Research paper

Individual differences in attentional modulation of cortical responses correlate with selective attention performance

Inyong Choi [a], Le Wang [a], Hari Bharadwaj [a,b], Barbara Shinn-Cunningham [a,b,*]

[a] Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA 02215, USA
[b] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

ARTICLE INFO

ABSTRACT

Many studies have shown that attention modulates the cortical representation of an auditory scene, emphasizing an attended source while suppressing competing sources. Yet, individual differences in the strength of this attentional modulation and their relationship with selective attention ability are poorly understood. Here, we ask whether differences in how strongly attention modulates cortical responses reflect differences in normal-hearing listeners' selective auditory attention ability. We asked listeners to attend to one of three competing melodies and identify its pitch contour while we measured cortical electroencephalographic responses. The three melodies were either from widely separated pitch ranges ("easy trials"), or from a narrow, overlapping pitch range ("hard trials"). The melodies started at slightly different times; listeners attended either the leading or lagging melody. Because of the timing of the onsets, the leading melody drew attention exogenously. In contrast, attending the lagging melody required listeners to direct top–down attention volitionally. We quantified how attention amplified auditory N1 response to the attended melody and found large individual differences in the N1 amplification, even though only correctly answered trials were used to quantify the ERP gain. Importantly, listeners with the strongest amplification of N1 response to the lagging melody in the easy trials were the best performers across other types of trials. Our results raise the possibility that individual differences in the strength of top–down gain control reflect inherent differences in the ability to control top–down attention.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Over the past few decades, electroencephalographic (EEG) studies have demonstrated that selective auditory attention modulates event-related potentials (ERPs) generated by neural activity in auditory cortex (Hillyard et al., 1973, 1998; Picton and Hillyard, 1974; Woldorff et al., 1993; Choi et al., 2013). These studies suggest that attention operates as a form of sensory gain-control, amplifying the representation of an attended object and suppressing the representation of ignored objects. Recent physiological studies using electrode recordings inside the scalp (Mesgarani and Chang, 2012; Zion Golumbic et al., 2013) as well as noninvasive magnetoencephalography (Chait et al., 2010; Xiang et al., 2010; Ding and Simon, 2012) and high-density EEG with source

localization (Kerlin et al., 2010; Power et al., 2011) support the theory that selective auditory attention involves a direct modulation of sensory responses.

Despite growing evidences that normal-hearing listeners differ greatly in their behavioral attention ability (Ruggles and Shinn-Cunningham, 2011), the factors contributing to these individual differences are not yet understood. Here, we explore whether individual differences in the strength of attentional modulation of auditory ERPs correlates with individual differences in selective attention performance.

Activity in sensory cortex has been shown to correlate with behavioral performance in some past studies. For instance, strong preparatory activity is associated with good performance on a visual detection task (Ress et al., 2000; Linkenkaer-Hansen et al., 2004). It is also reported that the strength of attentional modulation of sensory responses varies with task difficulty (Martínez-Trujillo and Treue, 2002; Atiani et al., 2009). Simultaneous MEG and behavioral studies using polyrhythmic tone sequences showed that behavioral performance improved and neural responses to a target stream were enhanced over time as

listeners attended the target (Elhilali et al., 2009; Xiang et al., 2010). Subjects' engagement with a task also influences both attentional modulation and behavior (Hill and Schölkopf, 2012). However, it is still not known whether individual differences in the strength of attentional modulation are consistent across conditions, suggesting that there are inherent differences in the efficacy of attentional control, or even whether individual differences in attentional modulation are present when a task is easy and all subjects perform well.

In the current study, we quantify how individual subjects' N1 responses to sources in a sound mixture are modulated by attention when selective attention is successfully performed. The strength of N1 modulation during successful deployment of selective attention is compared with behavioral performance for similar tasks, but for conditions varying in their levels of difficulty. The difficulty of the task was titrated by varying the pitch separation of competing melodies. The timing of the competing melodies ensured that one (the leading melody) was salient in the mixture and the other melody, started soon after the leading melody, was less salient. Because exogenous attention was drawn to the leading melody, focusing attention on the lagging melody required explicit top—down, endogenous attention. Listeners all performed well on the easy, different-pitch trials, even when the less-salient lagging melody was the target. This allowed us to quantify how strongly each individual subject modulated his or her ERPs when successfully directing endogenous attention to the lagging melody (compared to when they ignored both melodies). On same-pitch trials, where it was difficult to segregate the competing melodies, individual differences in behavioral performance were large, differentiating listener ability. Critically, we found that the strength of neuronal modulation in trials where top—down attention was successful was correlated with behavioral performance. These results support the idea that subjects vary in how effectively they gate sensory information and that such differences are related to behavioral ability.

2. Methods

2.1. Subjects

Eighteen normal hearing volunteers (six male, aged 20—35 years) participated in the experiments. All provided written informed consent as approved by the Boston University Institutional Review Board. Subjects were compensated at an hourly base rate, and received an additional bonus of $0.02 for each correct response (up to $8.64, depending on performance).

2.2. Auditory stimuli

Melodies were generated using Matlab (Mathworks, Natick, MA). Each note contained six harmonics added in cosine phase with magnitude inversely proportional to the frequency. Each note had a slowly decaying exponential time window (time constant 100 ms) and cosine squared onset (duration 10 ms) and offset (duration 100 ms) ramps to reduce spectral splatter.

