

Dimension-Selective Attention and Dimensional Salience Modulate Cortical Tracking of  
Acoustic Dimensions

Ashley E. Symons<sup>1</sup>, Fred Dick<sup>1,2</sup>, Adam T. Tierney<sup>1</sup>

<sup>1</sup>Department of Psychological Sciences, Birkbeck College, University of London

<sup>2</sup>Division of Psychology & Language Sciences, University College London

Correspondence to:

Ashley Symons

Department of Psychological Sciences

Birkbeck College, University of London

Malet Street

London WC1E 7HX

a.symons@bbk.ac.uk

## **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 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 systematically different rates. Inter-trial phase coherence (ITPC) and EEG signal amplitude at the rates of pitch and spectral change allowed us to measure 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. Neural entrainment to pitch changes was greater for sequences with larger compared to smaller pitch intervals, with no difference in entrainment to the spectral dimension. In Experiment 2, participants selectively attended to either the pitch or spectral dimension. Neural entrainment was stronger in response to the attended compared to unattended dimension for both pitch and spectral dimensions. These findings demonstrate that bottom-up and top-down attentional mechanisms enhance the cortical tracking of different acoustic dimensions within a single sound stream.

Key words: auditory, attention, salience, EEG, oscillations

## Introduction

Speech perception requires the mapping of continuous acoustic dimensions onto discrete linguistic categories. A central issue in speech perception is how listeners dynamically integrate and weight information from different acoustic dimensions during categorization. Prior work has shown that listeners weight acoustic dimensions according to the reliability with which each dimension distinguishes between categories (Holt et al., 2018; Toscano & 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 & 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 & 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, 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, listeners dynamically allocate attentional resources towards dimensions that are informative, and away from those that are less informative (Francis et al., 2000; Francis & Nusbaum, 2002; Gordon et al., 1993; Heald & 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). Although 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, Watson, & Meier, 2009; Blair, Watson, Walshe, et al., 2009; Carvalho & Goldstone, 2017; Chen et al., 2013; Kim & Rehder, 2011; Rehder & Hoffman, 2005a, 2005b; Zaki & 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 neural 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 & Poeppel, 2007; Obleser & 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 neural entrainment to target relative to distractor tone sequences (Elhilali et al., 2005; Laffere et al., 2020, 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). Neural entrainment 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 & 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 & 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 (Escera et al., 1998; Berti et al., 2004; Rinne et al., 2006; Schröger, 1996). Salient acoustic changes can also influence the degree of neural entrainment to acoustic streams, with decreased entrainment to attended streams following salient background sounds (Huang & Elhilali, 2020) and increased entrainment to 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 neural entrainment to increased salience along a dimension.

To investigate whether cortical tracking of acoustic dimensions is modulated by dimensional salience (Experiment 1) and dimension-selective attention (Experiment 2), we used a frequency tagging approach (e.g., Ding et al., 2016; Nozaradan et al., 2011), where changes in different dimensions are tagged to fixed rates. 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. 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 neural entrainment would be modulated by increased salience along the pitch dimension. Thus, we predicted that stronger inter-trial phase-locking 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 allowed us to test whether the effects of attention were due to attentional enhancement of the attended dimension, or suppression of the unattended dimension. We hypothesized that dimension-selective attention would result in enhanced neural entrainment to a dimension when it was attended, and potentially suppressed neural entrainment to a dimension when it was unattended.

## **Experiment**

### **Methods**

#### **Participants**

Twenty-nine participants (14 female, 15 male) between the ages of 19-59 with no known hearing impairments and no diagnosis of a language or learning disorder took part in this experiment. This sample size was powered to detect a medium effect ( $d = 0.5$ ) between 2-semitone versus 1-semitone pitch step size conditions ( $\alpha = 0.05$ ,  $\beta = 0.8$ ).

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 3-20 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.

## **Design**

In this experiment, participants listened to isochronous sequences of complex tones that varied in pitch and spectral peak frequency at different rates (see Figure 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 neural entrainment to 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).

## **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).

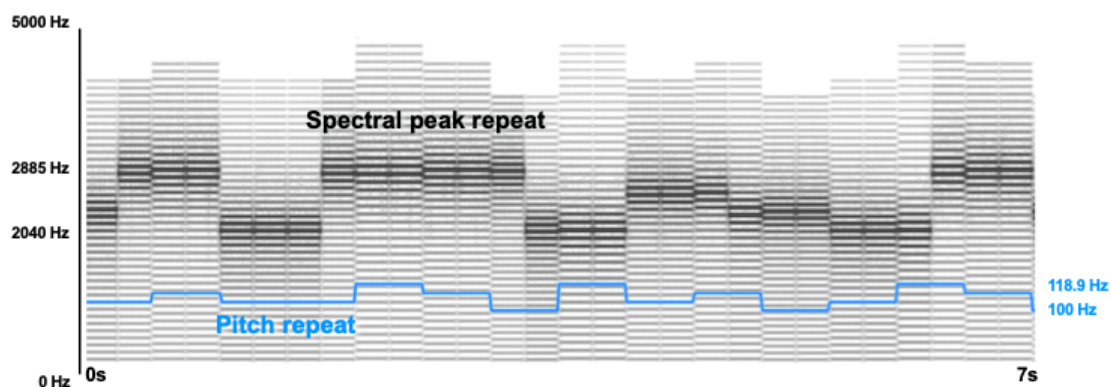
These two sets of tones were concatenated without pause to form 120-second sequences (480 tones). Crucially, pitch and spectral peak changed at different rates (1.33 Hz and 2 Hz). When pitch varied at 2 Hz, spectral peak varied at 1.33 Hz. When pitch varied at 1.33 Hz, spectral peak varied at 2 Hz (Figure 1). 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, with the exception of 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 (2Hz), 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 seconds 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.





