensions for each pitch step size (x-axis) across the frontocentral channels selected for analysis. Individual participant data are represented by light red/blue lines and the median is represented by the dark red/blue lines. Topographical plots below the x axis with the channels selected for analysis filled in black.

tral peak change ($W = 109$, $z = -1.689$, $p_{(corrected)} = 0.094$, $r = 0.331$). When comparing pitch versus spectral tracking in the 1-semitone and 2-semitone conditions, there was no significant difference in ITPC in the 1-semitone condition ($W = 178$, $z = 0.064$, $p_{(corrected)} = 0.960$, $r = 0.012$). In the 2-semitone condition, ITPC at the rate of pitch change was larger compared to the rate of spectral peak change ($W = 315$, $z = 3.540$, $p_{(corrected)} < 0.001$, $r = 0.694$).

In summary, we observed enhanced pitch tracking for larger (2-semitones) compared to smaller (1-semitone) pitch step sizes. In contrast, spectral peak tracking was unaffected by pitch step size.

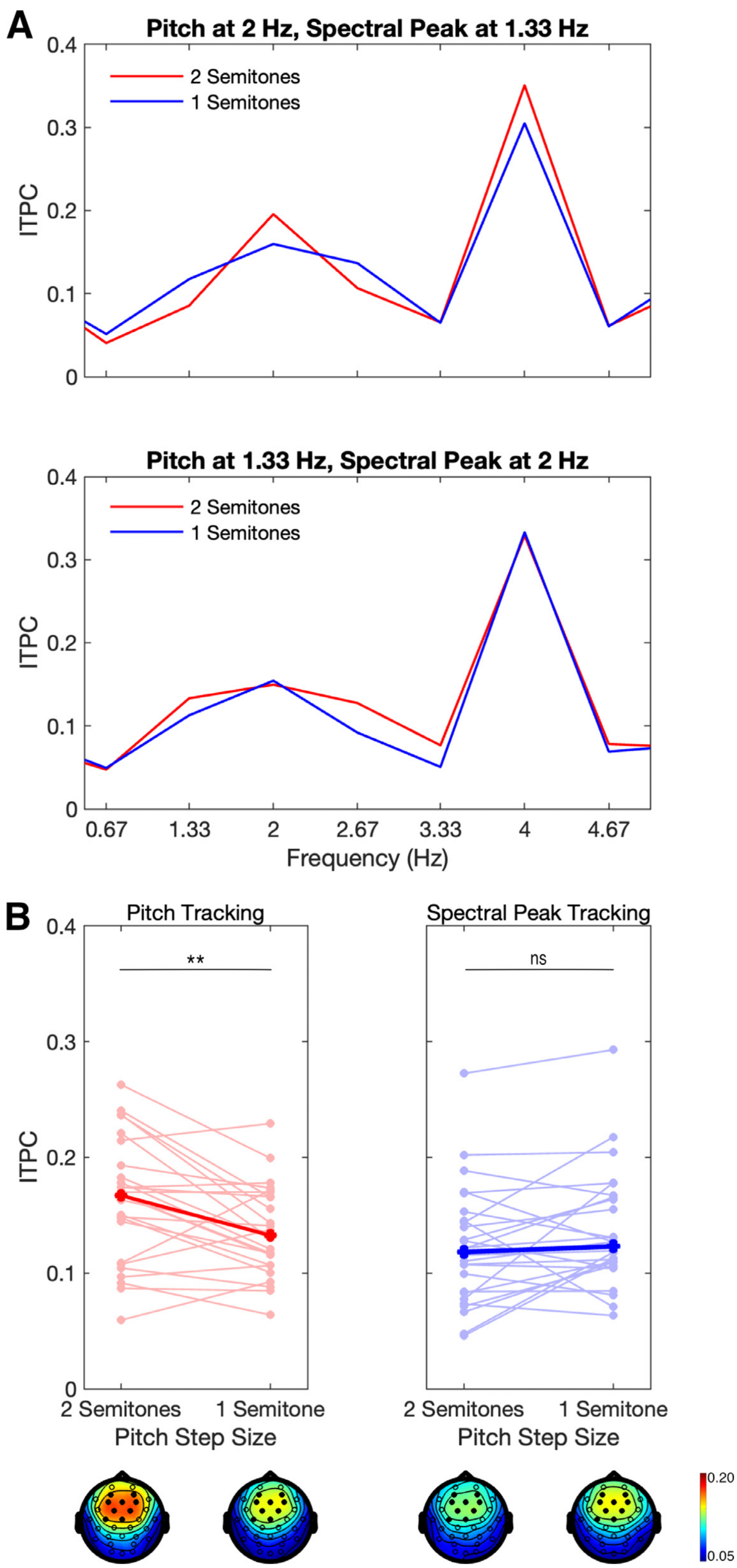

Fig. 3. (A) Median ITPC across a range of frequencies (0.5 Hz – 5 Hz) for 2 semitone (red) and 1 semitone (blue) pitch step size conditions at each of the dimension change rates. (B) ITPC at the frequencies corresponding to variations in the pitch (left) and spectral peak (right) dimensions for each pitch step size (x-axis) across the frontocentral channels selected for analysis. Individual participant data are represented by light red/blue lines and the median is represented by the dark red/blue lines. Topographic plots for each condition are displayed below the x-axis with selected channels filled in black.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Twenty-seven participants between the ages of 17–52 took part in the behavioral training tasks. This sample size was sufficient to provide 80% power to detect a medium effect size (d = 0.5) between conditions. Of these participants, 6 failed to reach the performance threshold required to pass the training. The final sample consisted of 21 participants (7 female, 14 male) between the ages of 17–52 (mean age = 34.71, standard deviation = 8.71) with no known hearing impairments and no language or learning disorders. Native languages spoken by participants included English (12), Italian (4), Arabic (1), Croatian (1), German (1), Hungarian (1), and Portuguese (1). Due to the challenging nature of the task, we primarily recruited individuals with musical training. Nineteen participants reported receiving some form of musical training, with the number of years of musical training ranging between 4 and 20 years (mean = 13.48, standard deviation = 4.87). This final sample size was sufficient to supply 71.5% power to detect a medium effect size. No further participants were excluded on the basis of behavioral or EEG data.

The Ethics Committee in the Department of Psychological Sciences at Birkbeck, University of London approved all experimental procedures. Informed consent was obtained from all participants. Participants were compensated for their participation in the form of course credits or payment at a standard rate.

3.1.2. Design

This experiment consisted of two phases: behavioral training and EEG testing. The behavioral training consisted of two tasks. The single dimension training task involved attending to variations in one dimension while the other dimension remained constant. The dimension-selective attention training task involved attending to one dimension while ignoring variations in the other dimension. In the EEG testing phase, participants completed a longer version of the dimension-selective attention task while EEG was recorded. In each task, there were three attention conditions: attend pitch, attend spectral peak, and a 'neutral' attention condition where participants performed the amplitude oddball detection task of Experiment 1, e.g., detecting occasional oddball tones that were reduced in amplitude relative to the other tones in the sequence. The rates of pitch and spectral peak change were counterbalanced within participants to ensure that any effect of attention was not due to differences in cortical tracking of different presentation rates. This resulted in an experimental design with three attention conditions (attend pitch, attend spectral peak, neutral) and, for neural measures, two phase-locked dimensions (pitch and spectral peak).

3.1.3. Stimuli

The results from Experiment 1 suggested that the salience of the pitch and spectral peak dimensions was relatively balanced, on average, in the 1-semitone pitch step size condition. Therefore, in this experiment, we used pitch step sizes of 1 semitone (100 Hz, 105.94 Hz, 112.25 Hz, 118.92 Hz) while the spectral peak step sizes varied by 2 semitones (2040.00 Hz, 2289.82 Hz, 2570.24 Hz, 2884.99 Hz).

3.1.3.1. Training stimuli. For the single dimension training, we created 48-second sequences in which only one dimension varied while the other dimension remained constant. In the attend pitch condition, the spectral peak was constant at 2040 Hz while F0 varied at either 1.33 Hz or 2 Hz. In the attend spectral peak condition, F0 was constant at 100 Hz while only the spectral peak varied every two (2 Hz) or three (1.33 Hz) tones. Eight repetitions, or instances in which the attended dimension did not change at the expected rate, were inserted into each sequence. In the 'neutral' attention condition, F0 was held constant at 100 Hz and spectral peak at 2040 Hz. In this condition only, two randomly selected tones were reduced in amplitude by −12 dB. This resulted in five conditions: attend pitch (pitch varying at 1.33 Hz), attend pitch (pitch varying at 2 Hz), attend spectral peak (spectral peak varying at 1.33 Hz), attend spectral peak (spectral peak varying at 2 Hz), and neutral attention (amplitude oddball detection). Five tone sequences were generated for each condition.

