Manuscript for

Musicality enhances selective attention: spatiotemporal insights from precise separation of simultaneous neural responses

1. ABSTRACT

We precisely resolved simultaneous neural responses to their driving stimuli and classified them to different selective attention conditions towards specific melodies within mixtures. Magnetoencephalographic findings reveal the influence of participants' musicality and task performance on the recruitment of cortical regions in bottom-up and top-down attentional processes. Across parietal regions, musicality and performance correlated positively with top-down attention at the left hemisphere, but negatively with bottom-up attention at the right. In the right prefrontal cortex, individuals who maintained selective attention over the tone duration performed better and were more musical. These results suggest that musical training boosts performance by enhancing top-down attention, reducing bottom-up distractions, and sustaining selective attention overtime, with these effects mediated by the frontoparietal cortex. This work advances our understanding of neural selective attention mechanisms in complex soundscapes and underscores the potential of musical training for cognitive enhancement. Our breakthrough approach effectively captures cognitive effects with frequency-tagging, creating new possibilities for scientific applications.

2. INTRODUCTION

Analyzing neurophysiological signals generated by multiple concurrent stimuli is challenging due to the difficulty in resolving brain signals according to their respective stimuli. In auditory neuroscience, an effective method to achieve this is crucial as simultaneous sounds are commonly encountered in naturalistic environments, such as a cocktail party scenario or an orchestra setting.

Frequency-tagging has been suggested as a viable solution to this problem by encoding each neural response with a precise identification tag which allows the neural activity to be separated and identified from a mixture[1,2]. However, this approach is notoriously tedious and difficult to implement, often producing auditory steady-state responses (ASSRs) with insufficient signal-to-noise ratios (SNR) needed to capture subtle cognitive effects, especially when multiple sounds are simultaneously present[3–5]. Firstly, auditory frequency-tagging is extremely intolerant towards small inconsistencies in timings (e.g. onset trigger time). To circumvent this limitation, we operated a specialized auditory setup that can achieve near-zero lags in sound delivery, and recorded brain activity at millisecond resolution using magnetoencephalography (MEG). Moreover, while the SNR can be improved with time-locked signal averaging, even minute inter-subject spatial variations can dilute and obscure potential effects with univariate analysis. To address this, we developed a multivariate approach using ASSR-specialized support vector machines that are capable of capturing cognitive and behavioral effects with high precision and sensitivity.

In this paper, we established the viability of this method with clear evidence from two experiments showing that our algorithms were able to (i) resolve neural signals according to their driving stimuli with absolute precision, (ii) localize and extract the effects of both top-down and bottom-up driven selective attention over key cortical areas, as well as (iii) achieve this on a continuous scale that correlated with individual musicality as well as performance scores on selective melody listening tasks.

Overall, the results demonstrate that the pattern of brain activity and direction of correlations differ significantly between top-down and bottom-up attention, especially at the frontoparietal cortices. The findings also indicate that musical training facilitates the maintenance and improvement of selective auditory attention over time through frontal mechanisms, providing us an insight as to how learning and expertise optimize brain function to accomplish cognitive goals.

3. RESULTS

General Task description

Participants listened to two simultaneous melodies of different pitch for a duration of 10 - 15 s per block. The low-pitched melody was frequency-tagged at 39 Hz and the high-pitched melody was frequency-tagged at 43 Hz to elicit corresponding neural responses at 39 and 43 Hz (Fig. 1).

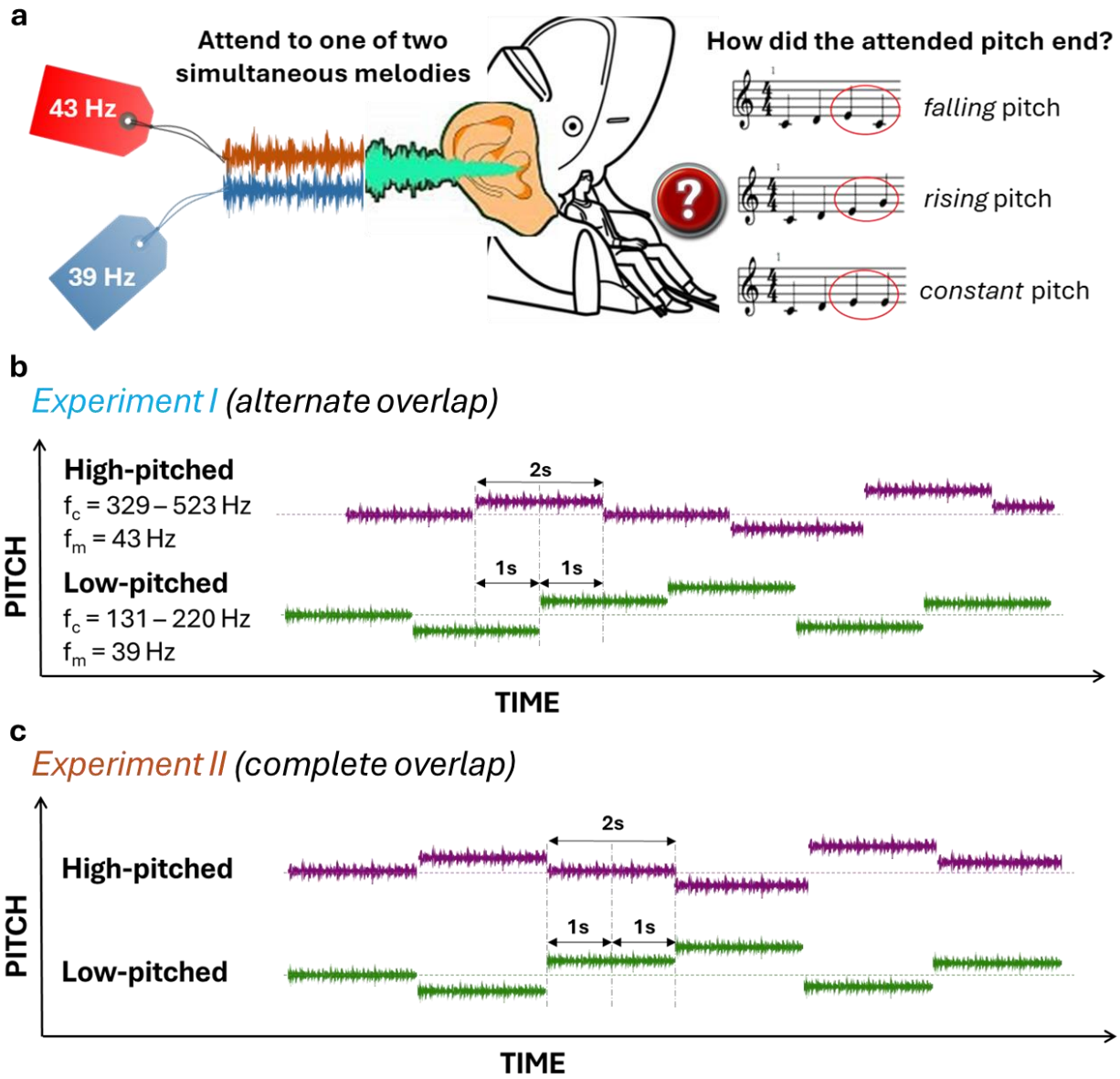

Fig. 1 | Illustrative description of experimental tasks.
a, Melody attention task. Participants listened to two simultaneous melodies that differed in pitch and were instructed to selectively attend to one. The low-pitched and high-pitched melodies were frequency-tagged at 39 and 43 Hz respectively. When the melody stopped, participants were required to report the last direction of pitch change for the attended melody, which was either falling, rising or constant. Identical sounds were presented to both ears, ensuring that the melodies could only be distinguished by pitch or timing. **b**, Experiment I: Alternately overlapping melodies. Tone onsets alternate between the low-pitched and high-pitched melodies. Each new tone induces bottom-up attention towards it, allowing the dissociation of top-down and bottom-up attention effects towards each melody. **c**, *Experiment II: Completely overlapping melodies*. Melody tones overlap completely, engaging bottom-up attention simultaneously towards both melodies. All other parameters including modulation frequencies (f_m) and carrier frequencies (f_c) were identical to experiment I.