*Figure 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).*

## **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 seconds), 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 in order 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.

## **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 k $\Omega$ . Triggers marking the beginning of each trial (every 6 tones, or 1.5 seconds) 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 neural entrainment were computed: inter-trial phase coherence (ITPC) and signal amplitude. Prior to calculation of ITPC, any remaining artifacts exceeding  $\pm 100 \mu\text{V}$  were rejected. 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. 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-second 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) in order 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.

## **Data Analysis**

### **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 seconds 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 were used to test for an effect of pitch step size on performance.

## **EEG Data**

For both phase and amplitude measures, 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 neural 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 (AF3, AF4, F3, F4, Fz, FC1, FC2, Cz, C3) for both phase and amplitude measures. 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 neural entrainment to pitch and spectral peak dimensions. Wilcoxon signed-rank tests were used to compare the effect of pitch step size on neural entrainment to pitch and spectral peak. False Discovery Rate was used to correct for multiple comparisons using the Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995). Processed data are available at: [osf.io/c5s6u/](https://osf.io/c5s6u/).

## **Results**

### **Behavioral**

As expected, performance was near ceiling in both the small pitch step size (median  $d'$  = 4.863) and large pitch step size (median  $d'$  = 4.599) conditions. The difference between the two conditions was statistically significant ( $V = 43$ ,  $z = -2.52$ ,  $p = 0.012$ ). The difference in behavioral performance was unexpected but could suggest that larger pitch step sizes are more distracting, resulting in worse performance. However, upon closer investigation, this difference may also be due to the different proximity of the quiet oddball tones in the different conditions. Specifically, we noticed that there were 3 instances where oddball tones occurred within the same 1.25 second window as another oddball tone. Two of these occurred in the 2-semitone pitch step size condition and one occurred in the 1-semitone pitch step size condition. Removing responses to both oddball tones within these time windows along with two extra randomly-selected responses to oddball tones in the 1-semitone condition (to ensure that the total proportion of hits and false alarms was balanced

across the two conditions) from all participants resulted in no significant difference in behavioral performance between small (median  $d$ -prime = 4.89) and large (median  $d$ -prime = 4.80) pitch step size conditions ( $V = 44$ ,  $z = -1.54$ ,  $p = 0.129$ ).

## EEG

Figure 2 displays neural entrainment to pitch and spectral dimensions at 1-semitone and 2-semitone pitch step sizes for ITPC (A) and signal amplitude (B). For both measures, neural entrainment to pitch changes was stronger for the 2-semitone compared to 1-semitone condition (ITPC:  $V = 284$ ,  $z = 2.760$ ,  $p_{(\text{corrected})} = 0.009$ , amplitude:  $V = 300$ ,  $z = 3.160$ ,  $p_{(\text{corrected})} = 0.004$ ). As predicted, neural entrainment to spectral peak changes presented at the same rates did not differ as a function of pitch condition (ITPC:  $V = 121$ ,  $z = -1.38$ ,  $p_{(\text{corrected})} = 0.231$ , amplitude:  $V = 148$ ,  $z = -0.698$ ,  $p_{(\text{corrected})} = 0.499$ ). In fact, there was no significant difference in neural entrainment to pitch versus spectral dimension in the 1-semitone condition (ITPC:  $V = 186$ ,  $z = 0.267$ ,  $p_{(\text{corrected})} = 0.803$ ; amplitude:  $V = 199$ ,  $z = 0.597$ ,  $p_{(\text{corrected})} = 0.754$ ), while in the 2-semitone condition neural entrainment was stronger to the pitch compared to spectral dimension (ITPC:  $V = 319$ ,  $z = 3.64$ ,  $p_{(\text{corrected})} < 0.001$ ; amplitude:  $V = 326$ ,  $z = 3.820$ ,  $p_{(\text{corrected})} < 0.001$ ).