For the dimension-selective attention training, 48-second sequences were created in which the two dimensions varied at different rates. In half of the sequences, pitch varied at 2 Hz and spectral peak varied at 1.33 Hz; in the other half of the sequences, spectral peak varied at 2 Hz and pitch at 1.33 Hz. Eight repetitions along each dimension were inserted into each training sequence. In addition, 3–4 tones were decreased in amplitude by −12 dB in all sequences. The tones that were decreased in amplitude were randomly selected, with the exception that they could not occur during a repetition, or at the tone immediately preceding or following a repetition. Six conditions were created by crossing attention condition (pitch, spectral peak, neutral) with dimension change rate (pitch and spectral peak): attend pitch (pitch at 1.33 Hz/spectral peak at 2 Hz), attend pitch (pitch at 2 Hz/spectral peak at 1.33 Hz), attend spectral (pitch at 1.33 Hz/spectral peak at 2 Hz), attend spectral (pitch at 2 Hz/spectral peak at 1.33 Hz), neutral (pitch at 1.33 Hz/spectral peak at 2 Hz), and neutral (pitch at 2 Hz, spectral peak at 1.33 Hz). Five tone sequences were generated for each condition.

3.1.3.2. EEG stimuli. The characteristics of the stimuli for the main EEG task were identical to those of the small pitch step size condition of Experiment 1. As with Experiment 1, the 250 ms tones were concatenated into 120 s sequences with pitch and spectral peak dimensions varying at different rates (1.33 Hz or 2 Hz). Twenty repetitions (instances in which the dimension did not change at the expected rate) were inserted into each sequence; these were task-relevant for the pitch and spectral attention conditions. Additionally, 3–4 randomly selected tones were decreased in volume by −12 dB; as previously, these were task-relevant in the neutral conditions. The locations of the quiet oddball tones were pseudorandomly selected, with the exceptions that the oddball tones could not occur in the first or last 1.5 s of each sequence, nor could they occur during or immediately before or after a repetition. For this experiment we also increased the minimum temporal interval between successive oddball tones such that there needed to be at least 6 standard-amplitude tones between successive oddball tones.

As with the dimension-selective training stimuli, six conditions were created by crossing attention condition (pitch, spectral peak, neutral) with dimension change rate (pitch and spectral peak): attend pitch (pitch at 1.33 Hz/spectral peak at 2 Hz), attend pitch (pitch at 2 Hz/spectral peak at 1.33 Hz), attend spectral (pitch at 1.33 Hz/spectral peak at 2 Hz), attend spectral (pitch at 2 Hz/spectral peak at 1.33 Hz), neutral (pitch at 1.33 Hz/spectral peak at 2 Hz), and neutral (pitch at 2 Hz, spectral peak at 1.33 Hz). Each of the six conditions consisted of four sequences of tones. Stimulus presentation was blocked by condition, with the order of conditions counterbalanced across participants.

3.1.4. Procedure

3.1.4.1. Behavioral training. Prior to the EEG task, participants completed two short behavioral training exercises outside of the EEG booth. In the first training exercise, participants listened to shortened sequences (48 s) in which only a single dimension varied. First, participants were familiarized with the two dimensions. The pitch dimension was described to participants as how 'high or low' the sound was. The spectral peak dimension was described to participants as 'brightness'. Participants were provided examples of the variations in each dimension that they could listen to multiple times before progressing to the task. For attend pitch and spectral peak conditions, participants were told which dimension to attend to, and how often that dimension would change. Their task was to press a button to respond when they detected repetitions in the attended dimension. In the neutral condition, participants were instructed to listen out for and respond to occasional quiet tones. Participants received feedback on their performance. Participant's total score was displayed on the screen, with +1 point for every correct detection and −1 for every false alarm. For the attend pitch and spectral peak conditions, if participants received a total score of < 7/8, they received another training block for that condition. If participants met or exceeded the threshold, they moved onto the next condition. For the neutral condition, if participants missed more than one quiet tone (out of two) or had more than one false alarm, they received another block of training on that condition.

In the second training exercise, participants listened to shortened sequences (48 s) in which *both* pitch and spectral peak dimensions changed at different rates. Quiet tones were also embedded in these training sequences. Prior to each sequence, participants were told which dimension to attend to, and how often that dimension would change. The task was the same as the single-dimension training, except that participants also had to ignore changes in the unattended dimension. Participants received visual feedback on their performance identical to that in the single dimension training. Participant's total score was displayed on the screen, with +1 point for every correct detection and −1 for every false alarm. For the attend pitch and spectral peak conditions, if participants received a total score of < 6/8, they received another training block for that condition. If participants met or exceeded the threshold, they moved onto the next condition. For the neutral condition, if participants missed more than one quiet tone (out of three or four) or had more than one false alarm, they received a second block of training on that condition.

Due to the COVID-19 pandemic, in-lab testing was suspended partway through the experiment. When testing could safely be resumed, the behavioral training tasks were moved from in-lab to online to minimize the amount of time during which the participants and researchers had to interact in the lab. Minor variations existed between the in-lab and online training and are described in the Supplementary Materials.

3.1.4.2. EEG testing. In the main EEG task, participants listened to 120-second sequences in which both pitch and spectral peak dimensions varied. There were 6 conditions (2 dimensional change rates x 3 attended dimensions). Stimulus presentation was blocked by attended dimension to minimize switching costs, with the order counterbalanced across participants. At the start of each block, participants were instructed to detect repetitions in the attended dimension (attend pitch and spectral peak conditions) or to detect the quiet tones. In the attend pitch and attend spectral blocks only, participants were also told how often they could expect the attended dimension to vary. For example, in an attend pitch block in which pitch varied at 2 Hz, participants would hear the instructions: "In this block, your task is to attend to pitch, which will change every 2 sounds. Press the trigger button when you hear a repetition in pitch." In contrast to the training, no visual feedback was provided. Participants made their responses by pressing the trigger on an Xbox One game controller. Stimuli were presented using PsychoPy3 (v 3.2.3) and the sound delivered via ER-3A insert earphones (Etymotic Research, Elk Grove Village, IL) at a level of 80 dB SPL. Each block lasted between 8 and 10 min. In between each sequence, participants had the option of taking a short self-paced break. The total duration of the EEG task was approximately 1 h.

3.1.5. Data acquisition and analysis

EEG data acquisition and preprocessing procedures were identical to those of Experiment 1.

3.1.5.1. Behavioral. In the attend pitch and attend spectral peak conditions the proportion of hits and false alarms was calculated for each attention condition. Hit rate was defined as responses occurring within 1.25 s following a repetition divided by the total number of repetitions. False alarm rate defined as the number of responses occurring outside of the 1.25 s target window, divided by the total number of non-repetitions (instances where the dimension changed at the expected rate). In the neutral condition, the proportion of hits and false alarms was computed in the same manner as in Experiment 1. All scores were converted to d-prime for analysis with the loglinear approach used to avoid infinite values (Hautus, 1995). Statistical analysis was conducted separately for the dimension-selective attention (attend pitch, attend spectral peak) and amplitude oddball detection tasks since behavioral performance on the two tasks was not directly comparable. A Wilcoxon signed-rank test was used to compare performance in the attend pitch and attend spectral peak conditions, with False Discovery Rate used to correct for multiple comparisons via the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995). Effect size (r) was estimated by dividing the z-statistic by the square root of the sample size (Fritz et al., 2012).

3.1.5.2. EEG. Signal amplitude was extracted over the 9 frontocentral channels identified in Experiment 1 (AF3, AF4, F3, F4, Fz, FC1, FC2, Cz, C3). The data were averaged across channels prior to statistical analysis. We then collapsed the data across the two different dimension change rates to determine the degree of cortical tracking of pitch and spectral peak dimensions in each attention condition. Wilcoxon signed-rank tests were used to compare pitch and spectral peak tracking across the different attention conditions (attend pitch, attend spectral peak, neutral). False Discovery Rate was used to correct for multiple comparisons using the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995). Effect size (r) was estimated by dividing the z-statistic by the square root of the sample size (Fritz et al., 2012). Processed data and stimuli are available at: osf.io/c5s6u/.