For both experiments (experiment I and II), participants' top-down selective attention was manipulated by instructing them to focus exclusively on either the low-pitched or high-pitched melody. In experiment I, the tone onsets alternate between the low-pitched and high-pitched melodies, drawing an effect of bottom-up attention towards each new tone. Thus, this experimental design allowed us to dissociate top-down and bottom-up attention effects towards each melody. In experiment II, the tone onsets for both melodies occurred concurrently, thereby engaging bottom-up attention simultaneously towards both melodies.

For both experiments, participants' task performance (measured by fraction of correct responses) was above chance level, verifying that selective attention was manipulated accordingly (N_I=28, m_I = 0.70, std_I=0.25; N_{II}=20, m_{II} = 0.77, std_{II}=0.21; two-tailed t-tests, $p < 0.001$;). Furthermore, similar to previous experiments[6–8] (see Extended Data Fig. 1 for description of previous experiments), task performance correlated positively with individual musicality in the current two experiments (Pearson correlation, $p < 0.001$). Musicality was measured by an adapted version of the Goldsmiths musical sophistication index[9,10]. Notably, the strength of correlation (Pearson r) between musicality and performance increased with task complexity (Fig. 2). The least complex task (experiment 0, r = 0.64) involved selectively attending to melodies that were completely separated in time and pitch. The moderately complex task (experiment I, r = 0.71) required attending to melodies that were partly separated in time and pitch. The most complex task (experiment II, r = 0.79) involved selectively attending to melodies that completely overlapped in time and could only be separated by pitch.

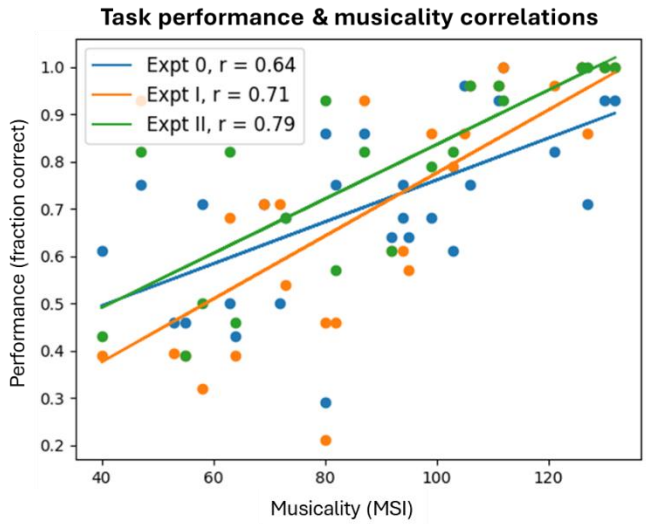

Fig. 2 | Correlations between task performance and musicality.
Task performance correlated positively (Pearson correlation, $p < 0.001$) with musicality across participants in present (Expt I and II) and past experiments (Expt 0). Musicality was measured by the Goldsmiths musical sophistication index[9,10] (MSI). Notably, the strength of correlation (Pearson r) between musicality and performance increased with task complexity. The least complex task (Expt 0, r = 0.64, N = 28) involved selectively attending to melodies that were completely separated in time and pitch. The moderately complex task (Expt I, r = 0.71, N = 28) required attending to melodies that were partly separated in time and pitch. The most complex task (Expt II, r = 0.79, N = 20) involved selectively attending to melodies that completely overlapped in time and could only be separated by pitch.

Selective attention to either melody was classified precisely at stimulus frequency tags

We trained a classifier on the frequency-tagged neural activity to discriminate between conditions in which attention was directed towards either the low-pitched or high-pitched melody. The area-under-curve (AUC) was computed to measure the classifier's ability to discriminate between these attention conditions, which was expected to increase with stronger engagement of selective attention.

For experiment I, peak discrimination of the classifier was achieved specifically at the set stimulus frequency tags of 39 and 43 Hz, but not at other frequencies. This was observed for both top-down (AUC = 0.62 (39 Hz), 0.60 (43 Hz); Fig. 3a – Left) and bottom-up attention (AUC = 0.54 (39 Hz), 0.63 (43 Hz); Fig. 3a – Right). Similarly for experiment II, peak discrimination of the classifier was observed at 39 and 43 Hz (AUC = 0.62 (39 & 43 Hz), obtained from the higher classification AUC

between the early half and late half of the tone at each frequency (Fig. 3b). The individual tone halves however, did not exhibit any discrimination peak at the expected frequencies (Extended Data Fig. 2). These results confirmed that we can reliably extract stimulus-specific neural responses at the predefined frequency-tags (i.e. 39 & 43 Hz), and with sufficient sensitivity to capture the effects of selective attention.

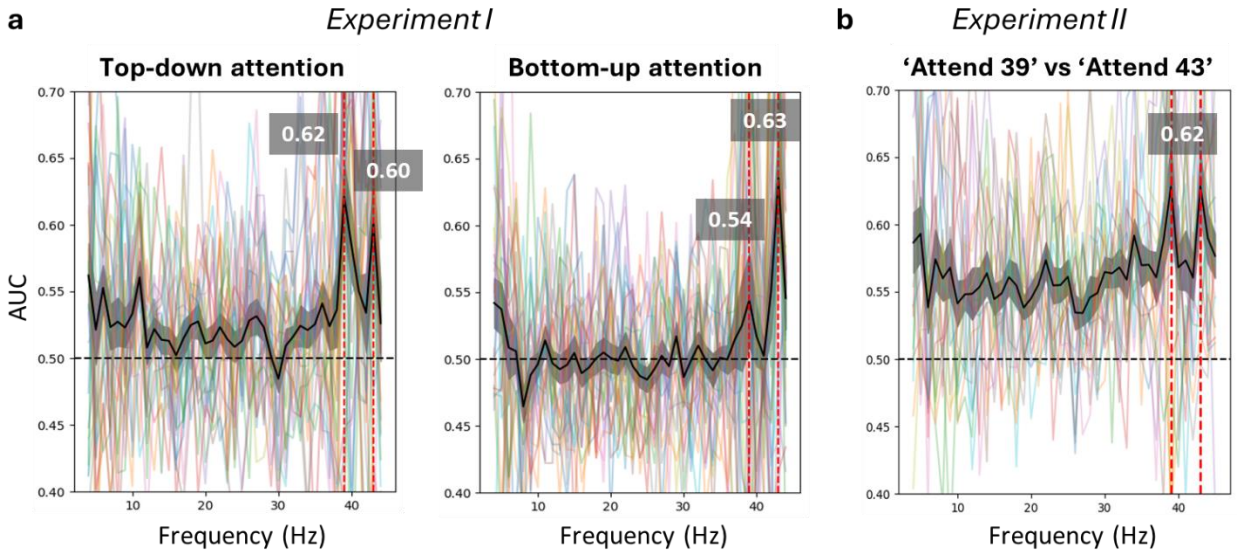

Fig. 3 | Classification of selective attention at frequency-tags
We trained a classifier on frequency-tagged neural activities to discriminate between conditions where attention was directed towards either the low-pitched or high-pitched melody. Classifier performance was evaluated using the area-under-curve (AUC), which was expected to increase with higher selective attention. Peak AUC values were specifically observed at the stimulus frequency tags of 39 and 43 Hz, but not at other frequencies. For each plot, the mean across participants is shown in black, with red vertical lines marking the AUC peaks at 39 and 43 Hz, displayed in gray boxes above. **a**, Experiment I (N = 25). Selective attention peaks were observed at 39 and 43 Hz for both top-down (left) and bottom-up attention (right). **b**, Experiment II (N = 20). Similar selective attention peaks were observed at 39 and 43 Hz. These results confirm that we have successfully extracted neural responses at the predefined frequency-tags and captured their attentional modulation. Shading indicates standard error of the mean (s.e.m.).

