Neuron
Article

The Spectrotemporal Filter Mechanism of Auditory Selective Attention

Peter Lakatos,[1,2,*] Gabriella Musacchia,[3] Monica N. O'Connel,[1] Arnaud Y. Falchier,[1] Daniel C. Javitt,[1,2] and Charles E. Schroeder[1,2]
[1]Cognitive Neuroscience and Schizophrenia Program, Nathan Kline Institute, Orangeburg, NY 10962, USA
[2]Department of Psychiatry, Columbia College of Physicians and Surgeons, New York, NY 10032, USA
[3]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ 07102, USA
*Correspondence: plakatos@nki.rfmh.org
http://dx.doi.org/10.1016/j.neuron.2012.11.034

SUMMARY

Although we have convincing evidence that attention to auditory stimuli modulates neuronal responses at or before the level of primary auditory cortex (A1), the underlying physiological mechanisms are unknown. We found that attending to rhythmic auditory streams resulted in the entrainment of ongoing oscillatory activity reflecting rhythmic excitability fluctuations in A1. Strikingly, although the rhythm of the entrained oscillations in A1 neuronal ensembles reflected the temporal structure of the attended stream, the phase depended on the attended frequency content. Counter-phase entrainment across differently tuned A1 regions resulted in both the amplification and sharpening of responses at attended time points, in essence acting as a spectrotemporal filter mechanism. Our data suggest that selective attention generates a dynamically evolving model of attended auditory stimulus streams in the form of modulatory subthreshold oscillations across tonotopically organized neuronal ensembles in A1 that enhances the representation of attended stimuli.

INTRODUCTION

Selective attention represents a fundamental cognitive capacity that allows the brain to enhance its internal representation of task-relevant events at the expense of irrelevant ones (Broadbent, 1958; Treisman, 1969; Desimone and Duncan, 1995). Even at its initial stage in primary auditory cortex (A1), auditory processing appears to be modulated by attention to specific features of auditory stimuli, such as frequency and time (Woldorff et al., 1993; Fritz et al., 2003, 2005; Bidet-Caulet et al., 2007; Atiani et al., 2009; Lakatos et al., 2009; Elhilali et al., 2009b; Jaramillo and Zador, 2011). Although the effects of selective attention on auditory responses are widely recognized, the specific neurophysiological mechanisms by which the brain is able to select task-relevant items along the fundamental organizing dimensions of auditory objects are not known. The main goal of our study was to investigate the physiological mechanisms underlying auditory selective attention in monkeys.

Recent studies have shown that when sensory inputs related to attended stimuli drive the cortex rhythmically, neuronal ensemble excitability fluctuates in a pattern that matches the temporal structure of these inputs, as reflected by a neuronal oscillation tied to the timing of attended events (Lakatos et al., 2008; Bosman et al., 2009; Luo et al., 2010; Mathewson et al., 2010; Saleh et al., 2010; Stefanics et al., 2010; Besle et al., 2011). This serves dual mechanistic purposes: sensory responses to attended stimuli are predictively amplified because the high excitability phase of the oscillation is aligned to the rhythmic inputs. At the same time responses to temporally offset stimuli are attenuated by the low excitability phase of the oscillation, in essence working as a temporal filter (Broadbent, 1958; Treisman, 1969; Large and Jones, 1999; Jones et al., 2002). However, this mechanism alone would not be efficient when relevant and irrelevant events considerably overlap temporally, as is often the case in a natural auditory environment (Kerlin et al., 2010; Ding and Simon, 2012; Mesgarani and Chang, 2012). It has been proposed that when multiple auditory streams are present, the dominant frequency content of the attended stream guides "temporal attention" and leads to the segregation of this stream from "background," ignored auditory stimuli (Shamma et al., 2011). Thus ideally, the mechanism of auditory selective attention should form a spectrotemporal filter, incorporating information about both the timing (rhythm) and the frequency content of the relevant auditory stimulus stream, in order to enhance the sensory representation of attended stimuli along these two feature dimensions. Although entrained neuronal oscillations likely form the temporal component of the filter, the neurophysiological mechanism that implements the spectral component of the filter is not yet clear. A recent study provided a potential key to this puzzle by showing that the sign of stimulus-related phase reset in A1 can be frequency-specific (O'Connell et al., 2011). Pure tones whose frequency corresponds to the preferentially processed frequency (best frequency [BF]) of a given A1 site reset local oscillatory activity to its high excitability phase, whereas tones whose frequency differs by about two octaves (non-BF tones) reset ongoing oscillations to their opposite, low excitability phases, enhancing the effects of sideband inhibition in A1.