On each trial, three isochronous melodies were presented (see Fig. 1A): a center melody that started first and was always ignored (interaural time difference or ITD of 0), one melody from the left (ITD of −100 μs), and one from the right (ITD of −100 μs). By design, the timing of the notes in the melodies was offset in time to allow us to isolate the ERPs evoked by each melody. The center melody consisted of three 1-s-duration notes. The leading melody started 0.6 s after the center melody and came from either the left or right, selected randomly with equal chance. The leading melody consisted of four notes, each of 0.6 s duration. The lagging melody started 0.15 s after the leading melody, from the opposite direction (three notes, each of 0.75 s duration). Given the timings of the three streams, we hypothesized that the onset of the leading melody would draw attention exogenously; in contrast, we expected the lagging melody to be hard to focus on, and to require strong volitional control, as 150 ms is near the limit for how long it
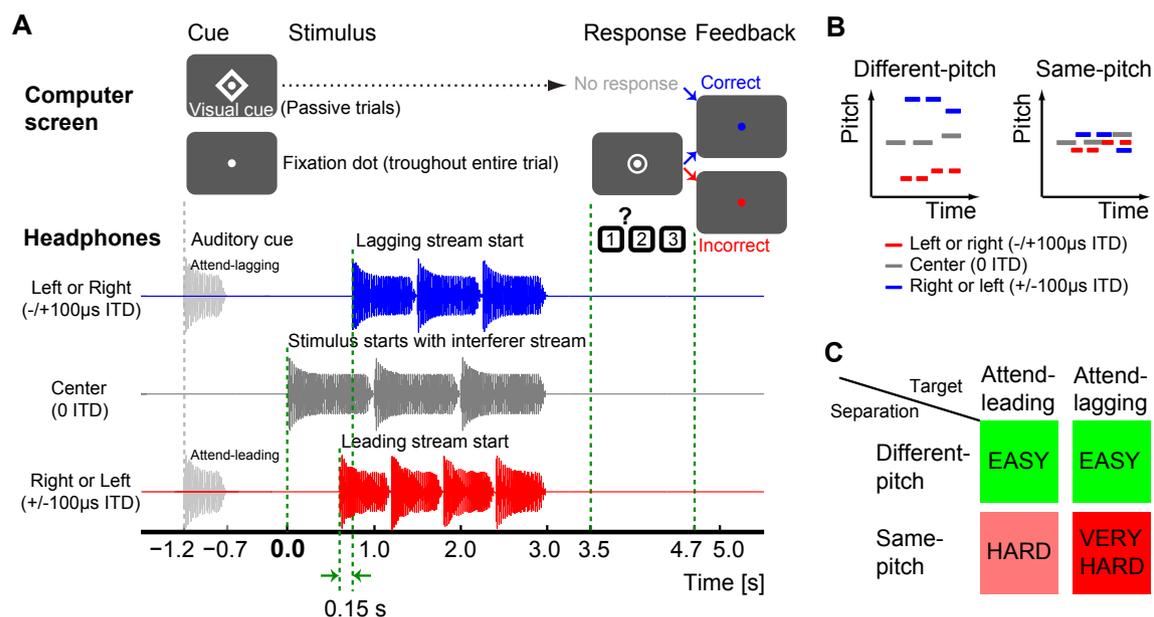
Fig. 1. Experimental design. A. Trial structure. Passive trials begin with a visual fixation point bracketed by "<>". Attention trials begin with an auditory cue indicating the location and pitch of the to-be-attended melody. The auditory mixture always begins with a center melody (gray waveform). The leading and lagging melodies then begin 0.60 and 0.75 s after the center (red and blue waveforms, respectively). The response period, marked by a circle surrounding the fixation point, begins at 3.5 s. Visual feedback was given for 300ms after the response period (a blue dot to indicate a correct answer or a red dot to indicate an incorrect answer). B. Schematic spectrogram of stimuli for the three melodies in different-pitch and same-pitch conditions. C. Illustration of the hypothesized difficulties of the four different trial types. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

takes to redirect attention from one stream to another (Best et al., 2008).

With equal chance on a given trial, the three melodies were constructed from either different or the same pitch ranges (see Fig. 1B). In "different-pitch" trials, the fundamental frequencies (f0s) of the notes in the left, middle, and the right melodies were in the ranges 600−726 Hz, 320−387 Hz, and 180−218 Hz, respectively. In "same-pitch" trials, all melodies were in the same range (320−387 Hz). Within each pitch range, there were three potential pitches (low, mid, and high). For leading and lagging melodies, on each trial two distinct pitches (low and mid, mid and high, or low and high) were randomly selected from the appropriate range to create ascending, descending, or zigzagging contours (chosen independently for the two melodies). For ascending trials, a transition point was selected randomly; all notes prior to the transition were set to the lower pitch and all subsequent notes set to the higher pitch. Similarly, descending trials were constructed from higher and then lower pitches. Zigzagging melodies were created by randomly selecting a transition point before the penultimate note. All notes prior to this were then randomly set to be either low or high, all the subsequent notes except for the final note were set to the opposite value (high or low), and the final note was set the same as the initial notes (i.e., a sequence like L-H-L-L was never generated to preclude listeners from identifying the contour shape before the final note). Note that the always-ignored center melody was constructed by simply randomly selecting notes from the three pitches in the 320−387 Hz range. For same-pitch trials, the contours were constructed to avoid cases in which temporally overlapping notes had the same pitch.

2.3. Experimental design

On each trial, listeners were presented with three concurrent melodies (left, right, and center). On 2/3 of the trials, an auditory cue indicated whether the listener should attend to the left or right melody; after the trial ended, listeners were then asked to identify the shape of that melodic contour (ascending, descending, or zigzagging). The remaining 1/3 of the trials were passive controls, where a preceding visual cue told the listener to ignore all three melodies and refrain from responding. We designed the task to maximize the degree to which behavioral differences were related to attentional control, rather than other factors. Specifically, the task required subjects to focus and maintain selective attention, but had low cognitive and memory demands, reducing the influence of these other cognitive processes on performance. The task required listeners to attend to isochronous melodies with different presentation rates. Moreover, all pitch differences within a melody were very salient. This ensured that peripheral sensory coding was not the primary factor limiting selective auditory attention performance.

The experiment was controlled using Matlab with the Psychtoolbox 3 extension (Brainard, 1997). Sound stimuli were presented using Etymotic (Elk Grove Village, IL) ER-1 insert headphones connected to a Tucker−Davis Technologies (TDT, Alachua, FL) System 3 unit. The TDT hardware also provided timing signals that were recorded along with the EEG responses. The stimulus sound level was fixed to 70 dB SPL (root-mean-squared).

Subjects sat in a sound-treated booth and fixed their gaze on a dot at the center of a computer screen throughout a trial. A trial began with either a 500-ms auditory cue (in selective-attention trials) or a 500-ms diamond-shaped visual cue (passive trials). The auditory cue had an ITD of either −100 μs or +100 μs to direct attention to the appropriate melody. The auditory cue had a pitch consistent with the target melody (f0 = 660 Hz for attend-left, different-pitch trials; f0 = 198 Hz for attend-right, different-pitch

trials; f0 = 352 Hz for same-pitch trials). 700 ms after the cue, the mixture of three melodies was presented (see Fig. 1A). 500 ms after the end of the auditory stimuli, a circle appeared for 1.2 s around the fixation dot to indicate the response period ("1", "2", or "3" for ascending, descending, or zigzagging, respectively). Answers made before or after the response period were marked as incorrect, as were any passive trials where a response was recorded. Visual feedback was given each block (see Fig. 1A).

Subjects repeated training blocks of 12 trials (4 passive trials and 8 auditory-attention trials) until they correctly responded to at least 11 of 12 melodies, presented alone. Each melody had either four 0.6-s-long notes or three 0.75-s-long notes with equal likelihood (like either the leading or lagging melodies in the regular trials, respectively). Seventeen subjects achieved this level within three blocks, but one subject failed to reach the criterion and was dismissed. The seventeen remaining subjects then completed a demo session consisting of two blocks of 12 trials. The first block presented different-pitch trials, and the second presented same-pitch trials. In this demo session, there were no performance criteria that had to be reached before proceeding to the main experiment.

In the main experiment, a total of 432 trials were presented (12 blocks of 36 trials), with 288 selective-attention and 144 passive trials, presented in random order. Each block contained an equal number of different-pitch (easy) and same-pitch (difficult) trials (18 each: 12 for attention trials, 6 for passive trials). After each block, a message on the computer screen displayed the number of correctly answered trials, the bonus earned ($0.02 times the number of correct responses) and bonus missed ($0.02 times the number of incorrect responses) to that point.