Temporal regions operate on bottom-up attention while frontal regions operate on top-down attention

We proceeded to examine how different brain regions are recruited in selective attentional processes, focusing on cortical areas that have been observed to contain ASSR sources in our previous studies[6–8]. Six regions of interest, namely the orbital gyrus (OrG), superior temporal gyrus (STG), and inferior parietal lobe (IPL) at each hemisphere, were identified and demarcated with a predefined atlas (Brainnetome Atlas[11]).

Among these six regions, bottom-up attention was discriminated above the chance level of 0.5 only at the right STG, while top-down attention was discriminated above chance at all regions ($p < 0.01$ for all significant cases, $n = 10000$, one-tailed permutation test, false discovery rate (FDR)-corrected for 12 tests) except the right IPL (Fig. 4). At the left and right OrG, the classifier could discriminate top-down attention significantly better than bottom-up attention ($p_{left} < 0.01$, $p_{right} < 0.001$, $n = 10000$, two-tailed permutation test, FDR-corrected for 6 tests). These results are in agreement with current literature[12–15] championing that bottom-up attention is triggered by lower-level automatic sensory mechanisms situated predominantly in the primary cortices such as the STG, while top-down attention recruits higher-level executive mechanisms located in the prefrontal cortex (Cohen 2014; Foster et al. 1994; S. Huang et al. 2012; Plakke and Romanski 2014).

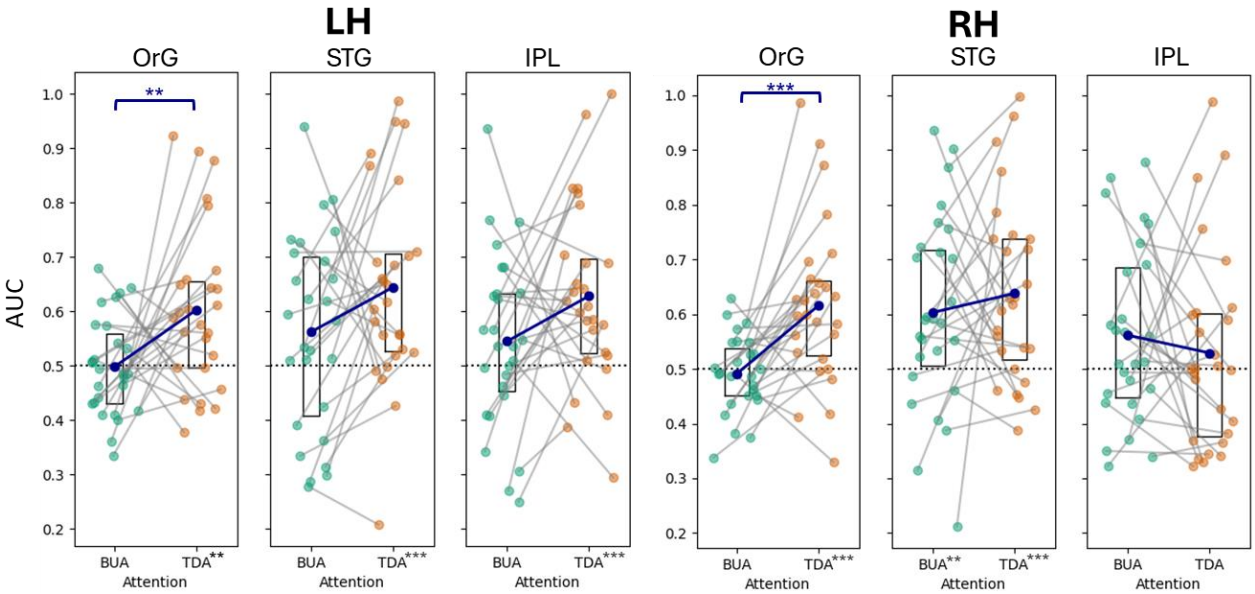

Fig. 4 | Top-down and bottom-up selective attention across cortical regions
Across participants (individual markers; N = 27), bottom-up attention was significantly discriminated above chance (dotted line) only at the right superior temporal gyrus (STG), while top-down attention was discriminated above chance at all regions except the right inferior parietal lobe (IPL). Significance levels above chance are indicated by black asterisks beside the corresponding attention condition on the x-axis. The y-axis denotes the AUC and is identical for all subplots. Box plots represent the 25th – 75th percentiles of the data, with the center dot indicating the mean. At the left and right orbital gyri (OrG), top-down attention was discriminated significantly better than bottom-up attention (blue brackets and asterisks). These results support the notion that bottom-up attention is triggered by lower-level automatic sensory mechanisms situated predominantly in the primary cortices such as the STG, while top-down attention recruits higher-level executive mechanisms located in the prefrontal cortex. All p-values are computed with permutation tests and false discovery rate (FDR)-corrected. ***p<0.001, **p<0.01, *p<0.05.

Top-down attention correlated positively with musicality and performance, while bottom-up attention correlated negatively

To examine the relationship between attention, musicality, and task performance, we correlated the AUC with participants' musicality scores as well as task performance. For musicality, a positive correlation with top-down attention was observed already at sensor level in experiment I (Pearson correlation, $r = 0.42$, $p < 0.05$; Extended Data Fig. 3a).

We proceeded to investigated which ASSR sources underlie this correlation, with our primary analysis centered on sources in the IPL, a region that has shown correlations with musicality in our previous study[6]. Indeed, the left IPL exhibited a significant positive correlation between top-down attention and task performance (Pearson correlation, $r = 0.49$, $p < 0.01$; Fig. 5a - Left). Further analysis of the remaining regions revealed that this correlation effect was also driven by the left OrG (Pearson correlation, $r = 0.40$, $p < 0.05$) and right STG (Pearson correlation, $r = 0.39$, $p < 0.05$), albeit to a smaller extent (Extended Data Fig. 3c).