As mentioned above, previous studies suggest that attention to a rhythmic event stream results in response gain due to periodic increase in excitability in anticipation of attended stimuli (Jones et al., 2002; Lakatos et al., 2008; Mathewson et al., 2010;

Stefanics et al., 2010). Thus it is possible that attention to rhythmic tone streams results in a synchronous entrainment of neuronal activity across differently tuned A1 neuronal ensembles and a general enhancement of excitability at times when attended stimuli are predicted to occur. A second possibility is that because phase reset is under strong attentional control (Lakatos et al., 2009), entrainment would only occur in A1 regions processing attended frequency content. Alternatively, phase reset could retain its frequency specificity observed in nonbehaving monkeys (O'Connell et al., 2011), and could entrain low frequency oscillations to opposing excitability phases in A1 regions processing attended versus ignored frequency content. The main goal of our present study was to answer these open questions by establishing the mechanism of auditory selective attention to frequency and rhythm (time) in primary auditory cortex. We hypothesize that when attention is directed toward one of two competing auditory streams, excitability across neuronal ensembles in A1 is modulated to enhance the representation of attended spectral content at predictable times. This basic two-dimensional spectrotemporal modulation of excitability can be extended to provide a mechanistic account for the brain's ability to preferentially represent more complex event streams such as that of the attended speaker in a cocktail party conversation (Kerlin et al., 2010; Ding and Simon, 2012; Mesgarani and Chang, 2012).

To test our hypothesis, we recorded laminar neuronal ensemble activity profiles from area A1 in three monkeys during the performance of an auditory selective attention task. We presented either two rhythmic auditory tone streams concurrently that differed in their rhythm and spectral content, or one of these streams in isolation. Every stream contained repetitive tones of one frequency, with occasional frequency deviants. The monkey was required to respond to the deviants in the single stream, or to the deviants in the cued stream in the dual stream (selective attention) condition. In the latter case, we found that the monkeys exclusively responded to deviants in one of the rhythmic streams, and never to deviants in both streams, confirming that just like humans, monkeys are able to segregate rhythmic auditory streams that differ in their rhythm and frequency content (Izumi, 2002). Analysis of the neuronal ensemble activity revealed the entrainment of ongoing subthreshold neuronal oscillations to the temporal structure of the attended stream, with opposing phases across differently tuned A1 neuronal ensembles. We found that these oscillations simultaneously sharpened and stabilized responses to attended stimuli, thereby increasing their contrast compared to stimuli in the background, ignored stream.

RESULTS

We recorded laminar profiles of field potentials and multiunit activity (MUA) from 32 A1 sites in three macaque monkeys using linear array multielectrodes. The sites were distributed evenly along the tonotopic axis of A1, with BFs ranging from 0.5 kHz to 32 kHz. To minimize the effects of volume conduction and more precisely define local synaptic current flow, we calculated one-dimensional current source density (CSD) profiles (Freeman and Nicholson, 1975) and carried out our analyses on the CSD waveforms and concomitant MUA.

The monkeys were trained to perform a classic auditory oddball task, detecting frequency "deviants" embedded in a rhythmic stream of "standard" pure tones with constant frequencies and stimulus onset asynchronies (SOAs). Frequency deviants (2–4 semitone difference from the standards) occurred randomly at 3–9 s intervals; when occurring in a cued (attended) stream, these were targets. The selective attention condition entailed the presentation of two concurrent streams differing in SOA (624.5 or 562.05 ms; rates of ~1.6 and 1.8 Hz) and tone frequency. Prior to dual stream presentation, the monkey was cued by presenting one stream alone. Eighty percent of the targets were paired with a juice reward, whereas 20% were not. In order to get the reward, monkeys had to stick out their tongues (lick), because the spout of the juicer was positioned away from the monkey's mouth. Licking on deviants both with and without paired reward was used to monitor performance (see Experimental Procedures).

In each experiment (multielectrode penetration), the frequency of standard tones was set to one of two values in different blocks: either matching the BF of the recording site (BF stream), or to a frequency value two octaves higher or lower than the BF of the recording site (non-BF stream). Previous studies indicate that frequency separations of two octaves or more promote stream segregation even at longer SOAs, like those used here (Van Noorden, 1975; Hartmann and Johnson, 1991). In five experiments, we used dual multielectrode recordings to collect data from two ipsilateral A1 sites simultaneously (Figure 1), allowing concurrent examination of BF stream and non-BF stream effects. In these experiments, we positioned the linear array multielectrodes 2 mm apart along the tonotopic axis of A1, because the tonotopic gradient in macaque A1 is ~1.0 mm/octave (Merzenich and Brugge, 1973; Kosaki et al., 1997). Therefore, setting the frequency of standard tones in our two stimulus streams such that they corresponded to the BF of the recording sites resulted in an approximately two octave difference.