2.4. EEG acquisition and analysis

EEG data was collected during the main experiment. A Biosemi ActiveTwo system was used to record EEG data at a sampling rate of 2048 Hz from 32 scalp electrode positions in the standard 10/20 configuration. Two extra electrodes were placed on the mastoids for reference. TDT hardware sent timing signals for all events, which were recorded in an additional channel. Recordings were re-referenced to the average of the two mastoid electrodes and then bandpass filtered from 2 to 50 Hz using a 2048 point FIR filter.

Different-pitch trials were the main focus of EEG analyses, as there were not enough correct responses to allow analysis of correctly answered same-pitch trials. For each different-pitch trial that was correctly answered, we extracted the epoch from −500 ms to 3 s relative to stimulus onset, then down-sampled to 256 Hz. Measurements were baseline corrected using the mean value from −200 ms to 0 ms. Trials whose maximum absolute peak voltage in the five front-central electrodes (AF3, AF4, F3, F4, and Fz) fell into the top 10% of values were rejected from further analysis. Same-pitch trials were also analyzed, separately, for comparison; however, even incorrectly answered trials were included in the analysis since the number of correctly answered trials was too small for some subjects (varying from 20 to 72 trials across subjects).

The attention trials were grouped into "attending-leading" and "attend-lagging" (4-note leading melody or 3-note lagging melody), regardless of target location. For the different-pitch condition, after preprocessing, there were a minimum of 53 trials and a maximum of 68 trials for each subject and attention condition. To fairly compare results, we selected the first 53 trials of each attention type before averaging to get ERPs. For each condition, the magnitude of the N1 ERP component was calculated from the individual-subject average ERPs for each electrode, computed by finding the local minimum within a fixed time window of 100– 200 ms after each note onset. For each subject, the magnitudes of

the N1 in the five front-central electrodes with the strongest auditory-evoked responses (AF3, AF4, F3, F4, and Fz) were averaged together; for brevity, we refer to this across-electrode average value simply as the N1 amplitude.

Inherent individual differences in overall ERP magnitude were large on an absolute scale. We therefore normalized each individual subject's N1 amplitudes by the amplitude of the N1 ERP response to the center-melody onset (the strongest N1, seen at 0.1 s in Fig. 3A) averaged over all attention conditions (note that this response did not depend on attentional condition).

For the same-pitch condition after artifact rejection, between 50 and 72 trials remained across subjects and attention conditions. The first 50 trials of each attention condition were selected and averaged before calculating auditory-evoked N1 amplitudes using the same subsequent steps described above.

For each subject and each attention condition (attend-leading, attend-lagging), we analyzed how attention modulated the N1 amplitudes associated with the attended and ignored melodies. By looking at appropriate time windows, we isolated the N1s elicited by notes in the leading melody and in the lagging melody. Then we quantified the strength of *lag amplification* for each subject as the average difference in lagging-melody N1 amplitudes in the attend-lagging condition minus the passive condition. *Lead amplification* was computed similarly, but for the leading melody. *Lead suppression* and *lag suppression* were calculated as the differences in the ignored-melody N1 amplitudes in the passive case minus when attending the other melody. Note that positive values of amplification and positive values of suppression both are consistent with a gain in the representation of the attended melody relative to the ignored melody (Hillyard et al., 1998; Chait et al., 2010; Choi et al., 2013).

3. Results

3.1. Selective attention ability varies across subjects

The percentage of correct responses in a given condition did not change across the 12 experimental blocks tested. Lines fitting performance as a function of run number for each subject had a mean slope not statistically different from zero (*t*-test; $P = 0.92, 0.09, 0.45$, and 0.09 for different-pitch attend-leading, different-pitch attend-lagging, same-pitch attend-leading, and same-pitch attend-lagging conditions, respectively), suggesting there were neither learning nor fatigue effects. There were also no statistically significant differences in performance for attend-left and attend-right conditions (*t*-test; $P = 0.37, 0.076$ for different-pitch and same-pitch conditions, respectively). Therefore, results were collapsed across experimental block as well as across direction of attention.

Performance was significantly worse for same-pitch compared to different-pitch stimuli ($P < 0.0001$, Wilcoxon rank sum test; compare right and left side of Fig. 2A). Subjects performed the selective attention task very well for the different-pitch stimuli, for both attend-leading and attend-lagging trials (left side of Fig. 2A). There was no statistical difference in performance between attend-leading and attend-lagging trials for the different-pitch stimuli (*t*-test, $P = 0.90$). For the harder, same-pitch stimuli, individual differences were accentuated and performance was significantly better in attend-leading trials than in attend-lagging trials (*t*-test, $P = 0.015$). Indeed, for all but one subject, performance was equivalent or lower in the attend-lagging, same-pitch condition than in the attend-leading, same-pitch condition.

Individual listeners' performance in the attend-leading and attend-lagging conditions was correlated both for the different-pitch stimuli (rank correlation Kendall $\tau_{17} = 0.71$, $P < 10^{-4}$, left

panel of Fig. 2B) and the same-pitch stimuli (Kendall $\tau_{17} = 0.78$, $P < 10^{-5}$, right panel in Fig. 2B). Thus, the ability to focus attention on one melody was closely related to the ability to focus on another melody in the same mixture.

When comparing performance for different-pitch and same-pitch stimuli, the relationship was not as strong as the lead-lag performance correlation within condition (for attend-leading, Kendall $\tau_{17} = 0.34$, $P = 0.075$; for attend-lagging, Kendall $\tau_{17} = 0.41$, $P = 0.031$; see Fig. 2C).

### 3.2. Attention modulates cortical responses

The onset of the leading melody caused a strong N1, regardless of attentional condition (see onset at first red dashed line in Fig. 3A). A repeated-measures ANOVA supports this observation, finding no significant effect of attentional condition on the initial N1 of the leading melody ($F_{2, 32} = 0.52$, $P = 0.60$ for different-pitch condition, $F_{2, 32} = 2.3$, $P = 0.12$ for same-pitch condition. See N1 amplitudes for Note 1 of the leading melody in Fig. 3B). Moreover, this N1 was larger in magnitude than any subsequent N1s, suggesting that exogenous attention always was drawn to the onset of the leading melody, regardless of attentional condition. Given this, we did not include the initial N1 to the leading melody in our subsequent analyses of N1 modulation.