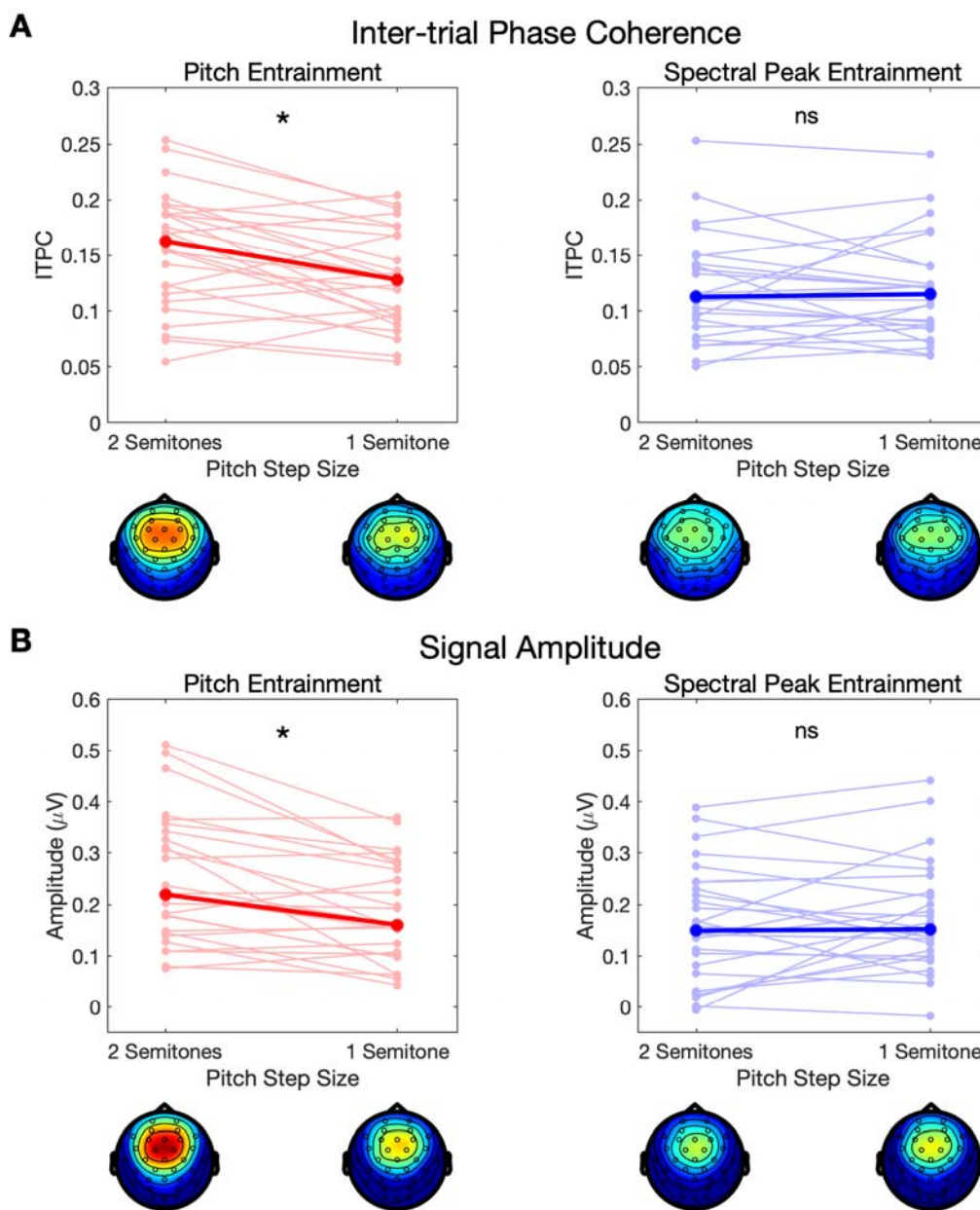


Figure 2. (A) Inter-trial phase coherence in response to variations in the pitch (red) and spectral peak (blue) dimensions for each pitch step size (x-axis). The values displayed represent the average across the frontocentral channels selected for analysis. Individual participants are represented in light red/blue lines and the median value represented by the dark red/blue lines, with topographical plots (scale: 0.05 – 0.2) for each condition displayed below the x-axis. (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, with topographical plots (scale = 0.05 – 0.3) below the x axis. For both measures, neural entrainment to the pitch dimension was stronger for the 2 compared to 1 semitone pitch step size condition. No effect of pitch step size was observed on neural entrainment to the spectral peak dimension.

## **Experiment 2**

### **Methods**

#### **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 4-20 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.

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.

#### **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 rate of pitch and spectral peak change was counterbalanced within participants to ensure that any effect of attention was not due to differences in neural entrainment to 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).

## **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).

## **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.

### **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 seconds 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), and 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.



## Procedure

### 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 seconds) 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 seconds) 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 part-way 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 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.

## **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 dimension change rates x 3 attended dimensions). Stimulus presentation was blocked by attended dimension in order 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 8-10 minutes. 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 hour.

## **Data Acquisition and Analysis**

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

### **2.1.5.1 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 seconds 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-second 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 & Hochberg, 1995).