Rhythmic, Frequency-Dependent Modulation of Baseline Excitability

To determine whether the previously observed fluctuation of excitability in response to rhythmic stimulus streams in A1 (Lakatos et al., 2005b; O'Connell et al., 2011) occurs in neuronal ensembles tuned to nonattended frequency content, and if it does, does it differ from predictive excitability modulation in neuronal ensembles processing attended frequency content, we first compared the neuronal activity related to non-BF and BF auditory streams presented alone (cueing trial blocks). Figure 1 displays data from a representative experiment where the laminar neuronal ensemble activity was recorded simultaneously in two A1 regions situated 2 mm apart on the superior temporal plane. We used streams of 5.7 and 16 kHz tones as stimuli to match the tuning of the recording sites. Upper color maps show laminar CSD responses related to attended 5.7 kHz stimuli, whereas lower color maps show laminar responses to attended 16 kHz stimuli. As expected, when the frequency of stimuli matched the BF of a given A1 region (BF streams, upper left and lower right CSD profiles in Figure 1), baseline excitability fluctuated such that just prior to attended stimuli, a source (blue in the CSD maps) over sink (red) pattern

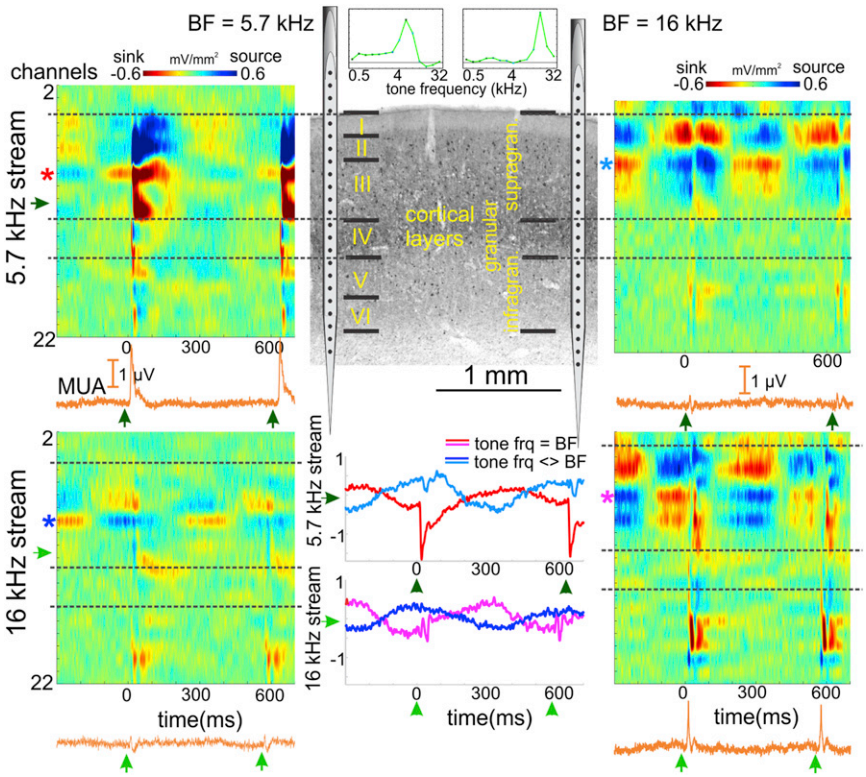

Figure 1. Laminar CSD Profiles and MUA in Response to Attended Rhythmic Auditory Stimulus Streams Recorded Concurrently in Two A1 Regions