The normalized N1 amplitudes corresponding to the leading-melody and lagging-melody were calculated separately for each subject, note, and attentional condition (see Fig. 3B). At the group level, focusing attention on a given different-pitch melody lead to amplification of the N1s corresponding to the notes 2-4 of the leading melody and all the notes of the lagging melody (see top panels of Fig. 3B; Wilcoxon rank-sum tests: attend-leading vs. passive conditions or attend-leading vs. attend-lagging conditions. One asterisk represents $P < 0.05$, two asterisks represent $P < 0.001$). However, when analyzing N1s of all trials for same-pitch melodies, none of the notes had amplification significantly different than zero (Wilcoxon rank-sum tests, $P > 0.05$ for all pairwise comparisons; see bottom panels of Fig. 3B). Rather, the N1 amplitudes corresponding to notes from same-pitch melodies had smaller means and larger across-subject variance than for notes in the different-pitch melodies.

The N1 amplitudes corresponding to notes 2−4 of the leading melody and the notes 1−3 of the lagging melody were averaged across notes. For the different-pitch condition, the attentional amplification effect at the group level held both for the leading melody N1s (Wilcoxon rank-sum tests: attend-leading vs. passive conditions, $P = 0.016$; attend-leading vs. attend-lagging conditions, $P = 0.0019$] and for the lagging-melody N1s (Wilcoxon rank-sum tests: attend-lagging vs. passive conditions, $P = 0.0059$; attend-lagging vs. attend-leading conditions, $P = 0.036$; see top panel of Fig. 3C). The N1 amplitudes were not statistically different for a melody when listening passively and when that melody was the distracter (Wilcoxon rank-sum tests: leading-melody passive vs. attend-lagging, $P = 0.27$; lagging-melody passive vs. attend-leading, $P = 0.89$). The N1 amplitudes computed across both correctly and incorrectly answered trials for the same-pitch condition did not show any attentional effect (see bottom panels of Fig. 3C) at the group level.

3.3. Neural amplification strength differs across subjects

Fig. 4 shows the distributions of amplification and suppression across subjects for attend-leading and attend-lagging conditions. For the different-pitch condition, of the four metrics considered, only lag amplification was significantly greater than zero at the group level (Wilcoxon signed rank test, $P = 0.00042$, not corrected
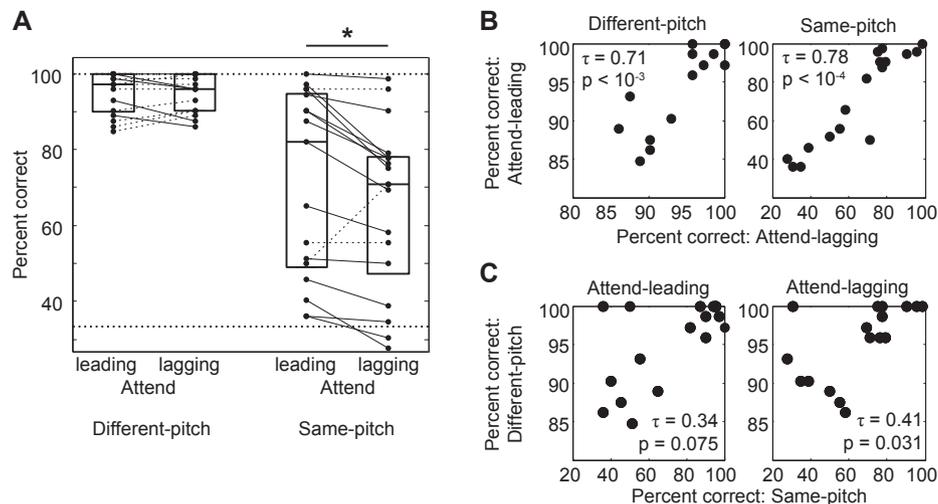
Fig. 2. There are consistent subject differences in performance across stimuli and attentional conditions. A. Summary of behavioral performance for the two types of stimuli (different-pitch or same-pitch) and attentional conditions (attend-leading or attend-lagging). Boxes denote the 25th − 75th percentile range; the horizontal bar in the center denotes the median. Each symbol represents one subject in one condition. Solid lines connect points for subjects who performed better for attend-leading than attend-lagging; dashed lines the reverse. Top and bottom horizontal dashed lines correspond to the perfect performance (100%) and chance (33%), respectively. In the different-pitch conditions, performance is uniformly high and not significantly different between attend-leading and attend-lagging conditions (at the group level). In the same-pitch conditions, performance is significantly worse in the attend-lagging condition than in the attend-leading condition (see text). B. Relationship between attend-leading and attend-lagging performance in different-pitch trials (left panel) and same-pitch trials (right panel). Performance for leading and lagging melodies is significantly correlated for both types of stimuli. C. Relationship between different-pitch and same-pitch performance in attend-leading trials (left panel) and attend-lagging trials (right panel). Performance is not significantly correlated for different-pitch and same-pitch stimuli when identifying the leading-melody contour (left panel) but significantly correlated when identifying the lagging-melody contour (right panel), albeit weakly.

for multiple-testing); the zero-median hypotheses could not be rejected for the other measures ($P = 0.062, 0.14$, and 0.94 for lead amplification, lead suppression, and lag suppression, respectively). Importantly, though, at the individual subject level, attention causes the neural representation of the attended melody to be enhanced compared to the neural representation of the competing melody for the majority of subjects (symbols are located above the negative diagonals in Fig. 4). Still, many listeners showed either negative amplification or suppression.

Amplification and suppression were not significantly correlated (see Fig. 4); listeners who strongly amplified the target melody did not strongly suppress the competing melody. Instead, if anything there was weak (albeit insignificant) negative relationship between amplification and suppression. Inter-subject variation of *lag amplification* was the smallest among all four metrics, but still quite large, ranging from near zero to 0.4 (40% of the size of the center-melody N1).

The analysis for the same-pitch trials did not show significant group-level attentional effect; neither amplification nor suppression was significantly greater than zero (see bottom panels of Fig. 4).

3.4. Behavioral ability correlates with selective amplification

Lead and *Lag amplification*, the two ERP metrics that showed a significant attention effect at the group level (see Section 3.2, Fig. 3C), were compared with behavioral performance for the four conditions (attend leading and lagging for each of different and same-pitch trials). Amplification of the lagging melody for different-pitch stimuli was positively correlated with behavioral performance in most conditions (see Fig. 5B). The correlation was stronger for performance in the same-pitch conditions (Kendall rank correlation $\tau_{17} = 0.48$, $P = 0.0082$ for attend-leading performance, $\tau_{17} = 0.52$, $P = 0.0039$ for attend-lagging performance; see bottom panels of Fig. 5B), which was the harder condition with the greater inter-subject variation in performance. *Lead amplification*,

in contrast, was not significantly correlated with performance (see Fig. 5A). When the Benjamini-Hochberg procedure was applied to control the false discovery rate (Benjamini and Hochberg, 1995), three out of the eight comparisons were found to have statistically significant correlations: the correlations between *lag amplification* and attend-leading performance for both the same-pitch and the different-pitch conditions, and the correlation between *lag amplification* and attend-lagging performance for the same-pitch condition (see Fig. 6 that shows the eight *P*-values in increasing order). When more conservative Bonferroni correction was applied, the only correlation found to be significant was between *lag amplification* and same-pitch, attend-lagging performance ($P < 0.00625$).