3.2. Results

3.2.1. Behavioral

3.2.1.1. Training. In the single dimension training, participants who passed the training required between 1 and 5 blocks of training to reach the performance threshold. Hit rate ranged from 0.63 to 1.0 (median = 1.0) in the attend spectral peak condition and 0.81–1.00 (median = 1.0) in the attend pitch condition. The number of false alarms ranged from 0 to 5 (median = 0) in the attend spectral peak condition and 0–2 (median = 0) in the attend pitch condition.[1] In the neutral condition, hit rate was at ceiling (1.0) for all participants with between 0 and 1 (median = 0) false alarms. In the dimension-selective attention training, hit rate ranged from 0.75 to 1.0 (median = 0.94) in the attend spectral peak condition and 0.86–1.0 (median = 0.94) in the attend pitch condition. The number of false alarms in both conditions ranged from 0 to 3 (median = 1). In the neutral condition, hit rate was at ceiling (1.0) for all participants with between 0 and 1 false alarms (median = 0). No significant differences were observed between attend pitch and attend spectral peak conditions on either task ($p > 0.05$). These data show that participants understood the task instructions and could selectively attend to the different acoustic dimensions.

3.2.1.2. EEG task. Fig. 4 shows task performance (d-prime) in attend pitch and attend spectral peak conditions. There was no significant difference in performance between the two conditions ($W = 162$, $z = 1.62$, $p = 0.111$, $r = 0.354$), suggesting that task difficulty was matched across the attend pitch (median d-prime = 3.216) and attend spectral peak (median d-prime = 2.914) tasks. In the neutral condition, performance was high (median d-prime = 4.857), comparable to the behavioral performance observed in Experiment 1.

3.2.2. EEG

Fig. 5 displays cortical tracking (amplitude) across a range of frequencies (Fig. 5A) and averaged across frequencies for each attention condition (Fig. 5B). Signal amplitude at the rate of pitch change was larger in the attend pitch compared to the attend spectral peak condition ($W = 188$, $z = 2.520$, $p_{(corrected)} = 0.017$, $r = 0.550$), as well as in the attend pitch compared to neutral condition ($W = 208$, $z = 3.215$, $p_{(corrected)} = 0.004$, $r = 0.702$). There was no significant difference in amplitude at the rate of pitch change between attend spectral peak and neutral conditions ($W = 162$, $z = 1.616$, $p_{(corrected)} = 0.133$, $r = 0.353$). At the rate of spectral peak change, amplitude was larger in the attend spectral peak compared to the attend pitch condition ($W = 187$, $z = 2.485$, $p_{(corrected)} = 0.017$, $r = 0.542$), and in the attend spectral peak compared to neutral condition ($W = 199$, $z = 2.902$, $p_{(corrected)} = 0.007$, $r = 0.633$). There was no significant difference in amplitude at the rate of spectral peak change between attend pitch and neutral conditions ($W = 123$, $z = 0.261$, $p_{(corrected)} = 0.812$, $r = 0.057$).

A similar overall pattern was observed for ITPC (Fig. 6). At the rate of pitch change, ITPC was marginally larger in the attend pitch

[1] One participant who failed to reach the performance threshold in the single dimension training but went on to reach the threshold in the dimension-selective attention training was included in this experiment.

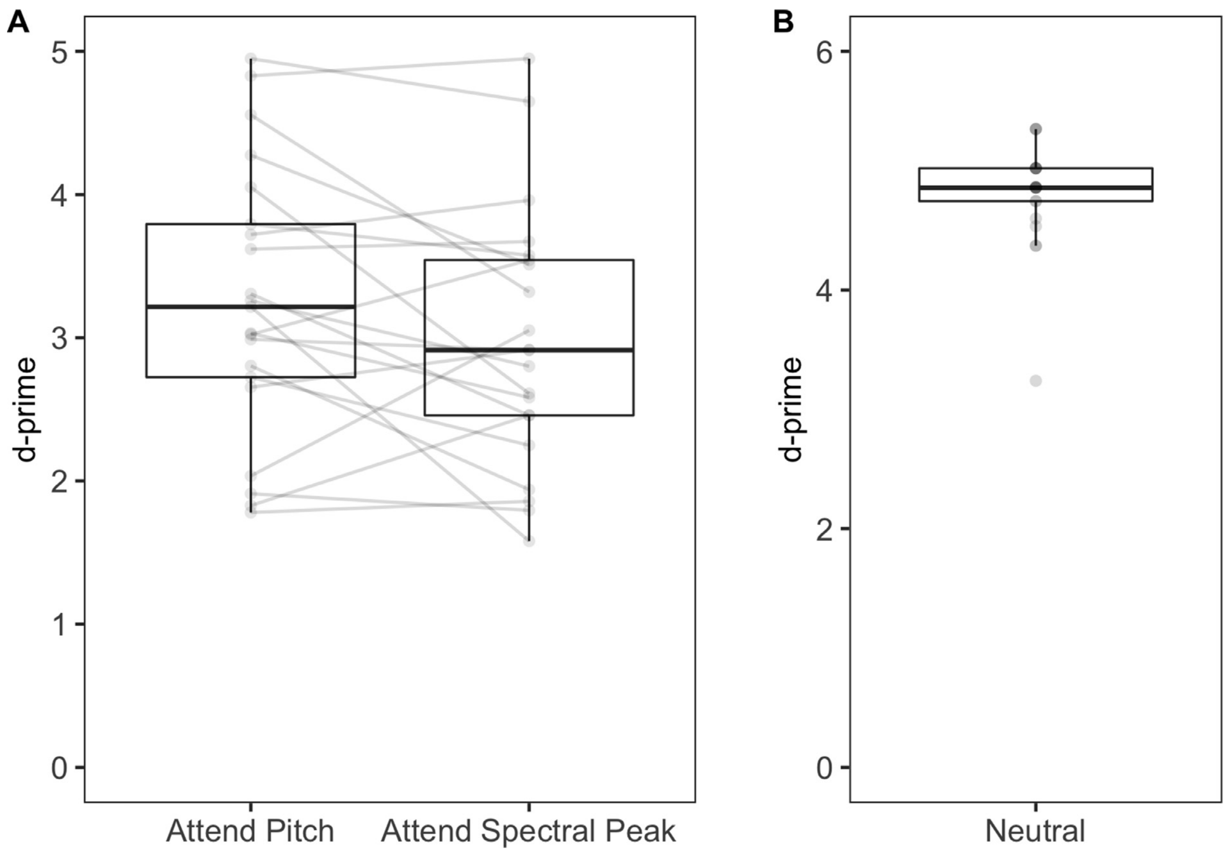

Fig. 4. Behavioral performance (d-prime) in the attend pitch and attend spectral peak conditions with lines connecting the data from individual subjects (A) and in the neutral condition (B). Note that the data from the attention tasks and the neutral tasks are plotted separately because the neutral task differed in the number of potential hits and false alarms. For each boxplot, the box represents the middle 50% of d-prime scores (Q1, median, Q3) and whiskers extend to the highest and lowest values within 1.5 times the inter-quartile range.

compared to the attend spectral peak condition (W = 176, z = 2.103, $p_{(corrected)}$ = 0.053, r = 0.459), as well as in the attend pitch compared to neutral condition (W = 202, z = 3.007, $p_{(corrected)}$ = 0.005, r = 0.656). The difference in ITPC between attend spectral peak and neutral conditions was not significant (W = 172, z = 1.954, $p_{(corrected)}$ = 0.06, r = 0.428). At the rate of spectral peak change, ITPC was larger in the attend spectral peak compared to the attend pitch condition (W = 185, z = 2.416, $p_{(corrected)}$ = 0.028, r = 0.523), and in the attend spectral peak compared to neutral condition (W = 212, z = 3.354, $p_{(corrected)}$ = 0.002, r = 0.732). The difference in ITPC at the rate of spectral peak change between attend pitch and neutral conditions was non-significant (W = 166, z = 1.755, $p_{(corrected)}$ = 0.082, r = 0.383).

In summary, for both pitch and spectral peak dimensions, we observed that attention enhanced cortical tracking of the attended dimension. However, there was no evidence that the unattended dimension was suppressed relative to the neutral attention condition.

3.2.3. Correlational analyses

Although this study was not explicitly designed to test for individual differences, we conducted exploratory correlational analyses to examine the relationship between behavioral performance, cortical tracking, and musical training to provide potential directions for future studies using this paradigm.

3.2.3.1. Relationship between behavioral and EEG data. To determine whether there was a relationship between behavioral performance and cortical tracking, we tested for a correlation between d-prime and the size of the attention effect in attend pitch and spectral peak conditions. Spearman's correlations were used with FDR correction for multiple correlations (Benjamini and Hochberg, 1995). In the attend pitch condition, the neural attention effect was measured by taking the difference in pitch tracking between conditions in which pitch was attended (attend pitch) and unattended (collapsed across attend spectral peak and neutral conditions). In the attend spectral peak condition, the neural attention effect was measured by taking the difference in spectral peak tracking between conditions in which spectral peak was attended (attend spectral peak) and unattended (collapsed across attend pitch and neutral conditions).