## **EEG**

ITPC and signal amplitude were 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 neural entrainment to pitch and spectral peak dimensions in each attention condition. Wilcoxon signed-rank tests were used to compare the different attention conditions (attend pitch, attend spectral peak, neutral) on pitch and spectral peak entrainment. False Discovery Rate was used to correct for multiple comparisons using the Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995). Processed data are available at: [osf.io/c5s6u/](https://osf.io/c5s6u/).

## **Results**

### **Behavioral**

#### **Training**

In the single dimension training, participants who passed the training required between 1-5 blocks of training to reach the performance threshold. Hit rate ranged from 0.63-1.0 (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 0-5 (median = 0) in the attend spectral

peak condition and 0-2 (median = 0) in the attend pitch condition.<sup>1</sup> In the neutral condition, hit rate was at ceiling (1.0) for all participants with between 0-1 (median = 0) false alarms. In the dimension-selective attention training, hit rate ranged from 0.75-1.0 (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 0-3 (median = 1). In the neutral condition, hit rate was at ceiling (1.0) for all participants with between 0-1 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.

### **EEG Task**

Figure 3 shows task performance (d-prime) in attend pitch and attend spectral peak conditions. There was no significant difference in performance between the two conditions ( $V = 162$ ,  $z = 1.08$ ,  $p = 0.111$ ), suggesting that, on balance, task difficulty was matched across the two dimensions. In the neutral condition, performance was high (median d-prime = 4.857), comparable to the behavioral performance observed in Experiment 1.

---

<sup>1</sup> One participant who failed to reach the performance threshold in the single dimension training but went onto reach the threshold in the dimension-selective attention training was included in this experiment.

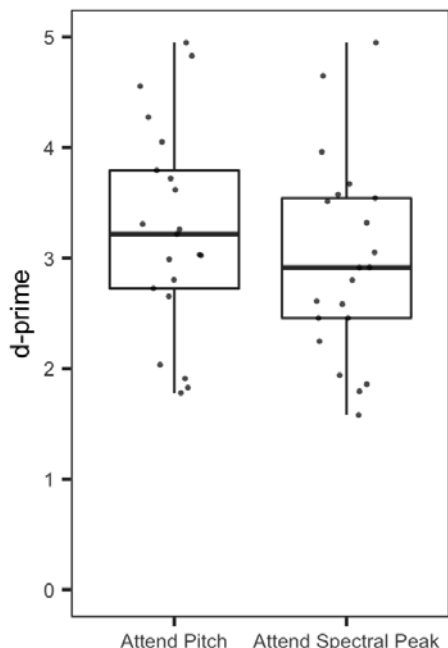


Figure 3. Behavioral performance ( $d$ -prime) in the attend pitch and attend spectral peak conditions. No significant differences in performance were observed between the two conditions.