Neither lead suppression nor lag suppression was significantly correlated with performance for any of the conditions. Lead and lag amplification measured from the same-pitch condition did not show significant correlation with behavioral performance either.

3.5. Other cognitive factors help account for individual differences

Listeners likely differed in their engagement, fatigue, general ability, and other high-level factors. To ameliorate such effects, we tried to motivate subjects by paying performance-based incentives and providing correct-answer feedback after each trial. This was true for passive trials, too: subjects received a pay bonus for correctly refraining from responding on passive trials; however, many subjects nonetheless responded on some passive trials.

False-hit rates (FHRs) on passive trials ranged from 0% up to 9.72%. Although the FHRs were not significantly correlated with performance for the different-pitch stimuli (see top panels of Fig. 7A; Kendall $\tau = -0.19$, $p = 0.36$ and Kendall $\tau = -0.25$, $p = 0.23$ for attend-leading and attend-lagging trials, respectively), they were significantly negatively correlated with performance for the difficult, same-pitch stimuli (see bottom panels of Fig. 7A; Kendall $\tau = -0.56$, $p = 0.0036$ and Kendall $\tau = -0.44$, $p = 0.020$ for attend-leading and attend-lagging trials, respectively). The FHRs were not
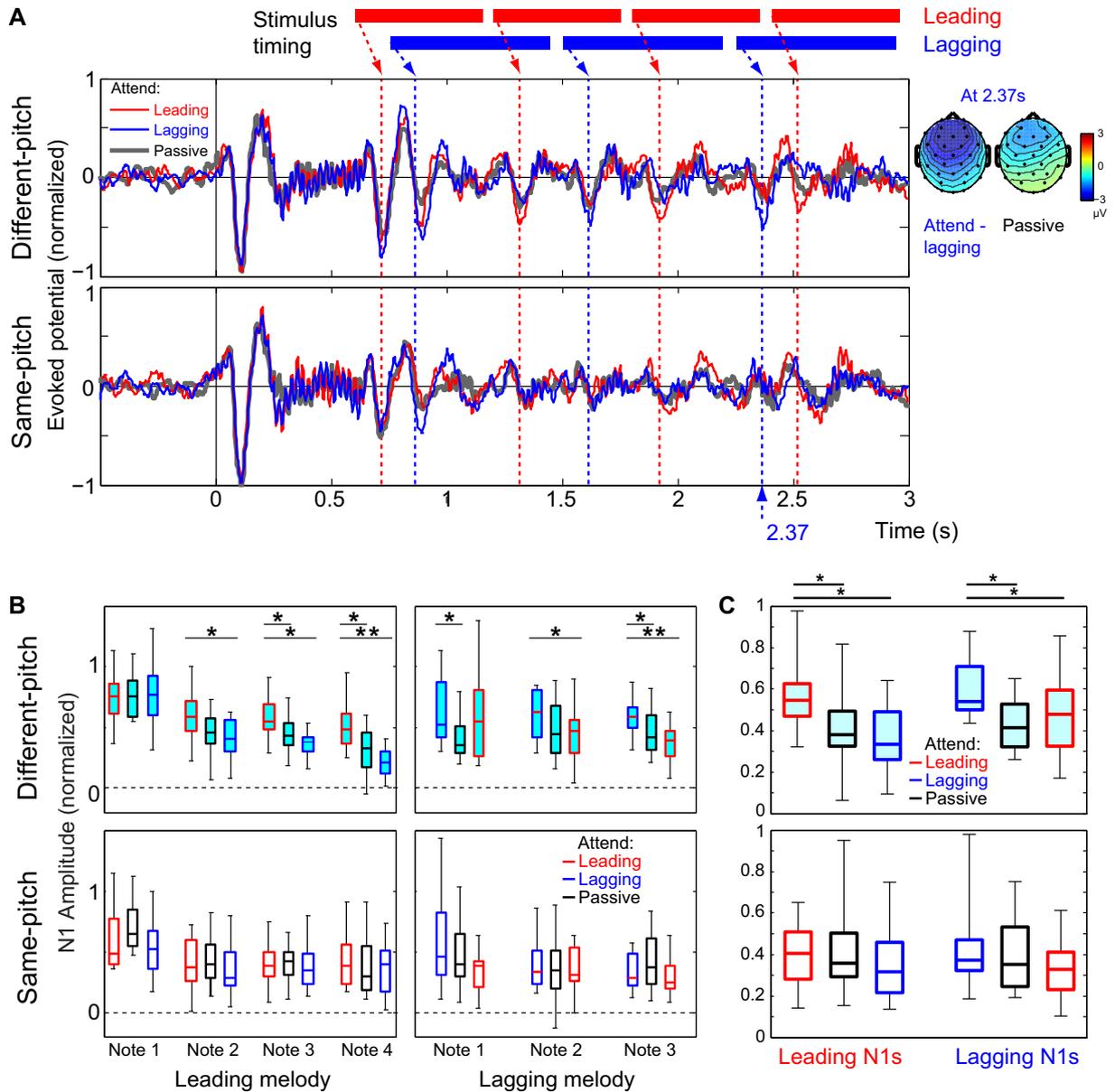
Fig. 3. Attention amplifies ERPs to attended melodies. A. Grand average ERP waveforms measured during correctly answered different-pitch trials (top) and all same-pitch trials (bottom). Note that there were not enough correctly answered same-pitch trials to support ERP analysis of only those trials; as a result, it is not fair to directly compare top and bottom panels. Red, blue, and gray waveforms represent attend-leading, attend-lagging, and passive trials, respectively. Note that each individual subject's ERPs were normalized by the magnitude of the N1 response to the center-melody onset, to account for inherent individual differences in N1 amplitudes, and then averaged across subjects. The example topographies on the right show the scalp distribution of N100 amplitude for the third note in the lagging stream (at 2.37 s, a lagging note N1), for the attend-lagging and passive conditions (left and right plots, respectively). ERPs magnitudes depend on attentional state. B. Average N1 amplitudes for individual notes in different attentional conditions (re: the subject-specific average N1 for that stimulus). Top panels are for correctly answered different-pitch trials while bottom panels are for all same-pitch trials. Boxes denote the 25th − 75th percentile range, the horizontal bar in the center denotes the median, and error bars span from maximum to minimum. Left panels: leading melody N1s. Right panels: lagging melody N1s. Box colors represent attentional state (attend-leading: red; attend-lagging: blue; passive: black). For different-pitch stimuli, for both leading- and lagging-melody N1s, ERPs are usually larger at the group level when listeners attended to that melody vs. either passive trials or when attending to the opposite melody. C. N1 amplitudes averaged across notes within each melody. Top panel: different-pitch stimuli (from correctly answered trials). Bottom panel: same-pitch stimuli (from all trials). Three left boxes: leading melody N1s. Three right boxes: lagging melody N1s. For different-pitch stimuli, N1 amplitudes to a given melody are larger when that melody is attended.

significantly correlated with the *lag amplification*, suggesting FHRs were independent of the strength of top−down amplification of the auditory cortical representation (see Fig. 7B). When the FHRs and *lag amplification* were combined as independent regressors, the correlation between predicted performance and observed performance in the same-pitch trials improved compared to when using lag amplification alone (compare Fig. 7C to bottom panels of Fig. 5B).