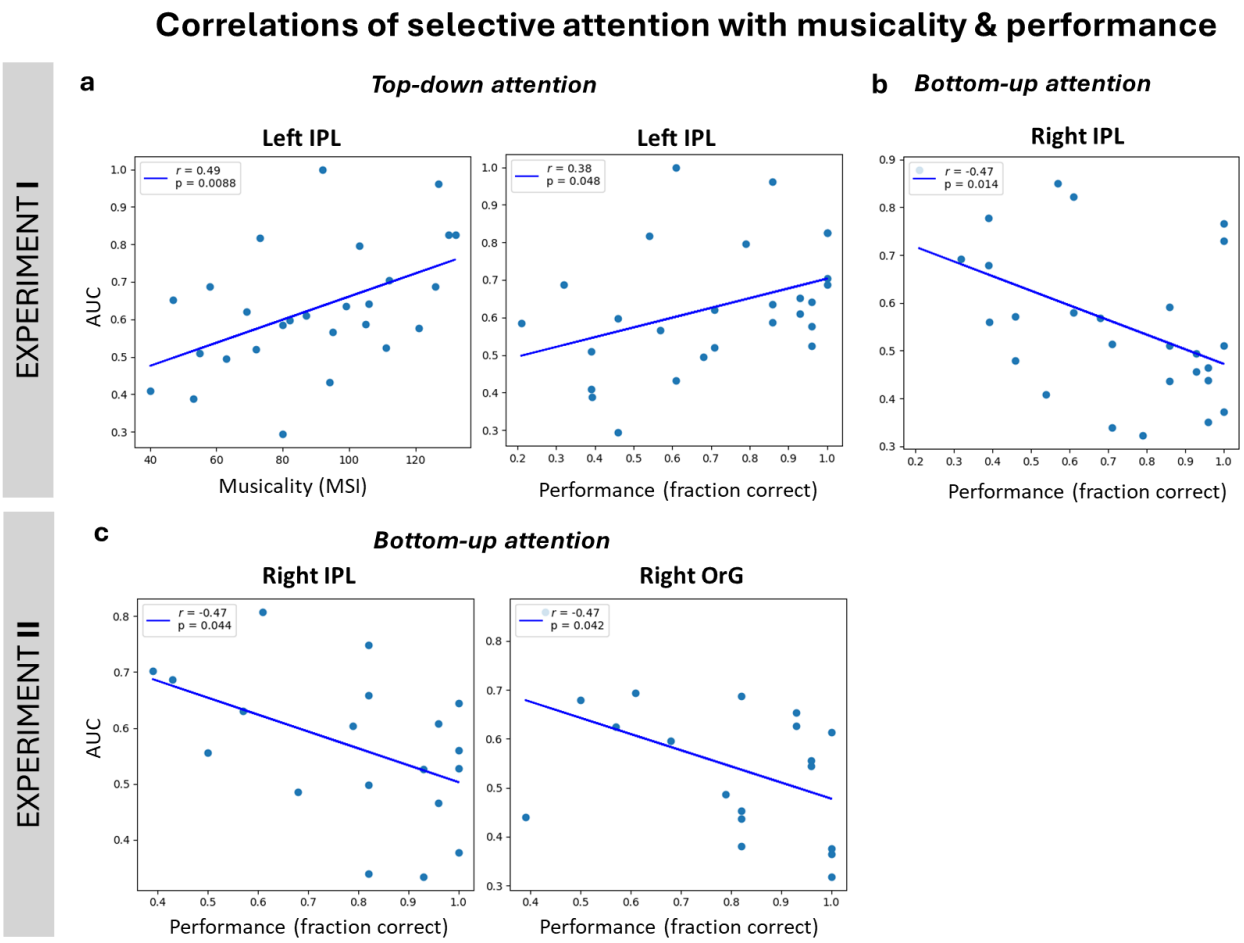

Fig. 5 | Correlations of selective attention with musicality & performance.
Scatterplots showing correlation between classifier area-under-curve (AUC) values against individual musicality or task performance. The AUC is plotted in the y-axis for all subfigures and reflects the degree of selective attention. **a**, In experiment I, top-down attention at the left inferior parietal lobe (IPL) shows positive correlations with both musicality (left; Pearson correlation, r=0.49, p<0.01) and performance (right; Pearson correlation, r=0.38, p<0.05) across 27 participants. **b**, Across the same participants, bottom-up attention correlated negatively with performance at the right IPL (Pearson correlation, r = -0.47, p < 0.05). **c**, Similarly, in experiment II, bottom-up attention correlated negatively with performance at the right IPL and right orbital gyrus (OrG) across 19 participants (Pearson correlation, r = -0.47, p < 0.05 for both regions).

For task performance, a positive correlation with top-down attention was also observed at the left IPL in experiment I (Pearson correlation, r = 0.38, p < 0.05; Fig. 5a - Right). Contrastingly, the correlation between bottom-up attention and task performance was negative at the right IPL (Pearson correlation, r = -0.47, p < 0.05; Fig. 5b). This negative correlation, although weaker, was already observed at sensor level (Pearson correlation, r = -0.45, p < 0.05; Extended Data Fig. 3b). Furthermore, in experiment II, the early half of the tone showed a negative correlation between bottom-up attention and task performance, but not the late half. This effect was also observed at the right IPL (Pearson correlation, r = -0.47, p< 0.05; Fig. 5c - Left), like in experiment I, as well as at the right OrG (Pearson correlation, r = -0.47, p < 0.05; Fig. 5c - Right). The early half of the tone was thought to elicit a stronger bottom-up attention effect (on both melodies) due to the perceptually distinct change in pitch. In contrast, the late half, being a continuation of the same tone, would not trigger a salient bottom-up effect. Hence, we infer that the negative correlation manifested from the effect of bottom-up attention rather than top-down attention which was unlikely to differ systematically between the two halves of the tone. Taken together, the results from both experiments provide compelling evidence that enhancements in selective attention across individuals, likely from

musical training, occur by enhancing top-down mechanisms while suppressing bottom-up mechanisms in the frontoparietal regions.

Musical training aids the maintenance and improvement of selective attention over time via frontal mechanisms, enhancing task performance

To investigate temporal dynamics of selective attention, we employed a sliding time window to calculate AUC values over the 2s tone duration in experiment II. The time of peak selective attention, corresponding to peak discrimination, appeared to follow a bimodal distribution across participants, with a separation at around 0.5 s (Fig. 6a). Participants were thus categorized into two equal groups based on their time of peak selective attention: Early attendees, whose selective attention peaked before 0.5 s from tone onset, and late attendees, whose selective attention peaked after 0.5 seconds. We found that late attendees performed significantly better ($\Delta = 0.19$, $p < 0.05$, $n = 10000$, two-tailed permutation test) at the task than early attendees (Fig. 6b). Moreover, late attendees tended to be more musical than early attendees, even though the difference nearly missed significance ($\Delta = 34.3$, $p = 0.057$, $n = 10000$, two-tailed permutation test).