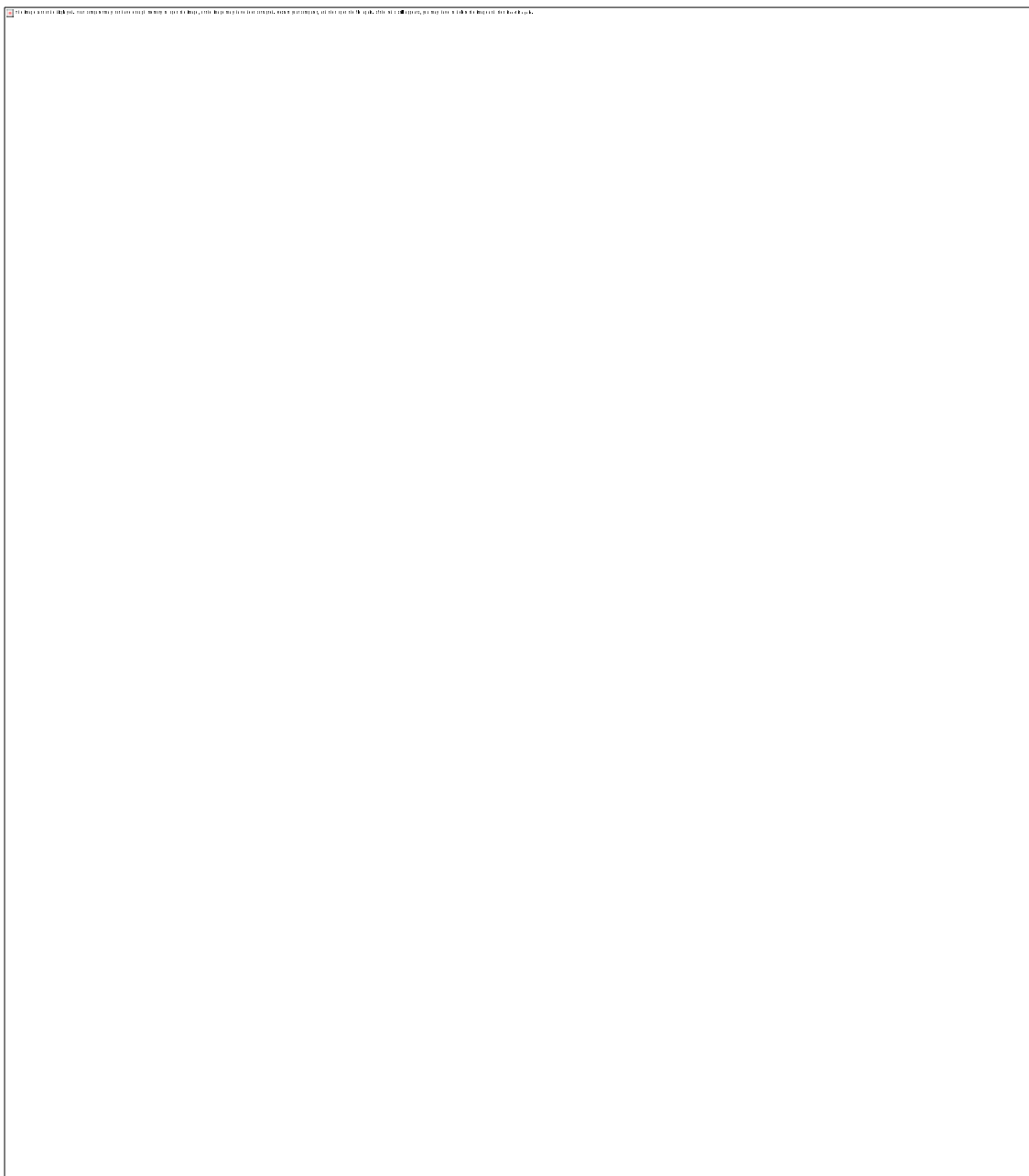
## EEG

As shown in Figure 4, neural entrainment to pitch was stronger in the attend pitch compared to the attend spectral peak condition (ITPC:  $V = 187$ ,  $z = 2.490$ ,  $p_{(\text{corrected})} = 0.017$ , amplitude:  $V = 188$ ,  $z = 2.520$ ,  $p_{(\text{corrected})} = 0.017$ ), as well as in the attend pitch compared to neutral condition (ITPC:  $V = 209$ ,  $z = 3.250$ ,  $p_{(\text{corrected})} = 0.002$ , amplitude:  $V = 208$ ,  $z = 3.220$ ,  $p_{(\text{corrected})} = 0.002$ ). There was also significantly greater neural entrainment to pitch in the attend spectral peak versus the neutral conditions as assessed via ITPC ( $V = 179$ ,  $z = 2.210$ ,  $p_{(\text{corrected})} = 0.035$ ). However, this effect was not observed for signal amplitude ( $V = 162$ ,  $z = 1.620$ ,  $p_{(\text{corrected})} = 0.133$ ).

Neural entrainment to the spectral peak dimension was stronger in the attend spectral peak compared to the attend pitch condition (ITPC:  $V = 193$ ,  $z = 2.690$ ,  $p_{(\text{corrected})} = 0.013$ , amplitude:  $V = 187$ ,  $z = 2.490$ ,  $p_{(\text{corrected})} = 0.017$ ), and in the attend spectral peak compared to neutral condition (ITPC:  $V = 209$ ,  $z = 3.250$ ,  $p_{(\text{corrected})} = 0.002$ , amplitude:  $V = 199$ ,  $z = 2.900$ ,  $p_{(\text{corrected})} = 0.007$ ). There was no significant difference in neural entrainment to the

spectral dimension between attend pitch and neutral conditions (ITPC:  $V = 141$ ,  $z = 0.886$ ,  $p_{(\text{corrected})} = 0.428$ , amplitude:  $V = 123$ ,  $z = 0.261$ ,  $p_{(\text{corrected})} = 0.812$ ).

Thus, for both pitch and spectral peak dimensions, attention enhanced neural entrainment to the attended dimension. However, there was no evidence that the unattended dimension was suppressed relative to the neutral attention condition.



*Figure 4. (A) Inter-trial phase coherence in response to variations in the pitch (left) and spectral peak (right) dimensions for each attention condition (x-axis). The values displayed represent the average across the frontocentral channels identified in Experiment 1. Individual participants are represented in light red/blue lines and the median value represented by the dark red/blue lines, with topographical plots (scale: 0.05 – 0.2) for each condition displayed below the x-axis. (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, with topographical plots (scale = 0.05 – 0.3) below the x axis.*

## **Correlational analyses**

### **Relationship between Behavioral and EEG Data**

To determine whether there was a relationship between behavioral performance and neural entrainment, 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 & Hochberg, 1995). In the attend pitch condition, the neural attention effect was measured by taking the difference in pitch entrainment 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 entrainment between conditions in which spectral peak was attended (attend spectral peak) and unattended (collapsed across attend pitch and neutral conditions). Correlations were conducted for both ITPC and amplitude data.