The significant contributions of both *lag amplification* and FHRs for predicting performance was reflected in a 2-way ANOVA with both included as regressors. Both factors contributed significantly to predictions of performance (when predicting attend-leading same-pitch performance, $F_{1,14} = 10.19$, $P = 0.0065$ for lag amplification and $F_{1,14} = 10.21$, $P = 0.0065$ for FHRs; when predicting attend-lagging same-pitch performance, $F_{1,14} = 15.11$, $P = 0.0016$ for lag amplification and $F_{1,14} = 6.23$, $P = 0.026$ for FHRs). These results
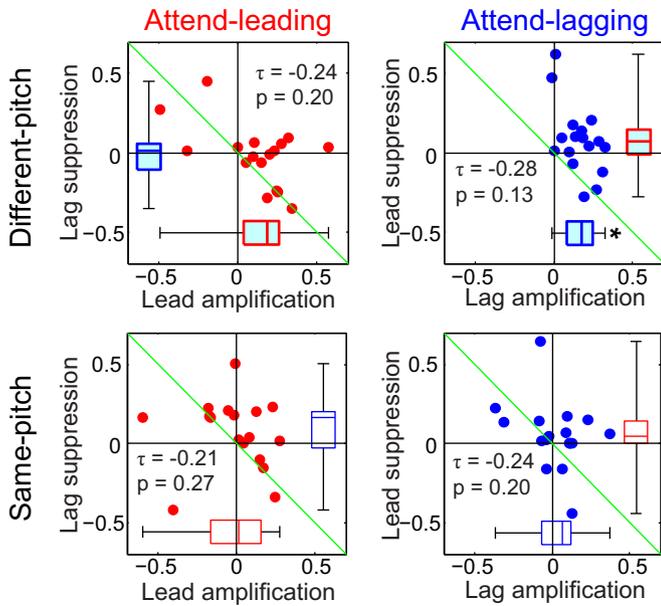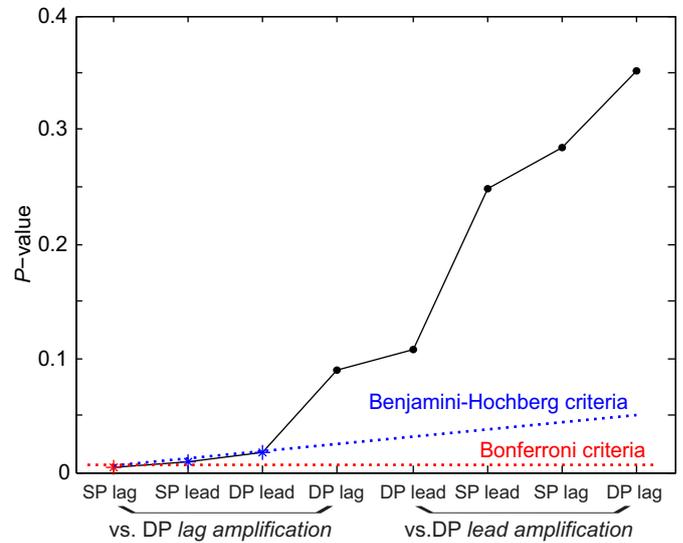
Fig. 4. N1 amplification and suppression of individual listeners. Target-melody-N1 amplification and interferer-melody-N1 suppression for attend-leading (left panel) and attend-lagging (right panel) conditions for different-pitch (top panels) and same-pitch stimuli (bottom panels). Each dot represents a single subject. Boxes show the across-subject distributions, giving the 25th – 75th percentile range, with the horizontal bar in the center denoting the median, and the error bars spanning from maximum to minimum. Only different-pitch lagging-melody amplification has a median value that is statistically different from zero (positive). There is no significant correlation between amplification and suppression.

Fig. 6. P-values from the eight correlation analyses in the increasing order. The blue dashed line represents the Benjamini-Hochberg criteria, while the red dashed line represents the Bonferroni criteria. The P-values obtained from comparisons between lag amplification and the same-pitch performance, and between lag amplification and the different-pitch, attend-leading performance satisfy Benjamini-Hochberg criteria. Only the P-value obtained from the comparison between lag amplification and the same-pitch, attend-lagging performance satisfies Bonferroni criteria ($P < 0.00625$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

suggest that individual differences in attention depend both on differences in cortical gain control and other independent cognitive factors.

4. Discussion

We designed a simple melody identification task that requires listeners to focus and sustain selective attention, mimicking key aspects of the demands made on listeners in everyday social settings. However, we designed our task to have very low memory load in order to emphasize differences in the ability to direct selective attention, which is controlled by executive control networks in the cortex (Giesbrecht et al., 2003; Knudsen, 2007; Hill and Miller, 2010; Shamma and Micheyl, 2010).