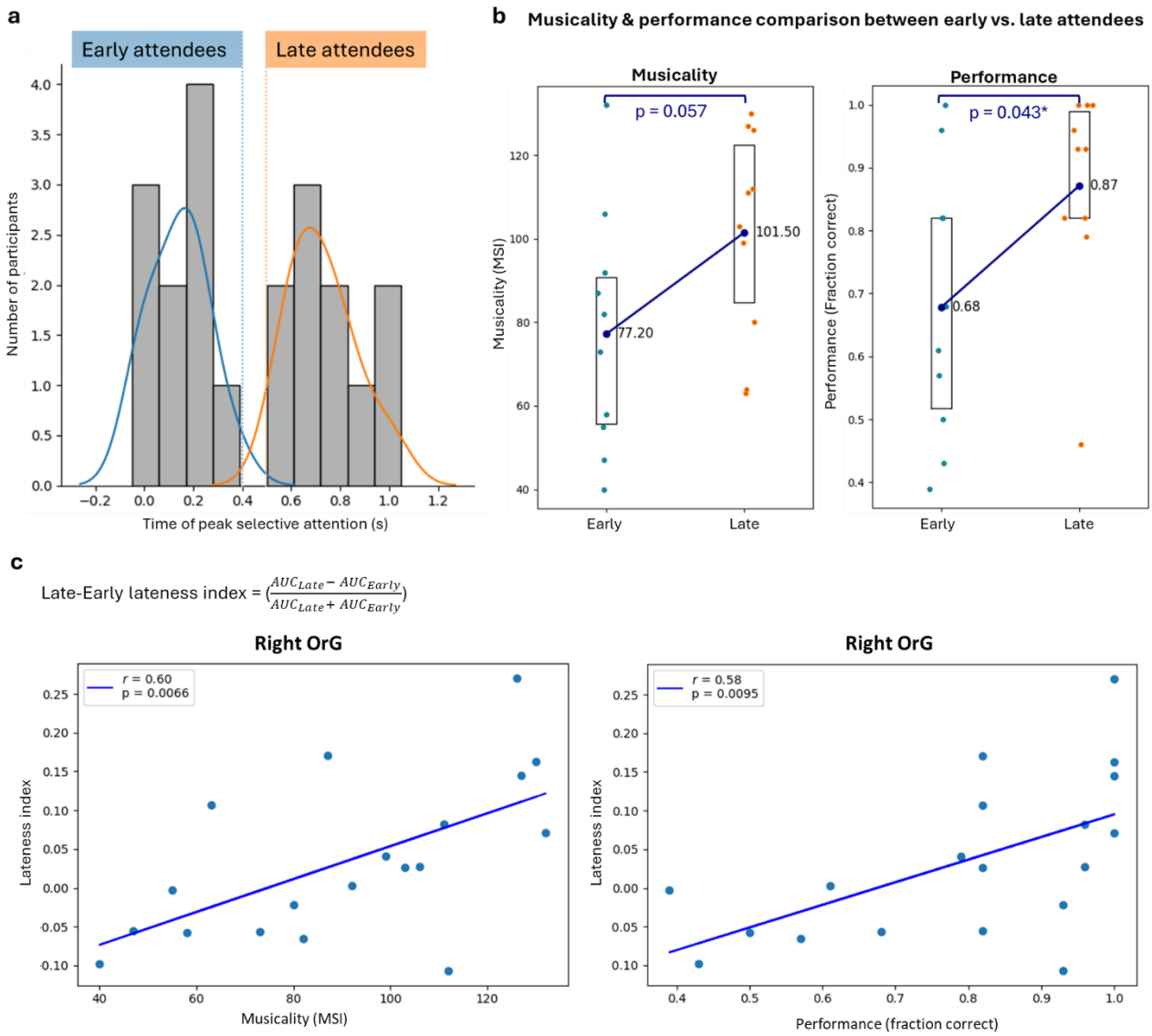

Fig. 6 | Musicality and performance across early and late attendees.
a, Histogram of the time of peak selective attention across participants. We computed the classifier AUC over the 2s tone duration using a sliding time window and extracted the time of peak selective attention, corresponding to maximum AUC, for each of the 20 participants in experiment II. The distribution appeared to be bimodal, dividing participants equally into "early" (blue) or "late" (orange) attendees depending on whether their attention peaked before of after 0.5 s respectively. Colored lines represent normal distributions fitted to each group of attendees. **b**, Musicality & performance comparison between early vs. late attendees. Late attendees performed significantly better ($\Delta = 0.19$, $p < 0.05$, n = 10000, two-tailed permutation test) at the task than early attendees. Moreover, late attendees tended to be more musical than early attendees, even though the difference nearly missed significance ($\Delta = 34.3$, $p = 0.057$, n = 10000, two-tailed permutation test). Box plots represent the 25th – 75th percentiles of the data, with the center dot indicating the mean. **c**, Correlational analysis of lateness index with musicality and performance. For each participant, we computed a 'lateness index' that reflects the relative strength of selective attention in the late half of the tone compared to the early half. The lateness index correlated positively with both musicality (left) and task performance (right) at the right orbital gyrus (OrG) (Pearson correlation, $p < 0.01$ for musicality & performance). Taken together, these results suggest that musical training sharpens the neural mechanisms for sustaining or improving auditory selective attention over time, particularly in the right prefrontal cortex, thereby boosting task performance.

At source level, instead of employing a sliding time window (to save on computational costs), we directly compared selective attention in the early half of the tone with its late half. To quantify the difference, we computed a 'lateness index' for each participant, reflecting the relative strength of selective attention in the late half compared to the early half. Our findings at the right OrG revealed positive correlations between the lateness index and both musicality as well as task performance across participants (Pearson correlation, $p < 0.01$ for musicality & performance; Fig. 6c). Essentially, participants who had better musical skills and task performance paid higher selective attention during the late half compared to the early half of the tone. Conversely, participants who were less

musical and performed poorly could not maintain selective attention throughout the tone duration. These results suggest that musical training sharpens the neural mechanisms for sustaining or improving auditory selective attention over time, particularly in the right prefrontal cortex, thereby boosting task performance.

4. DISCUSSION

The combined findings across both experiments demonstrate that our implementation of frequency-tagging in simultaneous stimuli is both reliable and consistent. Using our unique repeated-splitting support vector machines, we successfully isolated and extracted the effect of selective attention to a target melody within a mixture of two melodies, both at sensor and source levels. Our findings elucidate the differential recruitment of cortical regions in bottom-up and top-down attentional processes, aligning with established theories of attentional control which posit that top-down attention engages higher-level, goal-directed executive mechanisms orchestrated by prefrontal regions, while bottom-up attention is driven by automatic lower-level sensory mechanisms predominantly situated in the primary cortices[12–15]. To the best of our knowledge, this is the first successful classification of simultaneous ASSRs according to different cognitive states. This breakthrough opens up new possibilities for implementing frequency-tagging in cognitive or clinical neuroscience.

Frequency-tagging facilitates the use of simultaneous sounds in neuroscience research, creating environments which more accurately reflect the complexity of natural soundscapes and ensuring that experimental findings are more applicable to real-life cognitive phenomena. Our approach of applying frequency-tagging at full modulation depth, while eliciting stronger signals, noticeably deteriorates sound quality and naturalness. Decreasing the modulation depth mitigates the degradation in sound quality but may render the ASSRs too weak for extracting cognitive effects within realistic time constraints[1,2]. Therefore, it is imperative for future frequency-tagging endeavors to conduct pilot tests aimed at minimizing the modulation depth while preserving sufficient signal strength to capture the desired effects. A potential limitation of this work is that replicating our auditory setup (see *Stimulus presentation* under **Methods**) to achieve near zero-lag is tedious and costly, requiring the purchase and setup of new specialized hardware and software if not already accessible. As a more implementable alternative, we recommend recording the auditory output into a separate channel (MISC) of the MEG system to precisely track the time of stimulus delivery. Importantly, researchers should take into account the Nyquist frequency requirements to capture the auditory signal with sufficiently high fidelity by setting the sampling rate to at least twice the maximum carrier frequency of the auditory stimulus.

Our study provides compelling evidence that top-down and bottom-up attentional mechanisms are differentially influenced by musicality and task performance. We found a positive correlation between top-down attention and both musicality and task performance, particularly in the left IPL. This suggests that musical training enhances top-down attentional mechanisms, which are crucial for goal-directed behavior and cognitive control, leading to better task performance. Conversely, bottom-up attention showed a negative correlation with task performance in the right IPL, indicating that higher sensitivity to distractions via bottom-up attentional mechanisms may impede performance. Together, these insights contribute to a more nuanced understanding of the

neural substrates underlying selective attention, highlighting the distinct yet complementary roles of the left and right parietal cortices in top-down and bottom-up attention. Consistent with our hypothesis, the involvement of the parietal cortex is unsurprising as previous studies[16–18] from our lab and others have demonstrated its role in musical training and the ability to perform music-related tasks. Our analysis of the temporal dynamics of selective attention revealed that participants whose selective attention peaked later during the tone performed significantly better and were more musical than those who peaked earlier. Subsequent correlational analysis corroborated with these results, revealing that the stronger attention is during the late tone half, relative to the earlier tone half, the more musical and better performing the participant was. This effect was evident at the right prefrontal cortex, an established center for attentional control, aligning with the hypothesis that musical training sharpens neural mechanisms for sustaining or improving selective attention over time, thereby boosting task performance.