All data shown in the figure relates to attended streams in the single stream (cueing) condition. As the tuning curves (green traces in the middle) based on MUA response amplitudes to a range of different frequency pure tones show, the BFs of the two simultaneously recorded sites were 5.7 and 16 kHz, thus the monkeys were required to discriminate deviants among streams of 5.7 and 16 kHz standard tones. Responses related to the 5.7 kHz stream are shown on top, whereas responses to the 16 kHz stream are on the bottom. Laminar boundaries (dotted black lines) were determined based on functional criteria. Color maps display averaged laminar CSD profiles showing responses to attended BF (top left for the 5.7 kHz site and bottom right for the 16 kHz site) and non-BF stimulus streams. Orange traces below are concomitantly recorded MUA averaged across all cortical layers. Overlaid traces in the middle show concurrently recorded supragranular CSD activity of the two A1 sites in response to 5.7 kHz (upper) and 16 kHz (lower) auditory tone streams at the lower supragranular laminar locations marked by the colored stars (the upper CSD component in the supragranular layers represents passive or return current, see Figure S1). Although red and magenta traces are responses of the two regions when their BF corresponded to the attended tone frequency (BF streams), dark and light blue traces show responses to attended streams in cases when tone frequency did not match the BF of the recording site. Note the rhythmic opposite sign baseline fluctuation in the laminar MUA (orange traces) as well, especially in the 16 kHz site on the right. See also Figure S1.

characterized the CSD activity of the supragranular layers, indicating a high excitability phase of rhythmic neuronal ensemble activity (see Figure S1 available online; Lakatos et al., 2005b, 2008). In other words if the preferentially processed frequency of a given A1 neuronal ensemble is attended, excitability is predictably increased just prior to the time when attended stimuli are predicted to occur. In sharp contrast, when the monkey attended to streams of stimuli whose frequency did not match the BF of the recording site (non-BF streams, lower left and upper right CSD profiles in Figure 1), baseline fluctuation is present, but opposite in sign. Overlay of the CSD waveforms (Figure 1, middle) from simultaneously recorded corresponding supragranular sites (marked by colored asterisks on the left margin of each profile) clearly illustrates this phase opposition. In keeping with the notion that the opposed phases represent high and low excitability states of the local neuronal ensemble, we found that local neuronal firing (orange traces in Figure 1) related to attended BF versus non-BF streams also fluctuates rhythmically but in opposite directions. In the immediate prestimulus period (−100 to 0 ms), MUA significantly increased in 24/32 experiments when attending to BF streams, and decreased in 29/32 experiments when attending to non-BF streams (Wilcoxon rank sum, $p < 0.01$) compared to MUA in-between stimuli (−400 to −200 ms).

Underscoring the systematic nature of these baseline excitability fluctuations for attended streams, Figure 2A shows that δ band intertrial coherence (ITC), quantifying phase-similarity across trials, was significant in all experiments regardless of the relation of attended stimulus frequency to the BF of a site (Rayleigh's test, $p < 0.01$). Thus, non-BF stimulus streams are just as effective in rhythmically modulating excitability as BF streams, and attending to either results in a constant excitability phase at response onset across stimulus presentations. Critically, however, mean phase distributions near stimulus onset are frequency-specific (Figure 2B): in the case of BF streams, mean phases pooled around the negative peak of the δ frequency baseline fluctuation indicating a high excitability phase (ϕ mean = 2.33 radians, ϕ deviants = 0.62), whereas in the case of non-BF streams, the rhythmic baseline fluctuation at stimulus onset was around its opposite, low excitability phase (ϕ mean = −0.74 radians, ϕ deviants = 0.64). The distribution of mean δ phases was significantly different in the two cases (Fisher's nonparametric test for the equality of circular means, $p < 0.01$).

Entrainment of Ongoing Oscillatory Activity versus Evoked Type Responses

In theory, two distinct mechanisms could generate the rhythmic excitability fluctuations we observe: modulation of the frequency and phase of ongoing neuronal oscillations that aligns them to the temporal structure of an attended stream (entrainment), or generation of "evoked type" activity at the input rate. Because modulatory (phase reset) and evoked type responses

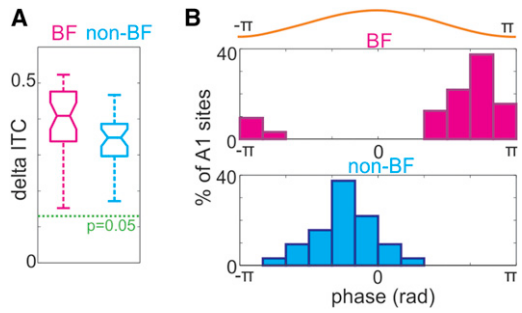

Figure 2. Pooled δ ITC and Mean δ Phase Related to Attended BF and Non-BF Stimulus Streams

(A) Boxplots show supragranular δ ITC (at the δ frequency that corresponded to the repetition rate in auditory streams) at the time of stimulus presentation across all experiments (n = 32). Green line denotes the value above which ITC is significant, calculated using Rayleigh's test (p = 0.01).