These analyses did not show any significant relationship between d and cortical tracking in either the attend pitch condition (ITPC: rho = 0.348, $p_{(corrected)}$ = 0.244; amplitude: rho = 0.392, $p_{(corrected)}$ = 0.244) or attend spectral peak condition (ITPC: rho = 0.166, $p_{(corrected)}$ = 0.629; amplitude: rho = 0.104, $p_{(corrected)}$ = 0.654).

3.2.3.2. Relationship with musical training. In this experiment, we predominantly recruited participants who received musical training. This provided the opportunity to conduct exploratory analyses to test whether there was a relationship between dimension-selective attention and years of musical training.

As shown in Fig. 7, there was a significant positive correlation between musical training and d-prime in the attend pitch condition (rho = 0.55, $p_{(corrected)}$ = 0.013) and attend spectral peak condition (rho = 0.53, $p_{(corrected)}$ = 0.013).

To explore whether this relationship with musical training was observed in the neural data, we correlated years of musical training with the attention effect in attend pitch and attend spectral peak conditions. Results showed a significant correlation between years of musical training and the ITPC attention effect for pitch (rho = 0.589,

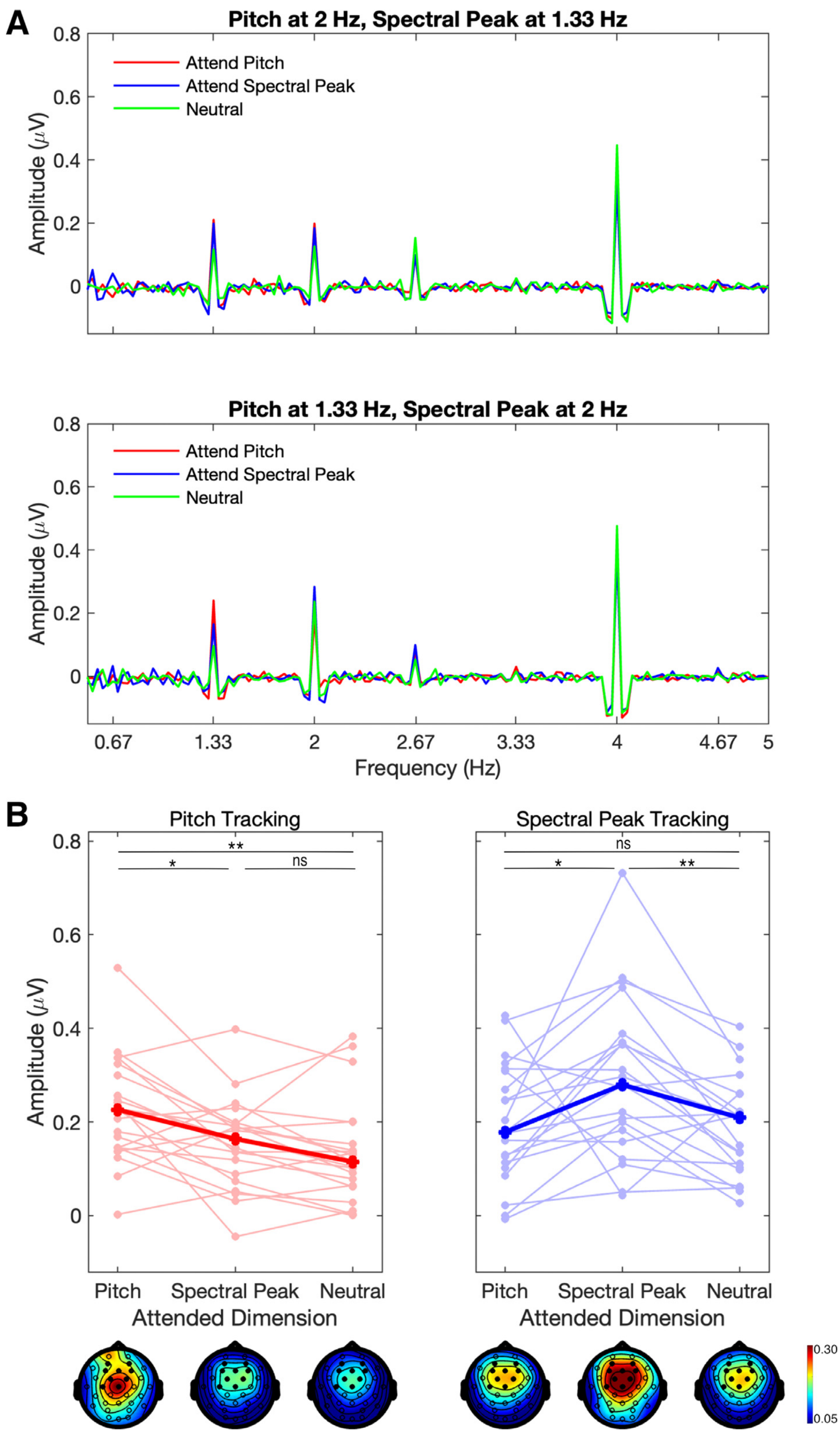

Fig. 5. (A) Median EEG signal amplitude across a range of frequencies (0.5 Hz – 5 Hz) in the attend pitch (red), attend spectral peak (blue) and neutral (green) conditions at each of the two dimension change rates. (B) Amplitude at the frequencies corresponding to variations in the pitch (left) and spectral peak (right) dimensions for each attention condition across the frontocentral channels selected for analysis. Individual participant data are represented by light red/blue lines and the median is represented by the dark red/blue lines. Topographical plots are displayed below the x axis, with the selected channels filled in black.

$p_{(corrected)}$ = 0.020). However, the correlation between years of musical training and the amplitude effect for pitch was not significant for amplitude (rho = 0.413, $p_{(corrected)}$ = 0.125). There was no significant correlation between years of musical training and the attention effect for spectral peak with either ITPC (rho = −0.044, $p_{(corrected)}$ = 0.848) or amplitude (rho = −0.099, $p_{(corrected)}$ = 0.848).

4. Discussion

4.1. Main findings

Here we demonstrate that neural tracking of acoustic dimensions can be modulated by both bottom-up salience and top-down atten-