These analyses did not show any significant relationship between d-prime and neural entrainment in either the attend pitch condition (ITPC:  $\rho = 0.484$ ,  $p_{(\text{corrected})} = 0.104$ ; amplitude:  $\rho = 0.392$ ,  $p_{(\text{corrected})} = 0.157$ ) or attend spectral peak condition (ITPC:  $\rho = -0.136$ ,  $p_{(\text{corrected})} = 0.654$ ; amplitude:  $\rho = 0.104$ ,  $p_{(\text{corrected})} = 0.654$ ).

### **Relationship with musical training**

In this experiment, we predominantly recruited participants who received musical training. This provided the opportunity to examine whether there was a relationship between dimension-selective attention and years of musical training.

As shown in Figure 5, there was a significant positive correlation between musical training and d-prime in the attend pitch condition ( $\rho = 0.55$ ,  $p_{(\text{corrected})} = 0.010$ ) and attend spectral peak condition ( $\rho = 0.53$ ,  $p_{(\text{corrected})} = 0.013$ ).

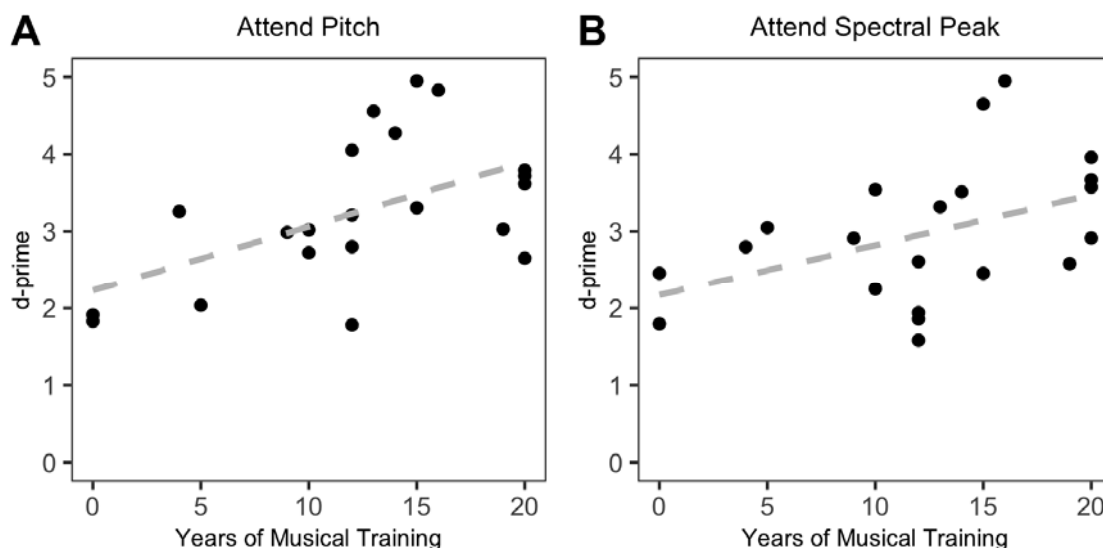


Figure 5. Relationship between years of musical training and d-prime on the dimension-selective attention task in attend pitch (A) and attend spectral peak (B) conditions. The correlation between years of musical training and d-prime was significant in the attend pitch condition, suggesting that individuals with more musical training were better at selectively attending to both acoustic dimensions.

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. However, there was no significant correlation between years of musical training and the attention effect for pitch (ITPC:  $\rho = 0.432$ ,  $p_{(\text{corrected})} = 0.125$ ; amplitude:  $\rho = 0.413$ ,  $p_{(\text{corrected})} = 0.125$ ) or spectral peak (ITPC:  $\rho = -0.044$ ,  $p_{(\text{corrected})} = 0.848$ ; amplitude:  $\rho = -0.099$ ,  $p_{(\text{corrected})} = 0.848$ ) after correcting for multiple correlations.

## Discussion

In two studies, we investigated the neural correlates of dimensional salience and selective attention within a single sound stream that varied in pitch and spectral peak frequency. Consistent with our hypotheses, both dimensional salience and dimension-selective attention modulated neural entrainment. Specifically, Experiment 1 showed stronger neural



entrainment to changes along the pitch dimension 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 entrainment to simultaneously occurring changes in a different dimension, spectral peak frequency.

In Experiment 2, we found stronger entrainment for attended compared to unattended dimensions. These findings are consistent with previous studies showing that both bottom-up and top-down attentional mechanisms modulate neural entrainment to distinct auditory objects (Elhilali et al., 2005; Huang & Elhilali, 2020; Kaya et al., 2020). Here we extend these findings to show that dimensional salience and dimension-selective attention can enhance neural entrainment to different acoustic dimensions within a single sound stream.