(B) Histograms show the distribution of mean supragranular δ phases at the time of stimulus presentation in response to BF and non-BF streams across all experiments. The orange trace on the top displays an oscillatory cycle for reference. Mean phases are pooled around the negative peak in response to BF streams, whereas they are clustered around the opposite phase, the positive peak in response to non-BF stimulus streams.

(lemniscal feed-forward excitation and inhibition) can each result in opposite sign responses to BF versus non-BF pure tones (Wehr and Zador, 2003; Zhang et al., 2003; Tan et al., 2004; O'Connell et al., 2011), both of these mechanisms are plausible.

If entrainment is the mechanism of the rhythmic baseline fluctuation observed in averaged waveforms (Figure 1), it should have a laminar profile similar to that of ongoing oscillatory activity, because by definition, entrainment only modulates the frequency and phase of ongoing oscillations. As Figure 1 illustrates, rhythmic δ frequency baseline fluctuation related to attended stimulus streams was largest in amplitude in the supragranular layers, which corresponds to the site of maximal amplitude ongoing δ oscillatory activity (Lakatos et al., 2005a). Although this qualitative feature of the data does not unequivocally argue for entrainment as opposed to an evoked mechanism, the two findings presented below do.

First, the amplitude of δ oscillatory activity did not change significantly during stimulus presentation: the amplitude of supragranular δ band (0.75–2.5 Hz) oscillatory activity before (mean = 0.49 mV/mm^2, SD = 0.26), during (BF streams: mean = 0.52 mV/mm^2, SD = 0.28, non-BF streams: mean = 0.48 mV/mm^2, SD = 0.25) and after (mean = 0.47 mV/mm^2, SD = 0.22) auditory stream presentation was not significantly different across recordings (Kruskal-Wallis test, $p > 0.01$). A paired t test comparing prestimulus δ amplitude with that measured during stimulation also did not reveal any significant difference ($p > 0.01$). It is important to note that the unchanged amplitude of ongoing activity during entrainment is the reason we refer to entrained oscillatory activity as subthreshold, however, both ongoing and entrained oscillations are only subthreshold on the neuronal ensemble level. From the point of view of a single neuron, due to an increased probability of spike generation in the depolarized state, the membrane potential fluctuations reflected in ongoing and entrained neuronal oscillations can and do result in incidental firing as illustrated by phase-related changes in the MUA (Figure S1D). However, this firing is sparse and under physiological conditions does not result in an activation of higher level targets due to the lack of synchrony, as illustrated by the fact that normally we do not perceive ongoing oscillatory activity.

The most compelling argument for entrainment is that its modulatory effect on ongoing oscillations outlasts rhythmic stimulation. Figure 3A depicts long epochs of supragranular CSD activity covering a period of auditory stimulation and extending 5 s after its end. The single trial CSD segments were averaged across all 32 experiments for attended BF stimuli presented at a 1.6 Hz (upper red) and for attended non-BF stimuli that were presented at a 1.8 Hz rate (lower blue). In the upper trace, along with suprathreshold-evoked responses to BF auditory stimuli (blue drop lines) a rhythmic baseline fluctuation is present. It also appears that after the last stimulus (time > 0), the oscillation continues: negative peaks coincide with time points when stimuli would occur if the stimulation had continued (red drop lines). Histograms show that at these time points, δ oscillatory phase is nonrandom across trials (experiments): phases are still pooled around π corresponding to the negative peak (the high excitability phase of δ oscillatory activity in supragranular sites). In the lower trace, an apparent oscillation is aligned with its positive peak to the timing of non-BF stimuli which again continues after stimulus presentation ends. This is confirmed by the significantly biased δ phase distribution that outlasts stimulation by several cycles. As a consequence, the amplitude spectra of ongoing δ activity in the 2–4 s time interval following stimulation (Figure 3B) show prominent δ peaks that correspond to the rhythm of slower (1.6 Hz) and faster (1.8 Hz) stimulus streams. Figure 3B and statistical analyses (see above) show that prestream δ oscillatory activity has the same overall amplitude as δ activity following the stimulus stream, but it is more "spread out" over different frequencies; poststream δ energy is still concentrated at the frequency that corresponds to the repetition rate of the attended stimulus stream. The finding that the frequency (and phase) of ongoing oscillatory activity reflects the temporal structure (and frequency content) of the attended rhythmic stream indicates that the structure of ongoing oscillatory activity was modulated by entrainment, so that excitability fluctuations become nonrandom in relation to the timing of attended events.