We intermingled two different stimulus types and looked for within-subject consistency across the trials: easy, "different-pitch" stimuli in which the competing melodies were separated in pitch and hard, "same-pitch" stimuli in which the pitches of the competing melodies overlapped. Differences in top–down modulation of ERPs were robustly quantified from the easy trials, where
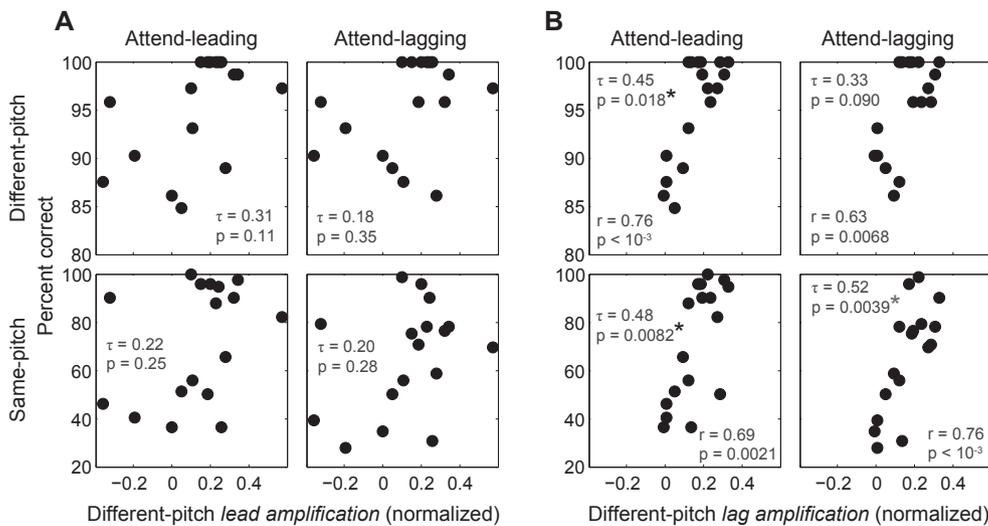
Fig. 5. Behavioral performance is related to lag amplification. A. Relationship between performance in the four conditions and the lead amplification in the different-pitch, attend-lagging condition. There is no statistically significant relationship in any correlation. B. Relationship between performance in the four conditions and the lag amplification in the different-pitch, attend-lagging condition. The correlation is significant for all types of trials except the different-pitch attend-lagging condition.
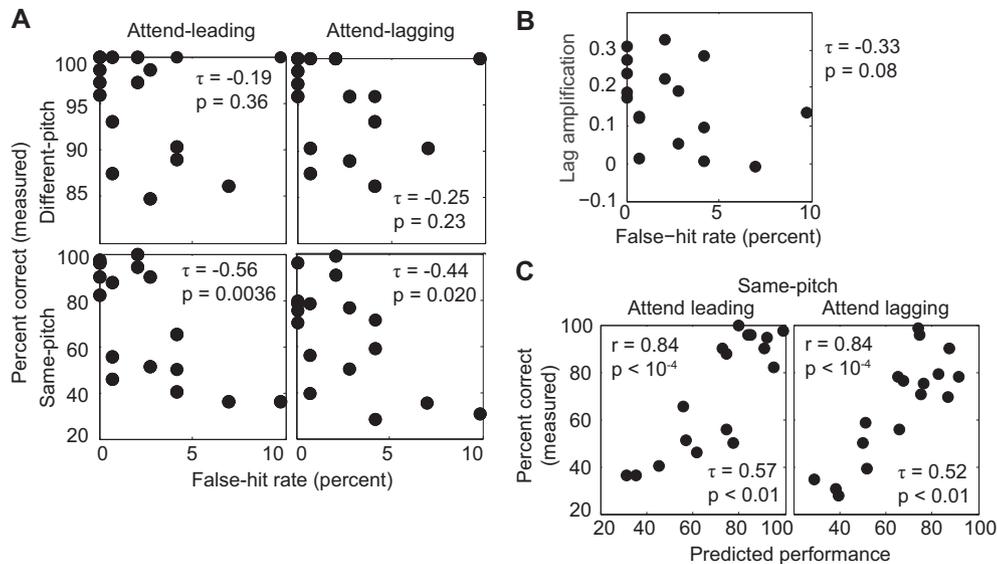
Fig. 7. False-hit rates are related to performance in some conditions. A. Relationship between performance in the four conditions and the false-hit rate (FHR). Performance is significantly correlated with FHR for same-pitch stimuli (bottom panels). B. Amplification of the N1 responses to the lagging melody in the different-pitch, attend-lagging condition plotted against the FHR in the passive condition. There is no statistically significant correlation. C. Relationship between performance and the linear prediction of performance using regressors of lag amplification and the FHRs. The FHR accounts for additional inter-subject variation that is not captured by the neural amplification (compare to bottom panels of Fig. 5B).

all listeners successfully performed the task. However, when the task was easy, all subjects performed well, limiting variability in performance and the power of the analysis. Conversely, when the task was hard, although inter-subject performance differences should be large, it was hard to gather a sufficient number of trials in which listeners successfully deployed selective auditory attention to analyze the ERPs for correctly answered trials alone. When analyzing *all* trials, neural modulation of the auditory responses was weak.

We found that behavioral performance is related to the strength of modulation of cortical ERPs in the easy condition; specifically, subjects who showed the largest amplification of the N1 responses to the lagging melody when it was attended also performed best in the difficult attentional tasks. While lag amplification strength is related to behavioral ability, lead amplification is not. This may reflect the fact that the leading melody drew attention exogenously; in contrast, focusing on the lagging melody required top–down, endogenous control. For this reason, we expected ERPs measured in the different-pitch, attend-lagging condition to be the best measure of individual differences in how attention shaped cortical responses. Thus, our finding that lag amplification predicts behavioral performance suggests that some of the individual variation in the ability to direct selective attention is related to differences in top–down cortical attentional processes.

For the different-pitch stimuli, we found large inter-subject differences in the pattern of attentional modulation of ERPs. Despite large inter-subject differences, for almost all subjects in both conditions, selective attention lead to an increase in the *relative* strength of the representation of the attended melody compared to that of the competing melody. Thus, consistent with past reports, at the group level, selective auditory attention operates as a form of selective gain control (Hillyard et al., 1973; Chait et al., 2010; Kerlin et al., 2010; Power et al., 2011; Choi et al., 2013). Although only the lag amplification was significantly greater than zero at the group level, the strengths of amplification and suppression were uncorrelated across subjects and varied significantly across subjects. These results suggest that listeners

differ in the efficacy of top–down gain control, as evidenced by differences in attentional modulation of cortical responses.

We only had enough correctly answered trials to quantify neural modulation for the easier, different-pitch stimuli. The strength of the lag amplification in the different-pitch trials was positively correlated with how effective listeners were at performing the selective attention task for both types of trials. The other metrics of attentional modulation (e.g. strength of lead suppression in the different-pitch trials) was not correlated with performance in any condition; however, this may simply be because listeners actively suppress all melodies in the passive condition (the reference for the amount of suppression in the attend conditions). Thus, only the neural modulation in the condition that demanded a volitional redirecting of attention (lag amplification in the attend-lagging condition) was correlated with performance across conditions. In other words, the listeners who most strongly amplified the neural representation of the target in the condition that required top–down control (where individual differences in cortical modulation could be measured) were consistently the best subjects across all conditions; however, this relationship is clearer in the challenging behavioral conditions where individual differences are most pronounced.

For the hard to separate, same-pitch stimuli, neural modulation was not robustly observed. In the same-pitch conditions, the mean N1 amplitude was small and inter-subject variance was large. The small amplitudes of the N1s may arise because the melodies overlapped in their pitch ranges, which could make the note onsets less clear (Martin et al., 1999; Martin and Stapells, 2005). We also found no significant attentional modulation of ERPs in the same-pitch conditions; however, we had to include responses from incorrectly answered trials to even calculate clear ERPs. Including results from incorrect trials likely leads to an underestimation of the strength of attentional modulation on successful trials. From the current results, we cannot conclude that attention does not alter the ERPs in the same-pitch trials; the sensitivity of our measures of modulation of N1 depends on many factors, including the quality of the EEG data.

Our behavioral results are consistent with other recent studies in showing that listeners with clinically normal hearing vary widely in how well they can deploy selective attention (Ruggles and Shinn-Cunningham, 2011; Anderson and Parbery-Clark, 2013). Many listeners who show no obvious hearing deficits seek treatment from audiologists because of difficulties in settings where selective attention is important (Moore et al., 2013). Such listeners are typically said to have "central auditory processing disorder," a weak diagnosis that is not linked to any specific mechanism or cause (Jerger and Musiek, 2000). Whereas differences in the fidelity of subcortical temporal coding contribute to differences in attentional task performance (Ruggles et al., 2011, 2012; Anderson and Parbery-Clark, 2013), the current results suggest that differences in cortical control also affect the ability to solve the cocktail party problem.