There are many potential factors that may drive dimensional salience. Here we focus on pitch, which is one among the most important acoustic dimensions for human language and music (Patel, 2010). Previous research has shown that deviations in pitch 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. Moreover, a recent EEG study has 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 (Kaya et al., 2020). Our results are consistent with these findings, showing that changes in the pitch step size elicit stronger neural entrainment specifically to pitch variations. An advantage of our paradigm is that it does not require deviant changes in an acoustic dimension to measure dimensional salience, but rather measures salience across entire recording blocks. To further establish this increase in neural entrainment as a marker of dimensional salience, future studies could present acoustically identical stimuli while altering the perceived salience of different acoustic dimensions within or between-subjects. For instance, this paradigm could provide a measure of dimensional salience before and after learning novel non-speech auditory categories that vary in the reliability of different acoustic dimensions (Holt & Lotto, 2006), or at different stages of second language acquisition. Alternatively, 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.

In addition to being modulated by dimensional salience, neural entrainment was also enhanced by dimension-selective attention. This effect of attention was demonstrated for two distinct dimensions – pitch and spectral peak. Previous research has shown that pitch and spectral dimensions can interact, with random variations in one dimension impeding discrimination of the other (Allen & Oxenham, 2014). However, our results provide one of the first demonstrations that attention can be directed to different acoustic dimensions. Whereas previous studies of auditory selective attention have established neural correlates of auditory stream segregation (e.g., Ding et al., 2016; Elhilali et al., 2005; Kerlin et al., 2010; Laffere et al., 2020, 2021; Zion Golumbic et al., 2013), here we show reliable neural correlates of attention to acoustic dimensions within a single sound stream.

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 entrainment to 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. This suggests that, at least in the context of the current study, the neural entrainment effects were 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 seen whether these results will generalize to other acoustic dimensions. Future work can examine the extent to which the effects of dimensional salience and dimension-selective attention 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. To explore this effect in a wider sample of participants, future studies could increase the amount of training provided and/or decrease the difficulty of the task by slowing the rate of tone presentation and decreasing the number of levels of each dimension.

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, 2008). 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 & 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.

The study of dimensional salience and dimension-selective attention has implications for speech perception and language acquisition. 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 & Nusbaum, 2002; Holt et al., 2018). However, research investigating the link between dimension-selective attention and cue weighting in speech perception is sparse, in part because of a lack of established measures of dimension-selective attention. The paradigms used in our studies could be coupled with studies of cue weighting in speech perception to establish an association between dimensional salience, selective attention and cue weighting. For example, a study could train participants to learn novel categories that differ along two (or more) dimensions and examine whether cue weights correlate with phase-locking to each dimension. Alternatively, research using speech stimuli could examine whether short-term changes in cue distribution result in related changes in neural entrainment.

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 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 & Zatorre, 2017; Parbery-Clark et al., 2009; Slater & Kraus, 2016; Zendel & Alain, 2012; but see Boebinger et al., 2015; Fuller et al., 2014; Madsen et al., 2017; 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 entrainment to variations in pitch salience in musicians who specialize in melodic instruments versus percussionists.

In conclusion, we provide evidence that dimensional salience and dimension-selective attention modulate neural entrainment, with stronger entrainment to 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.

### **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.

### **Declaration of Interest**

Declarations of interest: none

## References

- Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*(3), 1371–1379. <https://doi.org/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. *Journal of Comparative Physiology A*, *186*(5), 425–434. <https://doi.org/10.1007/s003590050442>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. <https://doi.org/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. *Experimental Psychology*, *51*(4), 249–257. <https://doi.org/10.1027/1618-3169.51.4.249>
- Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*(2), 330–336. <https://doi.org/10.1016/j.cognition.2009.04.008>
- Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196–1206. <https://doi.org/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. *The Journal of the Acoustical Society of America*, *137*(1), 378–387. <https://doi.org/10.1121/1.4904537>

- Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(11), 1699–1719. <https://doi.org/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. <https://doi.org/10.1016/j.neuropsychologia.2010.07.007>
- Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465. <https://doi.org/10.1121/1.3445785>
- 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, Perception, & Psychophysics*, *75*(2), 244–256. <https://doi.org/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. *Hearing Research*, *314*, 10–19. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1016/j.neuroimage.2017.07.056>

- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. <https://doi.org/10.1038/nn.4186>
- Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences*, *114*(51), 13579–13584. <https://doi.org/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 Biology*, *7*(6), e1000129. <https://doi.org/10.1371/journal.pbio.1000129>
- Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. <https://doi.org/10.1162/089892998562997>
- Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680. <https://doi.org/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. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251. <https://doi.org/10.1121/1.2945161>
- Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. <https://doi.org/10.1037/0096-1523.28.2.349>