To determine whether oscillatory entrainment to attended rhythmic stimulus streams occurs outside the δ frequency range, we presented blocks of 1 min long stimulus trains with presentation rates varying from 0.8 to 12 Hz in a subset of experiments (n = 10). Similar to other paradigms, the monkeys had to detect rarely occurring frequency deviants in a stream of standard tones. The data indicate that at most δ and θ range repetition rates (with the exception of 3.2 Hz) the baseline fluctuates in opposite phase when attended stimulus frequency matches the BF of the recordings site (magenta traces and histograms in Figure 4) versus when it does not (cyan in Figure 4), and that phase opposition is strongest specifically at repetition rates that correspond to the dominant frequency of δ and θ band ongoing oscillatory activity (Figures 4B and 4C). We also replicate the above described finding that the temporal structure of

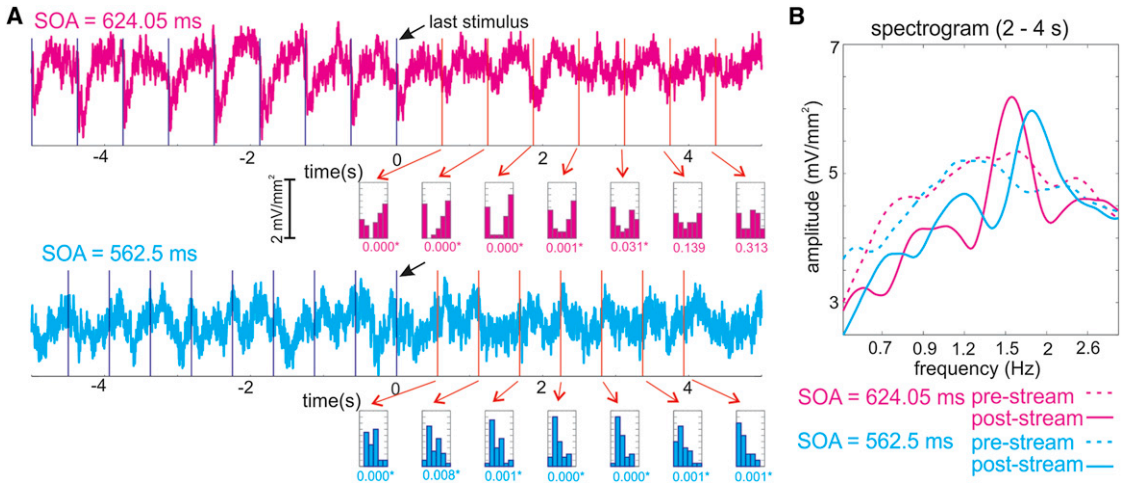

Figure 3. Responses to BF and Non-BF Stimulus Streams Centered on the Last Stimulus of Attended Rhythmic Auditory Streams and Averaged Across Experiments

(A) The upper magenta trace shows s 10 s long supragranular CSD segment centered on the last stimulus of BF tone streams with a 624.5 ms SOA, averaged across all experiments (n = 32). Stimuli are represented by blue lines, whereas red lines denote time points at which stimuli would have occurred if the stream would have continued. Histograms show the distribution of δ phases at these time points across all experiments. Below the histograms, the p value calculated using Rayleigh's test is displayed for each of the phase distributions. The lower blue trace shows the averaged supragranular CSD segment centered on the last stimulus of non-BF stimulus streams with a 562.05 ms SOA. Note that phase distributions at times when stimuli would have occurred after the stimulus streams ended are pooled around opposite phases than following BF stimulus streams. Nonetheless, in both cases δ phase distribution remains significantly biased for several δ cycles after auditory stimulus streams end, which is why the δ oscillation is visible in the averaged waveforms.

(B) Spectrograms of prestream (4–2 s prior to auditory streams) and poststream (2–4 s after auditory streams end) ongoing δ band oscillatory activity for stimulus streams presented with 624.5 ms (magenta traces) and 562.05 ms (blue traces) SOAs. The peaks in the poststimulus spectrograms at 1.6 and 1.8 Hz correspond to the slower and faster presentation rates.

ongoing oscillatory activity can be modulated to match attended stimulus structure (context) in these frequency bands, as evidenced by spectral peaks corresponding to the repetition rate of attended stimulus streams in the spectrum of ongoing activity following stimulus presentation (Figure 4D). Taken together these results indicate that both δ and θ frequency oscillations can entrain to rhythmic attended auditory stimuli, and that the phase of entrainment in these frequency ranges depends on attended frequency content. It will be interesting to examine the rules of entrainment in the case of more realistic, e.g., nested rhythmic stimulus structures, like in speech, in which case δ and θ entrainment might occur simultaneously and interdependently. It is also possible that auditory cortex neuronal ensemble activity entrains selectively in different frequency ranges when processing stress-timed versus syllable-timed languages (δ versus θ, respectively).