By tracking how attention towards each melody varies over time with our sliding window analysis, we can infer the attentional strategies employed by individual participants. Although beyond the scope of this paper, it is fascinating that we found correlations between attention towards the 39 and 43 Hz melodies overtime, particularly among participants with higher musicality (Extended Data Fig. 4). The data also suggests that these participants tended to perform better, although the difference was not significant. These observations might indicate that musical participants employ an organized strategy to manage attention towards simultaneous sounds, contributing to their better performance. Using a simple index like the Kendall's Tau, significant correlations between the 39 Hz and 43 Hz attention curves only emerged in five subjects. Visual inspection of the individual curves revealed more complex correlations with slight time lags between the curves for many other participants (Extended Data Fig. 5). We believe that these complex correlations can be more effectively characterized with a more appropriate index. Given that this area falls outside our primary expertise, we invite constructive suggestions from readers to enhance similar analyses in the future.

Overall, this study contributes to a deeper understanding of the neural substrates underlying selective attention and highlights the potential of musical training as a tool for cognitive enhancement. Further research could expand on these findings by exploring whether instrument-specific musical training impedes attentional control away from the trained instrument tones (e.g., whether a violinist finds it more challenging to ignore a violin tone), and elucidating the mechanisms involved. Additionally, biochemical techniques such as magnetic resonance spectroscopy can provide insights into the neurochemical basis, possibly involving GABA and glutamate neurotransmitters[19–23], that differentiate selective attention control between high and low performers, or musicians and non-musicians. Finally, although weaker sources outside the auditory cortex are typically overlooked in ASSR research, our current and past work has demonstrated that secondary sources in the frontoparietal regions can better capture cognitive effects such as selective attention and their correlation with behavioral measures (e.g. musicality, task performance). By emphasizing on their relevance n usefulness, we encourage future work to also consider these secondary ASSR sources in their analyses.

5. METHODS

5.1 Participants

28 participants took part in experiment I (18 – 49 years, mean age = 28.4, SD = 6.2; 9 female; 2 left-handed) and 20 participants took part in experiment II (21 – 49 years, mean age = 28.5, SD = 6.1; 9 female; 2 left-handed). All participants were fluent in English with self-reported normal hearing and participated voluntarily. The experiment was approved by the Regional Ethics Review Board in Stockholm (Dnr: 2017/998-31/2). Both written and oral informed consent were obtained from all participants prior to the experiments. All participants received a monetary compensation of SEK 600 (~ EUR 60).

5.2 Quantification of individual musicality

A subset of the Goldsmiths musical sophistication index (MSI) self-report questionnaire[9,10] containing 22 questions was used to estimate each participant's level of musical sophistication. A copy of the questionnaire used for this study can be found in Extended Data Fig. 6. To ensure relevance to the selective auditory attention task, we concentrated on the Perceptual Ability, Musical Training, and Singing Abilities subscales. The Emotions subscale was excluded because our melody stimuli lack emotional content, rendering this subscale irrelevant to the task. Additionally, several questions from the Active Engagement subcategory were omitted to prioritize musical aptitude and ability over exposure. For instance, questions such as "Music is kind of an addiction for me – I couldn't live without it," which do not directly impact musical ability, were excluded. This specific combination of questions was previously used in a similar selective auditory attention study[6] and demonstrated strong correlations between the resultant MSI and performance. The MSI quantifies a participant's musical skills, engagement, and behavior across multiple facets, making it ideal for testing a general population that includes both musicians and non-musicians. Among all participants, MSI scores ranged from 40 to 132, with a maximum possible score of 154 (m_I = 88.4, std_I=26.8; m_{II} = 89.4, std_{II} = 28.9).

5.3 Task

For both experiments I and II, participants were presented with two simultaneous melodies of different pitch and instructed to selectively attend to either one, following a cue, while following its pitch contour. Each melody was composed of a series of 2 s tones and lasted for 8 - 26 s. When the melody stopped, participants were required to report the last direction of pitch change for the attended melody, which was either falling, rising or constant, corresponding to a chance level of 33 % (see Fig. 1) with a button press. In each experiment, a total of 28 button responses were collected over approximately 10 minutes of MEG recording time for each participant. The fraction of correct button responses, out of 28, was computed as an index of task performance. To ensure that the sounds were separated based on features (in this case, pitch and timing) rather than location (i.e. left-versus-right), the stimuli was presented identically to both ears via insert earphones (model ADU1c, KAR Audio, Finland).

Experiment I. Tone onsets alternate between the low-pitched and high-pitched melodies, beginning either with either melody (order balanced across trials). As each new tone would draw an effect of

bottom-up attention towards it, this experimental design allowed us to dissociate top-down (from the cue) and bottom-up attention effects towards each melody.

Experiment II. Tone onsets for both melodies occurred concurrently, thereby engaging bottom-up attention simultaneously towards both melodies.

5.4 Stimuli

Frequency-tagging. The melodies were constructed with 2s long sinusoidal tones that were generated and amplitude-modulated using the Ableton Live 9 software (Berlin, Germany). The carrier frequencies of the low-pitched melody tones and high-pitched melody tones were 131 - 220 Hz and 329 – 523 Hz respectively. 256 tones were used to construct all 28 melodies of each pitch. For simplicity, only tones in the C major harmonic scale were used. Sinusoidal amplitude modulation of the tone was carried out by modifying its amplitude envelope, which corresponds to its raw sound volume, with regular increases and decreases according to a sine wave (see Extended Data Fig. 7). The rate of modulation, or modulation frequency (f_m), was 39 Hz for the low-pitched melody and 43 Hz for the high-pitched melody. The modulation frequency was centered around 40 Hz and the modulation depth (m) was set at 100% to maximize cortical ASSR power[1].

Stimulus presentation. The melody duration was randomized to last for 8 - 26 s each, reducing predictability of when it will end and thus maintaining participants' high attention throughout playback. The volume of the high-pitched melody was reduced by 10 Db relative to the low-pitched melody to balance subjective loudness differences across frequency ranges[24]. Identical stimuli were delivered to both ears through sound tubes, with the volume calibrated to approximately 75 dB SPL per ear using a sound meter (Type 2235, Brüel & Kjær, Nærum, Denmark) and slightly adjusted for individual comfort. To ensure that the stimulus was presented with less than 1 ms jitter, we used a specialized auditory setup comprising of the AudioFile Stimulus Processor (Cambridge Research Systems Limited) coupled with customized scripts in Presentation® (Version 18.3, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

5.5 Magnetoencephalography data acquisition & preprocessing

MEG measurements were carried out using a 306-channel whole-scalp neuromagnetometer system (Elekta TRIUXTM, Elekta Neuromag Oy, Helsinki, Finland). Data was recorded at a 1 kHz sampling rate, on-line bandpass filtered between 0.1-330 Hz and stored for off-line analysis. Horizontal eye-movements and eye-blinks were monitored using horizontal and vertical bipolar electrooculography electrodes. Cardiac activity was monitored with bipolar electrocardiography electrodes attached below the left and right clavicle. Internal active shielding was active during MEG recordings to suppress electromagnetic artefacts from the surrounding environment. In preparation for the MEG-measurement, each participant's head shape was digitized using a Polhemus Fastrak. The participant's head position and head movement were monitored during MEG recordings using head-position indicator coils. The acquired MEG data was pre-processed using MaxFilter[25,26] by applying temporal signal space separation (tSSS) to suppress artefacts from outside the MEG helmet and compensate for head movement during the recording, before being transformed to a default head position. The tSSS had a buffer length of 10 s and a cut-off correlation coefficient of 0.98.

5.6 MEG data analysis

Following preprocessing, subsequent MEG data analyses were done using the MNE-Python software package[27]. Firstly, we applied a 1 - 50 Hz bandpass filter to the continuous MEG data. Next, eye movements and heartbeats artefacts were automatically detected from electrooculographic and electrocardiography signals using independent components analysis (ICA, fastICA algorithm), validated visually, and removed for each participant. On average, we removed 4.9 components (SEM 0.2) across participants.