- 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? *Frontiers in Neuroscience*, 8, 179. <https://doi.org/10.3389/fnins.2014.00179>
- Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25(1), 1–42. <https://doi.org/10.1006/cogp.1993.1001>
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of  $d'$ . *Behavior Research Methods, Instruments, & Computers*, 27(1), 46–51. <https://doi.org/10.3758/BF03203619>
- Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8. <https://doi.org/10.3389/fnsys.2014.00035>
- Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119(5), 3059–3071. <https://doi.org/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. *Hearing Research*, 366, 50–64. <https://doi.org/10.1016/j.heares.2018.06.014>
- Horton, C., D’Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, 109(12), 3082–3093. <https://doi.org/10.1152/jn.01026.2012>
- Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, 141(3), 2163–2176. <https://doi.org/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.  
<https://doi.org/10.7554/eLife.52984>
- Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1939–1956. <https://doi.org/10.1037/a0025641>
- Idemaru, K., & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(3), 1009–1021. <https://doi.org/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. *The Journal of the Acoustical Society of America*, *132*(6), 3950–3964. <https://doi.org/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. *Proceedings of the National Academy of Sciences*, *114*(35), 9247–9254. <https://doi.org/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. *Journal of Experimental Psychology. General*, *149*(5), 914–934.  
<https://doi.org/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.  
<https://doi.org/10.7554/eLife.53539>

- Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*, 104481.  
<https://doi.org/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.  
<https://doi.org/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”. *The Journal of Neuroscience*, *30*(2), 620–628. <https://doi.org/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. *Journal of Phonetics*, *67*, 1–20.  
<https://doi.org/10.1016/j.wocn.2017.11.003>
- Kim, S., & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & Cognition*, *39*(4), 649–665.  
<https://doi.org/10.3758/s13421-010-0050-3>
- Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57.  
<https://doi.org/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. <https://doi.org/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.  
<https://doi.org/10.1016/j.neuroimage.2020.116717>

- 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. *Frontiers in Neuroscience, 10*.  
<https://doi.org/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.  
<https://doi.org/10.1016/j.neuron.2007.06.004>
- Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports, 7*(1), 12624. <https://doi.org/10.1038/s41598-017-12937-9>
- Marois, A., Labonté, K., Parent, M., & Vachon, F. (2018). Eyes have ears: Indexing the orienting response to sound using pupillometry. *International Journal of Psychophysiology, 123*, 152–162. <https://doi.org/10.1016/j.ijpsycho.2017.09.016>
- Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience, 31*(28), 10234–10240.  
<https://doi.org/10.1523/JNEUROSCI.0411-11.2011>
- Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences, 23*(11), 913–926.  
<https://doi.org/10.1016/j.tics.2019.08.004>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*.  
<https://doi.org/10.1155/2011/156869>

- Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, *114*(3), 1543–1549. <https://doi.org/10.1121/1.1598197>
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661. <https://doi.org/10.1097/AUD.0b013e3181b412e9>
- Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press Press.
- Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain and Language*, *213*, 104891. <https://doi.org/10.1016/j.bandl.2020.104891>
- Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. <https://doi.org/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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 811–829. <https://doi.org/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 Research*, *1077*(1), 135–143. <https://doi.org/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. <https://doi.org/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.

*Journal of the Association for Research in Otolaryngology*, 12(3), 395–405.

<https://doi.org/10.1007/s10162-010-0254-z>

Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience*, 8(6), 527–539.

<https://doi.org/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. *Biological Psychology*, 18(1), 33–39.

[https://doi.org/10.1016/0301-0511\(84\)90024-3](https://doi.org/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. *Cognitive Processing*,

17(1), 79–87. <https://doi.org/10.1007/s10339-015-0740-7>

Smith, J. O. (2007). *Introduction to Digital Filters: With Audio Applications*. Julius Smith.

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 Research*, 1205, 55–69. <https://doi.org/10.1016/j.brainres.2007.10.108>

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. *European Journal of Neuroscience*, 50(11), 3831–

3842. <https://doi.org/10.1111/ejn.14510>

Tierney, A., Rosen, S., & Dick, F. (2020). Speech-in-speech perception, nonverbal selective

attention, and musical training. *Journal of Experimental Psychology: Learning,*

*Memory, and Cognition*, 46(5), 968–979. <https://doi.org/10.1037/xlm0000767>

- Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34(3), 434–464. <https://doi.org/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. *Developmental Psychobiology*, 58(3), 382–392. <https://doi.org/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. *Proceedings of the National Academy of Sciences*, 111(25), E2606–E2615. <https://doi.org/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. *Current Biology*, 27(21), 3237–3247.e6. <https://doi.org/10.1016/j.cub.2017.09.014>
- 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. *Journal of Speech, Language, and Hearing Research*, 56(4), 1097–1107. [https://doi.org/10.1044/1092-4388\(2012/12-0086\)](https://doi.org/10.1044/1092-4388(2012/12-0086))
- Zaki, S. R., & Salmi, I. L. (2019). Sequence as context in category learning: An eyetracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(11), 1942–1954. <https://doi.org/10.1037/xlm0000693>
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, 27(2), 410–417. <https://doi.org/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. *Journal of Neuroscience*, 39(39), 7703–7714.

<https://doi.org/10.1523/JNEUROSCI.0776-19.2019>

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

<https://doi.org/10.1016/j.neuron.2012.12.037>

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, 28(3),

401-408.e5. <https://doi.org/10.1016/j.cub.2017.11.071>