Spatiotemporal Modulation of Ongoing Oscillatory Activity as a Spectrotemporal Filter Mechanism of Auditory Selective Attention in A1

To investigate whether frequency-dependent entrainment persists when monkeys have to selectively attend to one of two competing streams, and whether it plays a mechanistic role in auditory selective attention, we presented the monkeys with two simultaneous auditory stimulus streams that differed in their rhythm and frequency content. The monkeys were cued to attend to only one of the streams, and attending was confirmed behaviorally by selective responding to the deviants in the cued stream.

Figure 5A displays representative CSD (from the supragranular layers) and MUA (averaged across all layers) responses to attended and ignored stimulus streams from two concurrently recorded A1 sites. When the frequency content of the attended stimulus stream matched the preferred frequency (BF) of a given A1 site (2 kHz stream in 2 kHz region and 8 kHz stream in 8 kHz region), both CSD and MUA responses were larger than in response to stimuli in ignored streams. Another apparent difference between responses to attended and ignored streams is that although δ oscillations are entrained with their high excitability phases to attended stimuli (negative trending in CSD with concomitant prestimulus MUA increase), the baseline is flat for ignored stimulus streams. The lack of entrainment in this condition is also confirmed by the finding that δ phases (histogram insets) are random in relation to auditory stimuli. The same is true for attended versus ignored non-BF stimulus streams: whereas attended non-BF stimuli entrain δ oscillations, ignored stimuli have no detectable effect. Just as is the case of single stimulus streams, non-BF stimuli entrain δ oscillatory activity in counter phase to BF stimuli: the phase of entrained δ oscillation is near its positive peak at stimulus onset with a concomitant prestimulus decline in MUA, signaling low excitability. It is also apparent that compared to the effect of attention on BF responses, attending to stimuli with frequencies other than the BF of the neuronal ensemble has an opposite effect: both CSD and MUA response amplitudes are attenuated in the attended versus ignored condition.

Similar to the representative recording in Figure 5A, the pooled ITC data in Figure 5B show that although δ phase distribution