Experiment I data analysis. The continuous MEG data was segmented into 1 s long epochs from each tone onset. Since the tone onsets alternate between the low- and high-pitched melodies, bottom-up attention occurred on either melody tone but never on both. Thus, we defined all epochs into two conditions based on whether bottom-up attention was directed to the low- or high-pitched melody tone. Alternatively, we also defined all epoches into another two conditions based on whether the cue directed top-down attention to the low- or high-pitched melody. In both instances, 228 epochs were defined per condition per participant.

Experiment II data analysis. The first 0 - 1 s from each tone onset was defined as the early half of the tone, while the subsquent 1 - 2 s was defined as the late half of the tone, defining two conditions by time. Subsequently, each of the two time conditions were analyzed separately for top-down attention according to the cue, categorizing the epochs into attend low- or high-pitched for each of the early and late half. As a result, about 114 epochs were defined per condition per participant.

5.7 Classification of attention

Repeated-splitting support vector machine. To classify attention conditions at single-subject level, we designed a specialized "repeated-splitting" decoder that was highly sensitive to ASSRs. These customized algorithms were based on the scikit-learn package in python. For each condition, epochs were randomly divided into five groups and averaged within each group to produce five evoked ASSRs. Next, the five evoked ASSRs are fast fourier transformed (frequency resolution = 1 Hz) to acquire five power spectra. Spectral smoothing was performed using a Hann window in experiment I and a boxcar window in experiment II. For any contrast between two conditions, the corresponding power spectra (5 each condition) were fed into a binary linear support vector machine with all 306 sensors as features, and classified using a 5-fold cross validation method. We computed the area under the receiver operating characteristic curve (AUC) as a measure of discriminability between conditions (chance level = 0.5). Finally, we repeated the entire process from the initial group division step 1000 times and computed the mean AUC across repetitions. The AUC was computed independently for each frequency from 4 - 45 Hz. Fig. 7 outlines each of the steps described above.

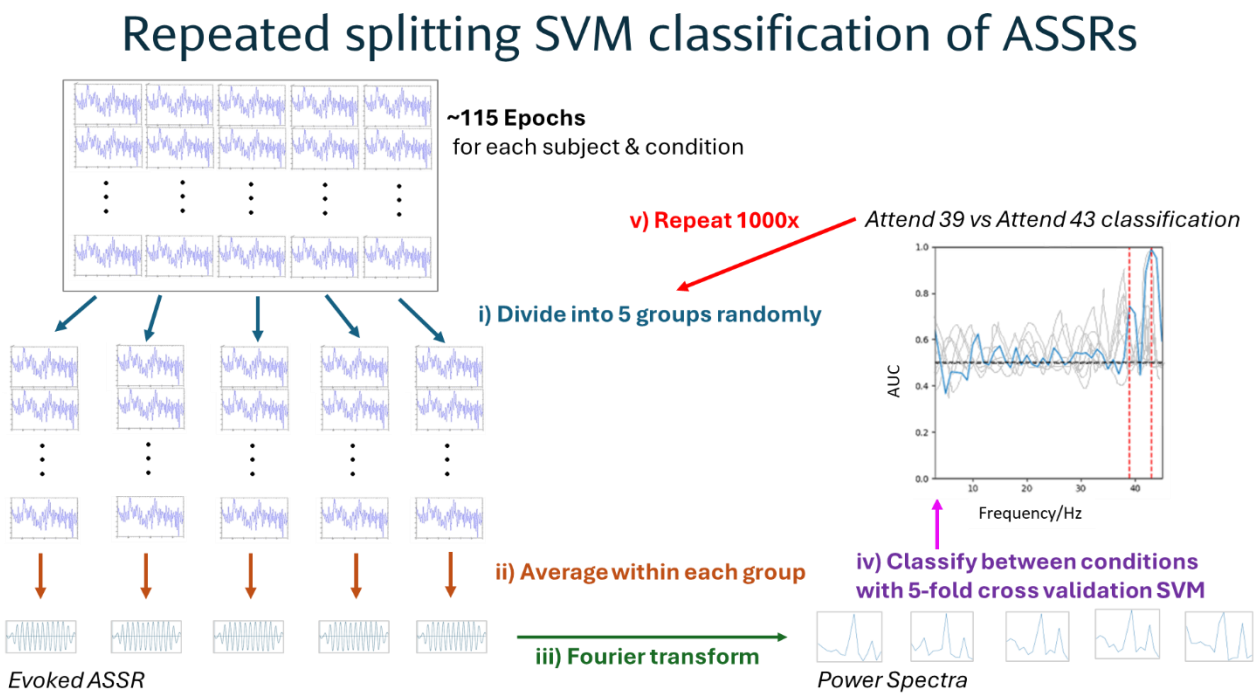

Fig. 7 | Single-subject repeated-splitting support vector machine pipeline for ASSR classification
For each contrast, epochs per condition were randomly divided into five groups and averaged within each group to produce five evoked ASSRs. Next, the five evoked ASSRs were fourier transformed to acquire five power spectra. For any contrast between two conditions, the corresponding power spectra (5 each condition) were classified via cross-validation, obtaining the area under the receiver operating characteristic curve (AUC) as a measure of discriminability between conditions (chance level = 0.5). Subsequently, we repeated the entire process from the initial group division step 1000 times and computed the mean AUC across repetitions. Note that the AUC was computed independently for each frequency from 4 - 45 Hz. For source analysis, the evoked ASSRs were localized to the cortical surface before fourier transformation (step iii). For each region of interest (ROI), only sources within the ROI were included for classification, combining the power at 39 and 43 Hz in feature space. As in sensor analysis, the mean AUC across 1000 repetitions was computed.

For *experiment I*, the AUC across frequency was then averaged across 25 participants (3 participants were excluded from further sensor-level analyses due to excessive noise; see Extended Data Fig. 8) to give the group-averaged AUC across frequency. For each participant in *experiment II*, the AUC across frequency was computed separately for the early and late half of the tone, before combining them by selecting the higher AUC between the two halves at each frequency. The resultant "best of two" AUC across frequency was then averaged across all 20 participants to give the group-averaged AUC across frequency.

Sliding window analysis. To investigate fluctuations in selective attention across time during experiment II, we employed a sliding window approach (fast fourier transform, window length = 1 s, step size = 50 ms, boxcar window) across the 2 s tone duration for the ASSRs, computing the power specifically at 39 and 43 Hz. Classification was performed independently for each time window, resulting in the AUC across time at 39 and 43 Hz, corresponding to selective attention towards the low-pitched and high-pitched melody respectively. For each participant, we determined the time of peak selective attention during which the AUC was maximum, regardless of which melody it corresponded to. Participants were split into two equal groups of ten based on their time of peak selective attention, which was before 0.5 s for early attendees and after 0.5 s for late attendees. Subsequently, we compared differences in musicality and performance between the early and late attendees using a two-tailed permutation test with 10 000 shuffles.

5.8 Structural Magnetic Resonance Imaging acquisition and processing

Anatomical MRIs were acquired using hi-res Sagittal T1 weighted 3D IR-SPGR (inversion recovery spoiled gradient echo) images by a GE MR750 3 Tesla scanner with the following pulse sequence parameters: 1 mm isotropic resolution, FoV 240 × 240 mm, acquisition matrix: 240 × 240, 180 slices 1 mm thick, bandwidth per pixel=347 Hz/pixel, Flip Angle=12 degrees, TI=400 ms, TE=2.4 ms, TR=5.5 ms resulting in a TR per slice of 1390 ms. Cortical reconstruction and volumetric segmentation of all participants' MRI was performed with the Freesurfer image analysis suite[28].