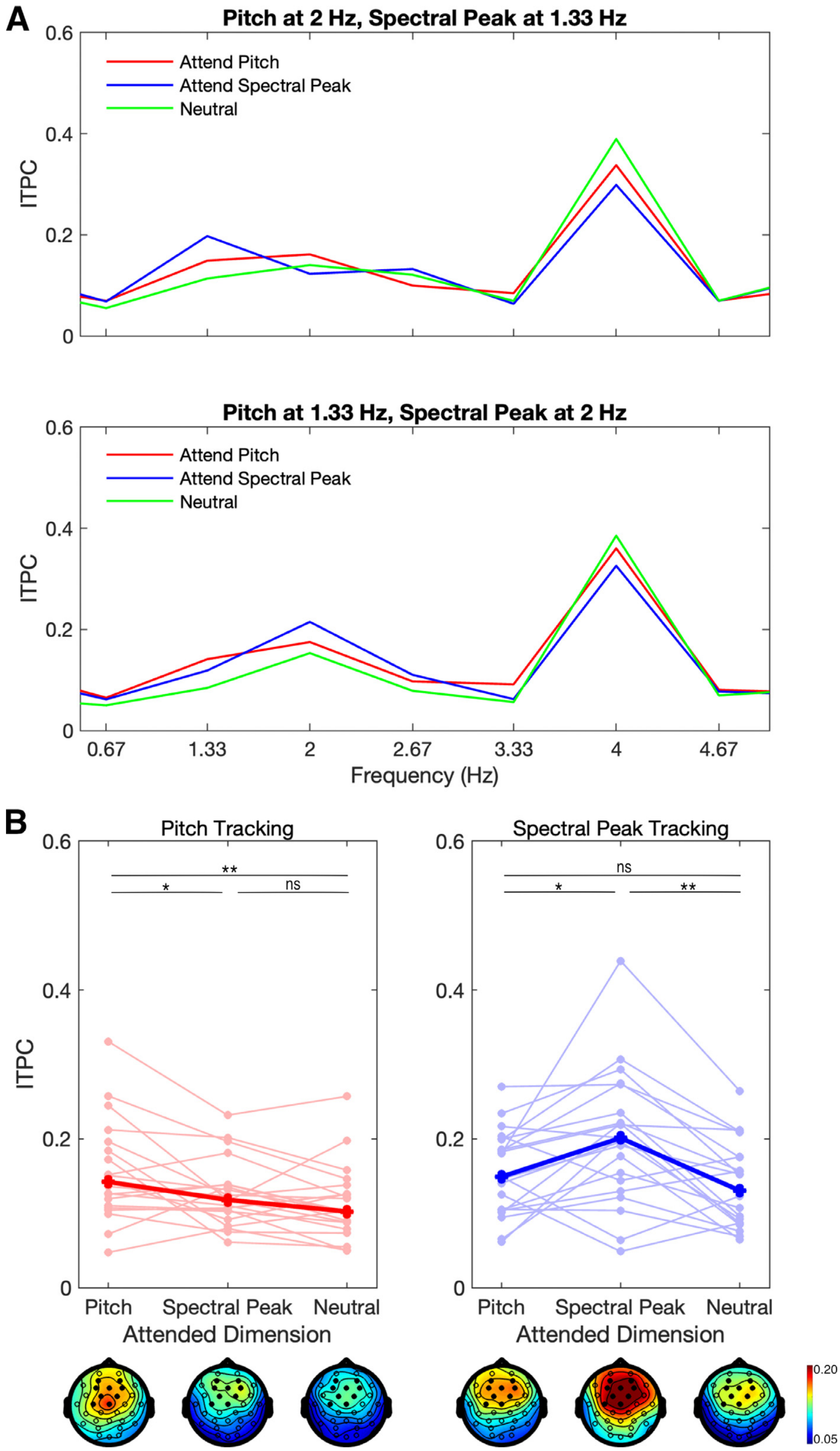

Fig. 6. (A) Median ITPC across a range of frequencies (0.5 Hz – 5 Hz) in the attend pitch (red), attend spectral peak (blue) and neutral (green) conditions at each of the two dimensions change rates. (B) ITPC at the frequencies corresponding to variations in the pitch (left) and spectral peak (right) dimensions for each attention condition across the frontocentral channels selected for analysis. Individual participant data are represented by light red/blue lines and the median is represented by the dark red/blue lines. Topographical plots are displayed below the x axis, with the selected channels filled in black.

tion. In two experiments, listeners heard single sound streams that varied in pitch and spectral peak frequency at fixed rates. Consistent with our hypotheses, both dimensional salience and dimension-selective attention modulated cortical tracking. Specifically, Experiment 1 showed stronger cortical tracking of pitch changes for tone sequences with high pitch salience (2-semitone pitch step sizes) compared to low pitch salience (1-semitone pitch step sizes). By contrast, relative pitch salience did not affect cortical tracking of simultaneously occurring changes in a different dimension, spectral peak frequency. In Experiment 2, we found stronger cortical tracking of attended compared to unattended dimensions, suggesting that selective attention can be directed to auditory dimensions. These findings show that attention can be directed to individual acoustic dimensions within a single auditory stream.

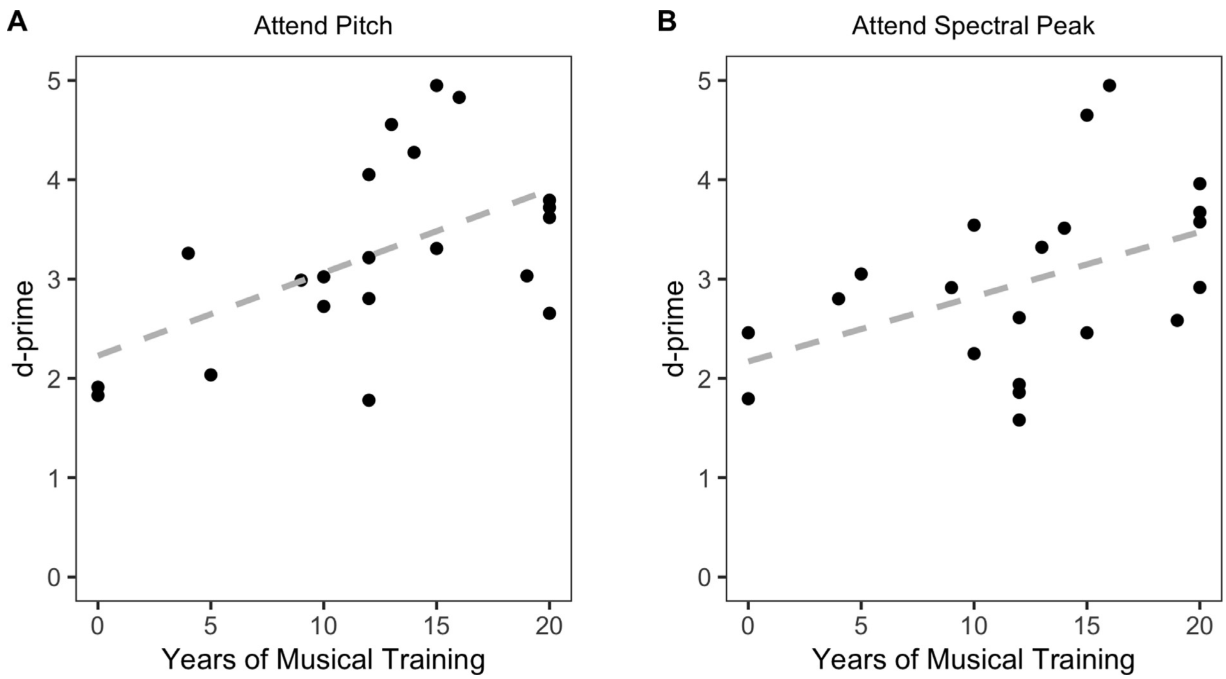

Fig. 7. Relationship between years of musical training and d-prime on the dimension-selective attention task in attend pitch (A) and attend spectral peak (B) conditions. The correlation between years of musical training and d-prime was significant in the attend pitch condition, suggesting that individuals with more musical training were better at selectively attending to both acoustic dimensions.

4.2. Dimensional salience

We find that dimensional salience can modulate neural tracking of dimensions: changes in pitch step size elicited stronger neural tracking of pitch variations, while neural tracking of spectral shape was unaffected. This finding builds on prior behavioral work showing that task-irrelevant variations in pitch can interfere with perception of timbre, and vice versa (Allen and Oxenham, 2014; Caruso and Balaban, 2014), suggesting that dimensions can compete with one another for attention, even when participants listen to a single auditory stream. Prior work has also shown that large changes in stimulus characteristics can capture attention to auditory streams (Berti et al., 2004; Escera et al., 1998; Marois et al., 2018; Siddle et al., 1984), with the size of pupil dilation responses (Marois et al., 2018; Wetzel et al., 2016) and ERP responses (Berti et al., 2004; Rinne et al., 2006; Schröger, 1996) varying in proportion to pitch step size. Similarly, in a recent frequency-tagging study, Kaya et al. (2020) observed an increase in power and cross-trial coherence in response to deviant tones with high (+6 semitones) compared to low (+2 semitone) pitch changes. Here we show that bottom-up stimulus features can guide attention even more precisely, directing listeners to focus on specific features within a single auditory stream. An advantage of our paradigm is that it does not necessarily require behavioral responses or deviant changes in an acoustic dimension to measure dimensional salience, but rather measures salience across entire recording blocks. Future work could make use of this measure to investigate factors driving dimensional salience within or between-subjects. For instance, this measure could be used to explore differences in dimensional salience that occur between speakers of different languages, such as pitch contour for speakers of tone languages versus non-tone languages.

4.3. Dimension-selective attention

In addition to being modulated by dimensional salience, cortical tracking was also enhanced by dimension-selective attention. Thus, top-down attention can modulate the cortical tracking of not just auditory objects (Elhilali et al., 2005; Huang and Elhilali, 2020; Kaya et al., 2020), but also acoustic dimensions within a single object. Although prior work has shown that attention can be directed to particular values along a dimension, such as high versus low frequencies (Dick et al., 2017) or particular points in time (Lange et al., 2003; Sanders and Astheimer, 2008), our findings show that attention can be directed to a dimension more globally, enhancing its neural representation across a broad range while leaving unaltered representations of other dimensions within the same object. Given that perceptual classification requires assessment of variability along multiple acoustic dimensions, including classification of speech (Toscano and McMurray, 2010), music (Prince, 2014), and environmental sounds (Lutfi and Liu, 2007), the ability to attend to a particular dimension may allow the listener to dynamically modulate the relative perceptual weights of the various features of an auditory object, prioritizing those features which are most informative for the task at hand.

In this respect, our results are broadly consistent with feature-based accounts of attention, which suggest that individuals can use features during the object-selection stage (Saenz et al., 2002); however, our results suggest that the role of auditory features is not just to guide attention to objects, but that features within an auditory stream can themselves be the *target* of auditory selective attention. Our study was not specifically designed to test object- versus feature-based accounts in the auditory domain, but future studies could adapt this paradigm to test specific predictions of object- and feature-based accounts. For instance, feature-based accounts predict that attention to a given dimension should enhance the processing of that dimension even for task-irrelevant objects, as has been observed in the visual domain (Adamian et al., 2020; Boehler et al., 2011; Chapman and Störmer, 2021; Saenz et al., 2002). Object-based accounts, on the other hand, predict that attention to a given dimension should enhance the processing of the unattended dimension within the same sound stream (Ernst et al., 2013; O'Craven et al., 1999). To test these predictions, future studies might use two multidimensional acoustic streams and instruct listeners to attend to a given dimension within one of the two streams. Enhanced cortical tracking of the attended dimension in the unattended stream would provide support for feature-based accounts while enhanced cortical tracking of both dimensions within a single stream would provide support for object-based accounts.