The claim that cortical amplification predicts listener ability bears further confirmation. It is possible that poor subcortical encoding induced poor cortical attentional modulation and behavioral performance at the same time. Future experiments are needed to explore contributions of subcortical-encoding acuity on the cortical attentional modulation and behavioral attentional ability. Also, although we designed our study to minimize effects of other cognitive factors such as memory load, etc., it remains unclear how much individual differences in cognitive status (e.g. drowsiness, motivation, etc.) contributed to the individual differences we found. We used the FHR (how often subjects inappropriately responded to passive trials) as a measure of subjects' overall alertness, and this helped explain performance differences not correlated with attentional modulation. However, there is likely a better and more direct measure for isolating the key factors contributing to inter-subject variability.

Interestingly, in the same-pitch trials, some listeners reported that they relied on temporal expectation to focus on the target melody because the task was too hard when depending on spatial cues alone. It may be that selective entrainment (Lakatos et al., 2008, 2013; Cravo et al., 2013) to the temporal structure of the attended melody causes amplification, which is very effective in the difficult, same-pitch trials. To tease apart whether different forms of selective attention give rise to different patterns of amplification, further experiments are needed. Understanding how differences in executive function contribute to the ability to solve the cocktail party could have a significant impact on audiological practice. Individual differences in the effectiveness of top–down control may also help explain differences in how well users can operate BCIs, whether based on auditory or visual modalities [e.g. see (Allison and Neuper, 2010; Hill and Schölkopf, 2012)].

5. Conclusions

Listeners vary in how well they can perform selective attention tasks. These performance differences are greatest in the most challenging stimulus conditions, such as when it is difficult to segregate a target sound stream from competing streams. At the group level, ERPs corresponding to a particular stream are enhanced when listeners are attending to that stream. This effect was most robust when the less salient lagging melody was attended; however, the strength of this attentional amplification varied across listeners. Importantly, the efficacy of top–down modulation when listeners attended the lagging stimulus correlated with subject performance in most conditions, suggesting that there are consistent subject-specific differences in how effectively listeners can focus selective auditory attention. An independent source of subject variability was seen in the ability of listeners to refrain from making false-alarm responses on control trials.

These results suggest that individual differences in selective attention ability are directly related to differences in the efficacy of executive cortical processes. Together with recent results showing that differences in the fidelity of neural coding of supra-threshold sound contribute to differences in selective attention ability, these results begin to shed light on how individual variation in both top–down and bottom–up processing affects the ability to solve the cocktail party problem.

References

Allison, B., Neuper, C., 2010. Could anyone use a BCI?. In: Brain-computer Interfaces. Springer, London, pp. 35–54.

Anderson, S., Parbery-Clark, A., 2013. Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. J. Speech Lang. Hear Res. 56, 31–44.

Atiani, S., Elhilali, M., David, S.V., Fritz, J.B., Shamma, S.A., 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron 61, 467–480.

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

Best, V., Ozmeral, E.J., Kopco, N., Shinn-Cunningham, B.G., 2008. Object continuity enhances selective auditory attention. Proc. Natl. Acad. Sci. U S A 105, 13174–13178.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.

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

Choi, I., Rajaram, S., Varghese, L.A., Shinn-Cunningham, B.G., 2013. Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. Front. Hum. Neurosci. 7, 1–19.

Cravo, A., Rohenkohl, G., Wyart, V., Nobre, A.C., 2013. Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. J. Neurosci. 33, 4002–4010.

Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. U S A 109, 11854–11859.

Elhilali, M., Xiang, J., Shamma, S.A., Simon, J.Z., 2009. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol. 7, e1000129.

Giesbrecht, B., Woldorff, M., Song, A., Mangun, G., 2003. Neural mechanisms of top-down control during spatial and feature attention. Neuroimage 19, 496–512.

Hill, K.T., Miller, L.M., 2010. Auditory attentional control and selection during cocktail party listening. Cereb. Cortex 20, 583–590.

Hill, N.J., Schölkopf, B., 2012. An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. J. Neural Eng. 9, 026011.

Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science 80 (182), 177–180.

Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. Lond B Biol. Sci. 353, 1257–1270.

Jerger, J., Musiek, F., 2000. Report of the consensus conference on the diagnosis of auditory processing disorders in school-aged children. J. Am. Acad. Audiol. 11, 467–474.

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

Knudsen, E.I., 2007. Fundamental components of attention. Annu Rev. Neurosci. 30, 57–78.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 80 (320), 110–113.

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

Linkenkaer-Hansen, K., Nikulin, V.V., Palva, S., Ilmoniemi, R.J., Palva, J.M., 2004. Prestimulus oscillations enhance psychophysical performance in humans. J. Neurosci. 24, 10186–10190.

Martin, B.A., Stapells, D.R., 2005. Effects of low-pass noise masking on auditory event-related potentials to speech. Ear Hear 26, 195–213.

Martin, B., Kurtzberg, D., Stapells, D., 1999. The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. J Speech. Lang. Hear Res. 42, 271–286.

Martınez-Trujillo, J., Treue, S., 2002. Attentional modulation strength in cortical area MT depends on stimulus contrast. Neuron 35, 365–370.

Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233–236.

Moore, D.R., Rosen, S., Bamiou, D.-E., Campbell, N.G., Sirimanna, T., 2013. Evolving concepts of developmental auditory processing disorder (APD): a British Society of Audiology APD special interest group "white paper". Int. J. Audiol. 52, 3–13.

Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials. II. Effects of attention. Electroencephalogr. Clin. Neurophysiol. 36, 191–199.

Power, A.J., Lalor, E.C., Reilly, R.B., 2011. Endogenous auditory spatial attention modulates obligatory sensory activity in auditory cortex. Cereb. Cortex 21, 1223–1230.

Ress, D., Backus, B.T., Heeger, D.J., 2000. Activity in primary visual cortex predicts performance in a visual detection task. Nat. Neurosci. 3, 940–945.

Ruggles, D., Bharadwaj, H., Shinn-Cunningham, B.G., 2011. Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. Proc. Natl. Acad. Sci. U S A 108, 15516–15521.

Ruggles, D., Bharadwaj, H., Shinn-Cunningham, B.G., 2012. Why middle-aged listeners have trouble hearing in everyday settings. Curr. Biol. 22, 1417–1422.

Ruggles, D., Shinn-Cunningham, B., 2011. Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. J. Assoc. Res. Otolaryngol. 12, 395–405.

Shamma, S.A., Micheyl, C., 2010. Behind the scenes of auditory perception. Curr. Opin. Neurobiol. 20, 361–366.

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

Xiang, J., Simon, J., Elhilali, M., 2010. Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. J. Neurosci. 30, 12084–12093.

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, 980–991.