5.9 MEG source reconstruction

Source localization of the MEG signals (combined planar gradiometers and magnetometers) was performed using a minimum-norm estimate[29] (MNE) distributed source model containing 20484 dipolar cortical sources. The MEG signals were coregistered onto each individual participant's anatomical MRIs using previously digitized points on the head surface during MEG acquisition. Based on our previous studies[6–8] on ASSR sources, we demarcated six regions of interest, namely the orbital gyrus (OrG), superior temporal gyrus (STG), and inferior parietal lobe (IPL) at each hemisphere, according to the Brainnetome Atlas[11].

5.10 Classification of selective attention (source level)

To classify attention conditions at individual source level, we incorporated the source localization step into the "repeated-splitting" support vector machine decoder (using all identical parameters previously described in 7 Classification of attention) as follows: For each contrast, epochs per condition were randomly divided into five groups and averaged within each group to produce five evoked ASSRs. The five evoked ASSRs were then source localized to produce five source estimations across the entire cortex. For each of the six regions of interest (ROIs), a subset of these sources was selected to contain only sources originating within the ROI. The data was then fast fourier transformed. The power at 39 and 43 Hz was combined in feature space before being classified into one of two conditions as we measured the decoding performance with the AUC. Finally, we repeated the entire process from the initial group division step 1000 times and computed the mean AUC across repetitions. One participant was excluded due to unsuccessful MRI collection in both experiments, resulting in a final sample of 27 participants in experiment I and 17 participants in experiment II for all source-level analysis.

Lateness Index. For experiment II, we compared the AUC between the early and late tone half to investigate fluctuations in selective attention across time at source level. This was done in place of the sliding window approach at sensor level to reduce computational costs. For each participant, we computed a lateness index using the formula: $\frac{AUC_{Late} - AUC_{Early}}{AUC_{Late} + AUC_{Early}}$ to quantify the relative strength of selective attention in the late compared to the early tone half. Following which, we carried out Pearson correlation tests between the lateness index and MSI scores, as well as between the lateness index and performance scores to examine which brain regions exhibit relationships between lateness and these behavioral measures. We reported two-tailed p values from these tests. Data normality was assessed using the Shapiro-Wilk test at a threshold of $p < 0.05$ for non-normal distributions and visually confirmed through quantile–quantile plots.

6. REFERENCES

1. Roß, B., Borgmann, C., Draganova, R., Roberts, L. E. & Pantev, C. A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J. Acoust. Soc. Am.* **108**, 679–691 (2000).
2. Picton, T. W., John, M. S., Dimitrijevic, A. & Purcell, D. Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *Int. J. Audiol.* **42**, 177–219 (2003).
3. John, M. S., Lins, O. G., Boucher, B. L. & Picton, T. W. Multiple Auditory Steady-state Responses (MASTER): Stimulus and Recording Parameters. *Int. J. Audiol.* **37**, 59–82 (1998).
4. Gander, P. E., Bosnyak, D. J. & Roberts, L. E. Evidence for modality-specific but not frequency-specific modulation of human primary auditory cortex by attention. *Hear. Res.* **268**, 213–226 (2010).
5. Riecke, L., Scharke, W., Valente, G. & Gutschalk, A. Sustained Selective Attention to Competing Amplitude-Modulations in Human Auditory Cortex. *PLoS ONE* **9**, e108045 (2014).
6. Manting, C. L., Gulyas, B., Ullén, F. & Lundqvist, D. Auditory steady-state responses during and after a stimulus: Cortical sources, and the influence of attention and musicality. *NeuroImage* **233**, 117962 (2021).
7. Manting, C. L., Gulyas, B., Ullén, F. & Lundqvist, D. Steady-state responses to concurrent melodies: source distribution, top-down, and bottom-up attention. *Cereb. Cortex* **33**, 3053–3066 (2023).
8. Manting, C. L., Andersen, L. M., Gulyas, B., Ullén, F. & Lundqvist, D. Attentional modulation of the auditory steady-state response across the cortex. *NeuroImage* **217**, 116930 (2020).
9. Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. Measuring the facets of musicality: The Goldsmiths Musical Sophistication Index (Gold-MSI). *Personal. Individ. Differ.* **60**, S35 (2014).
10. Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE* **9**, e89642 (2014).
11. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex N. Y. NY* **26**, 3508–3526 (2016).
12. Foster, J. K., Eskes, G. A. & Stuss, D. T. The cognitive neuropsychology of attention: A frontal lobe perspective. *Cogn. Neuropsychol.* **11**, 133–147 (1994).
13. Huang, S., Belliveau, J. W., Tengshe, C. & Ahveninen, J. Brain Networks of Novelty-Driven Involuntary and Cued Voluntary Auditory Attention Shifting. *PLoS ONE* **7**, e44062 (2012).
14. Cohen, R. A. Attention and the Frontal Cortex. in *The Neuropsychology of Attention* (ed. Cohen, R. A.) 335–379 (Springer US, Boston, MA, 2014). doi:10.1007/978-0-387-72639-7_13.
15. Plakke, B. & Romanski, L. M. Auditory connections and functions of prefrontal cortex. *Front. Neurosci.* **8**, (2014).
16. Rosenkranz, K., Williamon, A. & Rothwell, J. C. Motorcortical excitability and synaptic plasticity is enhanced in professional musicians. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 5200–5206 (2007).
17. Gärtner, H. *et al.* Brain morphometry shows effects of long-term musical practice in middle-aged keyboard players. *Front. Psychol.* **4**, 636 (2013).
18. de Aquino, M. P. B., Verdejo-Román, J., Pérez-García, M. & Pérez-García, P. Different role of the supplementary motor area and the insula between musicians and non-musicians in a controlled musical creativity task. *Sci. Rep.* **9**, 13006 (2019).
19. Sumner, P., Edden, R. A. E., Bompas, A., Evans, C. J. & Singh, K. D. More GABA, less distraction: a neurochemical predictor of motor decision speed. *Nat. Neurosci.* **13**, 825–827 (2010).
20. Falkenberg, L. E., Westerhausen, R., Specht, K. & Hugdahl, K. Resting-state glutamate level in the anterior cingulate predicts blood-oxygen level-dependent response to cognitive control. *Proc. Natl. Acad. Sci.* **109**, 5069–5073 (2012).
21. Kihara, K., Kondo, H. M. & Kawahara, J. I. Differential Contributions of GABA Concentration in Frontal and Parietal Regions to Individual Differences in Attentional Blink. *J. Neurosci.* **36**, 8895–8901 (2016).
22. von Düring, F. *et al.* Glutamate in Salience Network Predicts BOLD Response in Default Mode Network During Salience Processing. *Front. Behav. Neurosci.* **13**, (2019).
23. Dobri, S. G. J. & Ross, B. Total GABA level in human auditory cortex is associated with speech-in-noise understanding in older age. *NeuroImage* **225**, (2021).

24. Robinson, D. W. & Dadson, R. S. A re-determination of the equal-loudness relations for pure tones. *Br. J. Appl. Phys.* **7**, 166 (1956).
25. Taulu, S., Kajola, M. & Simola, J. Suppression of Interference and Artifacts by the Signal Space Separation Method. *Brain Topogr.* **16**, 269–275 (2004).
26. Taulu, S. & Simola, J. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* **51**, 1759 (2006).
27. Gramfort, A. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, (2013).
28. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).
29. Hämäläinen, M. S. & Ilmoniemi, R. J. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* **32**, 35–42 (1994).

7. ACKNOWLEDGEMENTS

Data for this study was collected at the Swedish National Facility for Magnetoencephalography, Karolinska Institutet, Sweden. This facility is supported by Knut and Alice Wallenberg Foundation (KAW2011.0207). The present studies were supported by the Swedish Foundation for Strategic Research (SBE 13-0115). The computations for source analysis were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre.

8. COMPETING INTERESTS

The authors declare no competing interests.