4.4. Applications to speech and music

The paradigms used in this study can be adapted to test specific predictions made by theoretical accounts of auditory categorization. For instance, attentional accounts of cue weighting suggest that listeners selectively attend to acoustic dimensions that are most diagnostic of category identity (Francis et al., 2000; Francis and Nusbaum, 2002; Holt et al., 2018). To test the role of dimension-selective attention in cue weighting, future studies could train participants to learn novel categories that differ along two orthogonal acoustic dimensions that vary in informativeness (e.g., Holt and Lotto, 2006). If dimensions that are diagnostic of category identity attract attention, we would expect increased cortical tracking of the acoustic dimensions following category learning, with greater tracking of more informative compared to less informative dimensions. Future research using this paradigm can also be used to explore whether attention is involved in the ability to adapt to short-term changes in the distribution of acoustic cues in speech. Prior work has shown that when the typical relationship between two orthogonal dimensions is reversed, listeners down-weight the secondary (less informative) dimension (e.g., Idemaru and Holt, 2011, 2014). If attention is involved, we would expect a decrease in the cortical tracking of the secondary dimension in response to short-term changes in cue distribution.

This paradigm could also be used to compare dimension-selective attention ability across different populations, such as individuals with and without musical training. In Experiment 2, we predominantly recruited participants who received musical training because of the difficulty of the task. This allowed us to test whether there was a relationship between years of musical training and dimension-selective attention. Musical training was associated with improved behavioral dimension-selective attention performance for both pitch and spectral peak dimensions. That this effect was not observed consistently in the neural data may have been due to lack of power, as this study was not explicitly designed to test for an effect of musicianship. Nonetheless, the behavioral effect suggests that musical training may be associated with robust dimension-selective attention. Evidence from previous research suggests that musical training is associated with less informational masking during tone detection (Oxenham et al., 2003), improved nonverbal auditory selective attention (Tierney et al., 2020), and better speech-in-speech perception in some studies (Clayton et al., 2016; Du and Zatorre, 2017; Parbery-Clark et al., 2009; Slater and Kraus, 2016; Zendel and Alain, 2012; but see Boebinger et al., 2015; Fuller et al., 2014; Ruggles et al., 2014). Our results provide an initial indication that this musician advantage might extend to dimension-selective attention. This paradigm could be used to further explore the relationship between musical training and dimension-selective attention. For example, future studies could compare trained musicians to non-musicians, or potentially compare cortical tracking of variations in pitch salience in musicians who specialize in melodic instruments versus percussionists.

4.5. Limitations and caveats

The inclusion of the neutral attention condition was aimed at testing the extent to which the effects of attention were driven by neural enhancement of the attended dimension versus suppression of the unattended dimension (Chait et al., 2010). Previous research of auditory object-based attention (Horton et al., 2013) has observed suppression effects of neural tracking of unattended stimuli. Suppression has also been suggested to play a role in attention to different acoustic dimensions (Costa-Faidella et al., 2017). However, we observed no evidence of suppression. It is possible that differences in task demands between the dimension-selective attention and neutral tasks may have influenced cortical tracking. The pattern of results observed in our study is unlikely to be driven solely by differences in task demands, as an effect of task difficulty might be expected to elicit an increase in cortical tracking at the 4-Hz tone presentation rate and result in a difference between the two conditions rather than the absence of a difference at the frequency tagged to the unattended dimension. Nonetheless, future work could remove any potential contribution of task demands by matching the difficulty of the neutral and dimension-selective attention tasks. Alternatively, the unattended dimension may not have represented a distraction in the current task, thus resulting in the absence of suppression. The current experimental design also does not allow us to rule out the possibility of automatic distractor suppression (Schneider et al., 2020). However, the results of the present study suggest that the effects of attention on cortical tracking may be driven by enhanced gain of the attended dimension rather than active suppression of the unattended dimension.

While our results show that attention can be directed to pitch and spectral peak, it remains to be investigated in future work whether these results generalize to other acoustic dimensions. Another potential limitation to the current study stems from the difficulty of the dimension-selective attention task. Behavioral results show that participants can selectively attend to pitch and spectral peak dimensions, with performance at above chance levels. However, dimension-selective attention appears to be a difficult task for many individuals. Of our initial sample of 27 participants, 6 failed to reach the required performance threshold on the training tasks. For this reason, we predominantly recruited individuals with musical training backgrounds. Therefore, future research will be needed to determine the extent to which these effects generalize to the general population. This could be achieved by a) increasing the amount of training provided and/or b) decreasing the difficulty of the task by slowing the rate of tone presentation and decreasing the number of levels of each dimension.

4.6. Individual variability

The variability in dimension-selective attention performance across individuals raises the question of whether and how dimension-selective attention can be trained. Previous research has shown that relatively short-term training can improve speech perception in noise (Whitton et al., 2014, 2017) and enhance electrophysiological indices of auditory spatial attention (Isbell et al., 2017; Stevens et al., 2008, 2013). Even two hours of training can boost behavioral and neural measures of auditory selective attention (Laffere et al., 2020). If training can improve object-based auditory attention, then it may be possible to train dimension-selective attention as well. Indirect support for this idea comes from studies showing that categorization training can alter the weight listeners place on different acoustic cues (Chandrasekaran et al., 2010; Francis et al., 2000, 2008). The paradigm used in the current study could provide a more direct measure of training-induced changes in dimension-selective attention. If training can boost dimension-selective attention, it could be used to help improve second-language learning by facilitating attention to acoustic dimensions relevant for categorization.

We also observed large individual differences in neural measures of dimensional salience and dimension-selective attention, consistent with previous studies showing large individual differences in object-based attention (Choi et al., 2014; Laffere et al., 2020; Ruggles and Shinn-Cunningham, 2011; Tierney et al., 2020). There are multiple potential factors that may contribute to these individual differences. On the one hand, individual differences in cortical tracking of a particular dimension may reflect variability in the precision with which that dimension is encoded by the auditory system. In this case, we would predict individuals who have more difficulty encoding a particular dimension (e.g., individuals with congenital amusia who have difficulty processing pitch) to show less passive cortical tracking of that dimension and potentially be less sensitive to variations in the salience of that dimension. On the other hand, variability in executive function or attentional control may contribute to differences in cortical tracking of attended dimensions. Since there is little research explicitly investigating dimension-selective attention, it is unclear whether and how this type of attention relates to other forms of auditory attention (e.g., object-based attention) as well as other executive functions. Understanding how these sensory and cognitive factors contribute to individual differences in dimensional salience and dimension-selective attention, and in turn, how these individual differences relate to auditory perception in natural listening environments remains a key topic for future research.

4.7. Conclusions

In conclusion, we provide evidence that dimensional salience and dimension-selective attention modulate cortical tracking, with greater tracking of salient and attended dimensions. This study offers a paradigm that can be used to explore the effects of dimensional salience and dimension-selective attention across multiple acoustic domains.

Declaration of Competing Interest

Declarations of interest: none.

CRediT authorship contribution statement

Ashley Symons: Software, Data curation, Formal analysis, Investigation; Methodology, Writing – original draft, Visualization. Fred Dick: Conceptualization, Methodology, Writing - review & editing, Visualization. Adam Tierney: Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing, Visualization, Supervision.

Acknowledgements

This work was supported by a Research Project Grant from the Leverhulme Trust [grant number RPG-2019–107] to ATT and a Reg and Molly Buck Award from SEMPRE to AES.

Data availability

The stimuli and processed data files are available at: osf.io/c5s6u/.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118544.

References

Adamian, N., Andersen, S.K., Hillyard, S.A., 2020. Parallel attentional facilitation of features and objects in early visual cortex. Psychophysiology 57 (3), e13498. doi:10.1111/psyp.13498.

Allen, E.J., Oxenham, A.J., 2014. Symmetric interactions and interference between pitch and timbre. J. Acoust. Soc. Am. 135 (3), 1371–1379. doi:10.1121/1.4863269.

Bala, A.D.S., Takahashi, T.T., 2000. Pupillary dilation response as an indicator of auditory discrimination in the barn owl. J. Compar. Physiol. A 186 (5), 425–434. doi:10.1007/s003590050442.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Berti, S., Roeber, U., Schröger, E., 2004. Bottom-up influences on working memory: behavioral and electrophysiological distraction varies with distractor strength. Exp Psychol 51 (4), 249–257. doi:10.1027/1618-3169.51.4.249.

Blair, M.R., Watson, M.R., Meier, K.M., 2009a. Errors, efficiency, and the interplay between attention and category learning. Cognition 112 (2), 330–336. doi:10.1016/j.cognition.2009.04.008.

Blair, M.R., Watson, M.R., Walshe, R.C., Maj, F., 2009b. Extremely selective attention: eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. J. Exper. Psychol. 35 (5), 1196–1206. doi:10.1037/a0016272.

Boebinger, D., Evans, S., Scott, S.K., Rosen, S., Lima, C.F., Manly, T., 2015. Musicians and non-musicians are equally adept at perceiving masked speech. J. Acoust. Soc. Am. 137 (1), 378–387. doi:10.1121/1.4904537.

Boehler, C.N., Schoenfeld, M.A., Heinze, H.-J., Hopf, J.-M., 2011. Object-based selection of irrelevant features is not confined to the attended object. J. Cogn. Neurosci. 23 (9), 2231–2239. doi:10.1162/jocn.2010.21558.

Caruso, V.C., Balaban, E., 2014. Pitch and timbre interfere when both are parametrically varied. PLoS One 9 (1), e87065. doi:10.1371/journal.pone.0087065.

Carvalho, P.F., Goldstone, R.L., 2017. The sequence of study changes what information is attended to, encoded, and remembered during category learning. J. Exp. Psychol. Learn Mem. Cogn. 43 (11), 1699–1719. doi:10.1037/xlm0000406.

Chait, M., de Cheveigné, A., Poeppel, D., Simon, J.Z., 2010. Neural dynamics of attending and ignoring in human auditory cortex. Neuropsychologia 48 (11), 3262–3271. doi:10.1016/j.neuropsychologia.2010.07.007.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., and Rosario, H.D. (2020). *pwr: Basic Functions for Power Analysis* (1.3-0) [Computer software]. https://CRAN.R-project.org/package=pwr

Chandrasekaran, B., Sampath, P.D., Wong, P.C.M., 2010. Individual variability in cue-weighting and lexical tone learning. J. Acoust. Soc. Am. 128 (1), 456–465. doi:10.1121/1.3445785.

Chapman, A.F., Störmer, V.S., 2021. Feature-based attention is not confined by object boundaries: spatially global enhancement of irrelevant features. Psychon. Bull. Rev. doi:10.3758/s13423-021-01897-x.

Chen, L., Meier, K.M., Blair, M.R., Watson, M.R., Wood, M.J., 2013. Temporal characteristics of overt attentional behavior during category learning. Attention, Percept. Psychophys. 75 (2), 244–256. doi:10.3758/s13414-012-0395-8.

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. doi:10.1016/j.heares.2014.04.008.

Clayton, K.K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A.D., Kidd, G., 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. PLoS One 11 (7), e0157638. doi:10.1371/journal.pone.0157638.

Costa-Faidella, J., Sussman, E.S., Escera, C., 2017. Selective entrainment of brain oscillations drives auditory perceptual organization. Neuroimage 159, 195–206. doi:10.1016/j.neuroimage.2017.07.056.

Dick, F.K., Lehet, M.I., Callaghan, M.F., Keller, T.A., Sereno, M.I., Holt, L.L., 2017. Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally directed attention and systematically related to cortical myeloarchitecture. J. Neurosci. 37 (50), 12187–12201. doi:10.1523/JNEUROSCI.1436-17.2017.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164. doi:10.1038/nn.4186.

Du, Y., Zatorre, R.J., 2017. Musical training sharpens and bonds ears and tongue to hear speech better. Proc. Natl. Acad. Sci. 114 (51), 13579–13584. doi:10.1073/pnas.1712223114.

Elhilali, M., Xiang, J., Shamma, S.A., Simon, J.Z., 2005. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol. 7 (6), e1000129. doi:10.1371/journal.pbio.1000129.

Ernst, Z.R., Boynton, G.M., Jazayeri, M., 2013. The spread of attention across features of a surface. J. Neurophysiol. 110 (10), 2426–2439. doi:10.1152/jn.00828.2012.

Escera, C., Alho, K., Winkler, I., Näätänen, R., 1998. Neural mechanisms of involuntary attention to acoustic novelty and change. J. Cogn. Neurosci. 10 (5), 590–604. doi:10.1162/089892998562997.

Francis, A.L., Baldwin, K., Nusbaum, H.C., 2000. Effects of training on attention to acoustic cues. Percept. Psychophys. 62 (8), 1668–1680. doi:10.3758/BF03212164.

Francis, A.L., Kaganovich, N., Driscoll-Huber, C., 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. J. Acoust. Soc. Am. 124 (2), 1234–1251. doi:10.1121/1.2945161.

Francis, A.L., Nusbaum, H.C., 2002. Selective attention and the acquisition of new phonetic categories. J. Exp. Psychol. Hum. Percept. Perform. 28 (2), 349–366. doi:10.1037//0096-1523.28.2.349.

Fritz, C.O., Morris, P.E., Richler, J.J., 2012. Effect size estimates: current use, calculations, and interpretation. J. Exper. Psychol. 141 (1), 2–18. doi:10.1037/a0024338.

Fuller, C.D., Galvin, J.J., Maat, B., Free, R.H., Başkent, D., 2014. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? Front. Neurosci. 8, 179. doi:10.3389/fnins.2014.00179.

Gordon, P.C., Eberhardt, J.L., Rueckl, J.G., 1993. Attentional modulation of the phonetic significance of acoustic cues. Cogn. Psychol. 25 (1), 1–42. doi:10.1006/cogp.1993.1001.

Hautus, M.J., 1995. Corrections for extreme proportions and their biasing effects on estimated values of d′. Behav. Res. Methods Instrum. Comput. 27 (1), 46–51. doi:10.3758/BF03203619.

Heald, S., Nusbaum, H.C., 2014. Speech perception as an active cognitive process. Front. Syst. Neurosci. 8. doi:10.3389/fnsys.2014.00035.

Holt, L.L., Lotto, A.J., 2006. Cue weighting in auditory categorization: implications for first and second language acquisition. J. Acoust. Soc. Am. 119 (5), 3059–3071. doi:10.1121/1.2188377.

Holt, L.L., Tierney, A.T., 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. doi:10.1016/j.heares.2018.06.014.

Horton, C., D'Zmura, M., Srinivasan, R, 2013. Suppression of competing speech through entrainment of cortical oscillations. J. Neurophysiol. 109 (12), 3082–3093. doi:10.1152/jn.01026.2012.

Huang, N., Elhilali, M., 2017. Auditory salience using natural soundscapes. J. Acoust. Soc. Am. 141 (3), 2163–2176. doi:10.1121/1.4979055.

Huang, N., Elhilali, M., 2020. Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. Elife 9, e52984. doi:10.7554/eLife.52984.

Idemaru, K., Holt, L.L., 2011. Word recognition reflects dimension-based statistical learning. J. Exp. Psychol. Hum. Percept. Perform. 37 (6), 1939–1956. doi:10.1037/a0025641.

Idemaru, K., Holt, L.L., 2014. Specificity of dimension-based statistical learning in word recognition. J. Exp. Psychol. Hum. Percept. Perform 40 (3), 1009–1021. doi:10.1037/a0035269.

Idemaru, K., Holt, L.L., Seltman, H., 2012. Individual differences in cue weights are stable across time: the case of Japanese stop lengths. J. Acoust. Soc. Am. 132 (6), 3950–3964. doi:10.1121/1.4765076.

Isbell, E., Stevens, C., Pakulak, E., Wray, A.H., Bell, T.A., Neville, H.J., 2017. Neuroplasticity of selective attention: research foundations and preliminary evidence for a gene by intervention interaction. Proc. Natl. Acad. Sci. 114 (35), 9247–9254. doi:10.1073/pnas.1707241114.

Jasmin, K., Dick, F., Holt, L.L., Tierney, A., 2019. Tailored perception: individuals' speech and music perception strategies fit their perceptual abilities. J. Exp. Psychol. Gen. 149 (5), 914–934. doi:10.1037/xge0000688.

Jasmin, K., Dick, F., Stewart, L., Tierney, A.T., 2020. Altered functional connectivity during speech perception in congenital amusia. Elife 9, e53539. doi:10.7554/eLife.53539.

Jasmin, K., Sun, H., Tierney, A.T., 2021. Effects of language experience on domain-general perceptual strategies. Cognition 206, 104481. doi:10.1016/j.cognition.2020.104481.

Kaya, E.M., Huang, N., Elhilali, M., 2020. Pitch, timbre and intensity interdependently modulate neural responses to salient sounds. Neuroscience 440, 1–14. doi:10.1016/j.neuroscience.2020.05.018.

Kerlin, J.R., Shahin, A.J., Miller, L.M., 2010. Attentional gain control of ongoing cortical speech representations in a "cocktail party". J. Neurosci. 30 (2), 620–628. doi:10.1523/JNEUROSCI.3631-09.2010.

Kim, D., Clayards, M., Goad, H., 2018. A longitudinal study of individual differences in the acquisition of new vowel contrasts. J. Phon. 67, 1–20. doi:10.1016/j.wocn.2017.11.003.

Kim, S., Rehder, B., 2011. How prior knowledge affects selective attention during category learning: an eyetracking study. Mem. Cognit. 39 (4), 649–665. doi:10.3758/s13421-010-0050-3.

Kong, E.J., Edwards, J., 2016. Individual differences in categorical perception of speech: cue weighting and executive function. J. Phon. 59, 40–57. doi:10.1016/j.wocn.2016.08.006.

Laffere, A., Dick, F., Holt, L.L., Tierney, A., 2021. Attentional modulation of neural entrainment to sound streams in children with and without ADHD. Neuroimage 224, 117396. doi:10.1016/j.neuroimage.2020.117396.

Laffere, A., Dick, F., Tierney, A., 2020. Effects of auditory selective attention on neural phase: individual differences and short-term training. Neuroimage 213, 116717. doi:10.1016/j.neuroimage.2020.116717.

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 (5), 806–817. doi:10.1111/1469-8986.00081.

Liao, H.-I., Yoneya, M., Kidani, S., Kashino, M., Furukawa, S., 2016. Human pupillary dilation response to deviant auditory stimuli: effects of stimulus properties and voluntary attention. Front. Neurosci. 10. doi:10.3389/fnins.2016.00043.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010. doi:10.1016/j.neuron.2007.06.004.

Lutfi, R.A., Liu, C.-J., 2007. Individual differences in source identification from synthesized impact sounds. J. Acoust. Soc. Am. 122 (2), 1017–1028. doi:10.1121/1.2751269.

Marois, A., Labonté, K., Parent, M., Vachon, F., 2018. Eyes have ears: indexing the orienting response to sound using pupillometry. Int. J. Psychophysiol. 123, 152–162. doi:10.1016/j.ijpsycho.2017.09.016.

Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. J. Neurosci. 31 (28), 10234–10240. doi:10.1523/JNEUROSCI.0411-11.2011.

Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. (Regul. Ed.) 23 (11), 913–926. doi:10.1016/j.tics.2019.08.004.

O'Craven, K.M., Downing, P.E., Kanwisher, N., 1999. FMRI evidence for objects as the units of attentional selection. Nature 401 (6753), 584–587. doi:10.1038/44134.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci 2011. doi:10.1155/2011/156869.

Oxenham, A.J., Fligor, B.J., Mason, C.R., Kidd, G., 2003. Informational masking and musical training. J. Acoust. Soc. Am. 114 (3), 1543–1549. doi:10.1121/1.1598197.

Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009. Musician enhancement for speech-in-noise. Ear Hear. 30 (6), 653–661. doi:10.1097/AUD.0b013e3181b412e9.

Prince, J.B., 2014. Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. J. Exp. Psychol. Hum. Percept. Perform. 40 (6), 2319–2337. doi:10.1037/a0038010.

Reetzke, R., Gnanateja, G.N., Chandrasekaran, B., 2021. Neural tracking of the speech envelope is differentially modulated by attention and language experience. Brain Lang. 213, 104891. doi:10.1016/j.bandl.2020.104891.

Rehder, B., Hoffman, A.B., 2005a. Eyetracking and selective attention in category learning. Cogn. Psychol. 51 (1), 1–41. doi:10.1016/j.cogpsych.2004.11.001.

Rehder, B., Hoffman, A.B., 2005b. Thirty-something categorization results explained: selective attention, eyetracking, and models of category learning. J. Exp. Psychol. Learn. Mem. Cogn. 31 (5), 811–829. doi:10.1037/0278-7393.31.5.811.

Rinne, T., Särkkä, A., Degerman, A., Schröger, E., Alho, K., 2006. Two separate mechanisms underlie auditory change detection and involuntary control of attention. Brain Res. 1077 (1), 135–143. doi:10.1016/j.brainres.2006.01.043.

Ruggles, D.R., Freyman, R.L., Oxenham, A.J., 2014. Influence of musical training on understanding voiced and whispered speech in noise. PLoS One 9 (1), e86980. doi:10.1371/journal.pone.0086980.

Ruggles, D., Shinn-Cunningham, B., 2011. Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. J. Assoc. Res. Otolaryngol. 12 (3), 395–405. doi:10.1007/s10162-010-0254-z.

Saenz, M., Buracas, G.T., Boynton, G.M., 2002. Global effects of feature-based attention in human visual cortex. Nat. Neurosci. 5 (7), 631–632. doi:10.1038/nn876.

Sanders, L.D., Astheimer, L.B., 2008. Temporally selective attention modulates early perceptual processing: event-related potential evidence. Percept. Psychophys. 70 (4), 732–742. doi:10.3758/PP.70.4.732.

Schneider, D., Herbst, S.K., Klatt, L.-I., Wöstmann, Malte, 2020. Target enhancement or distractor suppression? Functionally distinct alpha oscillations form the basis of attention. Eur. J. Neurosci. 00, 1–10. doi:10.1111/ejn.15309.

Schröger, E., 1996. A neural mechanism for involuntary attention shifts to changes in auditory stimulation. J. Cogn. Neurosci. 8 (6), 527–539. doi:10.1162/jocn.1996.8.6.527.

Siddle, D.A.T., Remington, B., Churchill, M., 1984. Effects of stimulus change on the electrodermal orienting response. Biol Psychol 18 (1), 33–39. doi:10.1016/0301-0511(84)90024-3.

Slater, J., Kraus, N., 2016. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. Cogn. Process. 17 (1), 79–87. doi:10.1007/s10339-015-0740-7.

Smith, J.O., 2007. Introduction to digital filters with audio applications, 2. Julius Smith https://ccrma.stanford.edu/~jos/filters/.

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. doi:10.1016/j.brainres.2007.10.108.

Stevens, C., Harn, B., Chard, D.J., 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 (1), 73–86. doi:10.1177/0022219411417877.

Teoh, E.S., Cappelloni, M.S., Lalor, E.C., 2019. Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. Eur. J. Neurosci. 50 (11), 3831–3842. doi:10.1111/ejn.14510.

Tierney, A., Rosen, S., Dick, F., 2020. Speech-in-speech perception, nonverbal selective attention, and musical training. J. Exp. Psychol. Learn Mem. Cogn. 46 (5), 968–979. doi:10.1037/xlm0000767.

Toscano, J.C., McMurray, B., 2010. Cue integration with categories: weighting acoustic cues in speech using unsupervised learning and distributional statistics. Cogn. Sci. 34 (3), 434–464. doi:10.1111/j.1551-6709.2009.01077.x.

Wetzel, N., Buttelmann, D., Schieler, A., Widmann, A., 2016. Infant and adult pupil dilation in response to unexpected sounds. Dev. Psychobiol. 58 (3), 382–392. doi:10.1002/dev.21377.

Whitton, J.P., Hancock, K.E., Polley, D.B., 2014. Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. Proc. Natl. Acad. Sci. 111 (25), E2606–E2615. doi:10.1073/pnas.1322184111.

Whitton, J.P., Hancock, K.E., Shannon, J.M., Polley, D.B., 2017. Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise. Curr. Biol. 27 (21), 3237–3247. doi:10.1016/j.cub.2017.09.014, .e6.

Winn, M.B., Chatterjee, M., Idsardi, W.J., 2013. Roles of voice onset time and F0 in stop consonant voicing perception: effects of masking noise and low-pass filtering. J. Speech Lang. Hear. Res. 56 (4), 1097–1107. doi:10.1044/1092-4388(2012/12-0086).

Zaki, S.R., Salmi, I.L., 2019. Sequence as context in category learning: an eyetracking study. J. Exp. Psychol. Learn. Mem. Cogn. 45 (11), 1942–1954. doi:10.1037/xlm0000693.

Zendel, B.R., Alain, C., 2012. Musicians experience less age-related decline in central auditory processing. Psychol. Aging 27 (2), 410–417. doi:10.1037/a0024816.

Zhao, S., Yum, N.W., Benjamin, L., Benhamou, E., Yoneya, M., Furukawa, S., Dick, F., Slaney, M., Chait, M., 2019. Rapid ocular responses are modulated by bottom-up-driven auditory salience. J. Neurosci. 39 (39), 7703–7714. doi:10.1523/JNEUROSCI.0776-19.2019.

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

Zoefel, B., Archer-Boyd, A., Davis, M.H., 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Curr. Biol. 28 (3), 401–408. doi:10.1016/j.cub.2017.11.071, .e5.