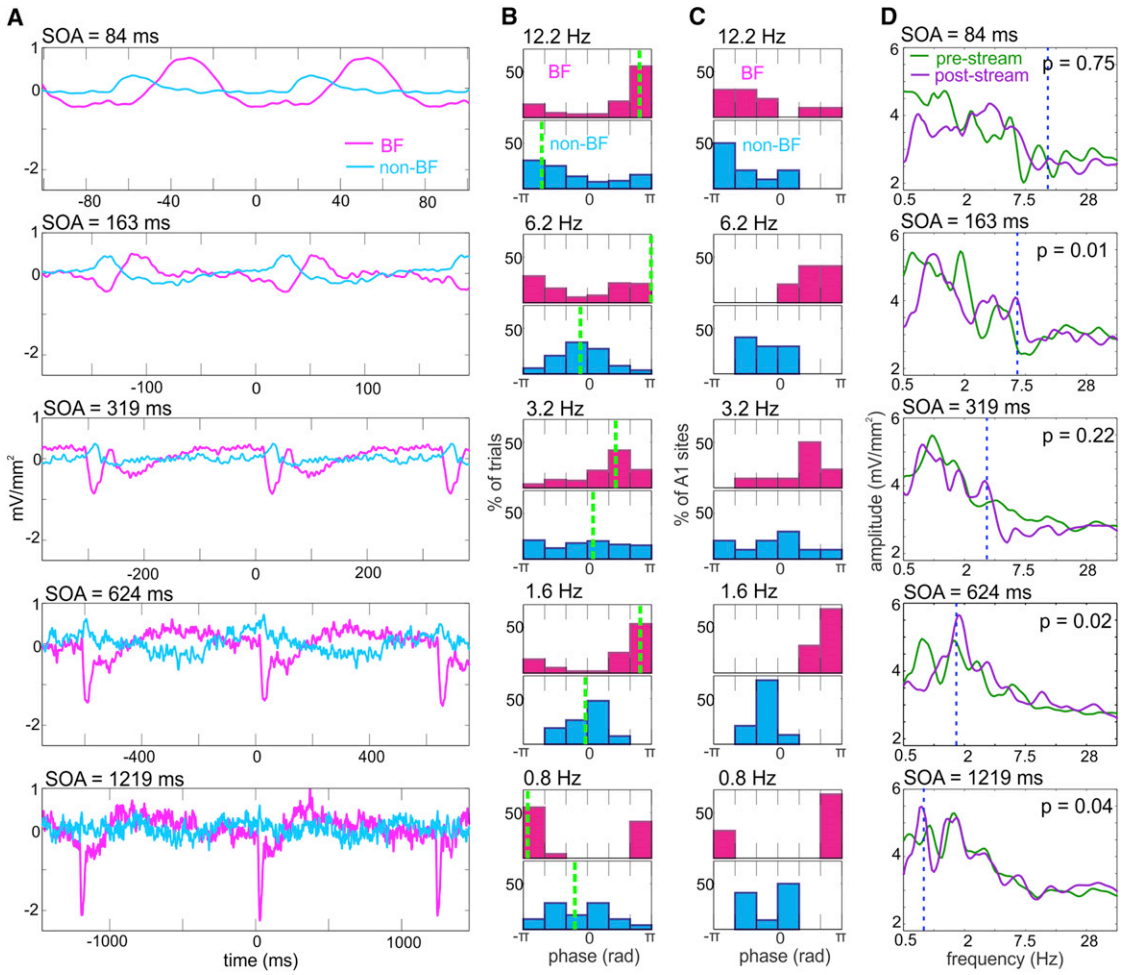

Figure 4. The Rhythm Specificity of Oscillatory Entrainment

(A) Averaged supragranular CSD related to stimuli presented at 12.2, 6.2. 3.2, 1.6, and 0.8 Hz rates (from top to bottom) in a representative experiment. The time frame of the graphs was adjusted such that 2.4 × SOA is displayed in all cases, which includes three event-related responses. It appears that in-between responses to stimuli, the baseline fluctuates in opposite phase when attended stimulus frequency matches the BF of the recordings site versus when it does not, with the exception of the highest presentation rate.

(B) Single trial phase distributions at the presentation rate from the same experiment show that phase opposition is strongest and phase distribution is most biased at 0.8, 1.6, and 6.2 Hz, which correspond to the frequency of dominant δ and θ oscillatory activity in primary auditory cortex. Although phase distribution is also biased in response to stimuli presented at 12 Hz, it appears that this is a result of a phasic component, not a sinusoidal oscillation (see A). Mean phases are indicated by the green dotted lines.

(C) Mean phases in response to attended BF and non-BF tone streams pooled across all experiments. Similar to the single trial data (B), phase opposition is strongest specifically at repetition rates that correspond to the dominant frequency of δ and θ band ongoing oscillatory activity.

(D) Averaged spectrograms of ongoing oscillatory activity recorded prior to (prestream, green) and following (poststream, lilac) the presentation of the 1 min long rhythmic stimulus streams. To determine whether poststream ongoing activity reflects the rhythm of the attended stimulus stream even three to four cycles after the stream ends (similar to results in Figure 3), the time frame in which ongoing activity was measured was adjusted based on the SOA within the streams: it was set to −4 × SOA – −2 × SOA preceding and 2 × SOA – 4 × SOA following each rhythmic stream. The presentation rate of streams is indicated by blue dotted lines in the spectrograms, and p values are associated with the statistical comparison of pre- versus poststimulus amplitudes across experiments at these frequencies (Wilcoxon signed rank). It is apparent that poststream ongoing neuronal activity reflects the structure of attended rhythmic streams at rates from 0.8–6.2 Hz (not at 12.2 Hz), and that similar to phase opposition (B and C), this effect is strongest at frequencies that match the frequency of dominant ongoing oscillatory activity in the sub-α frequency range.

See also Figures S2 and S3.

was significantly biased related to attended stimulus streams, it was generally random (BF: 81%, non-BF: 94%) in relation to the timing of stimuli in ignored stimulus streams. This indicates that when stimulus streams are presented simultaneously, only the attended stream entrains δ oscillatory activity, even if the frequency of attended auditory stimuli does not match the BF of a given A1 region. Thus strikingly, entrainment in a given A1 region is determined by top-down influences rather than the stimulus preference of the neuronal ensemble. As a result, the frequency and phase of entrained oscillatory activity in the

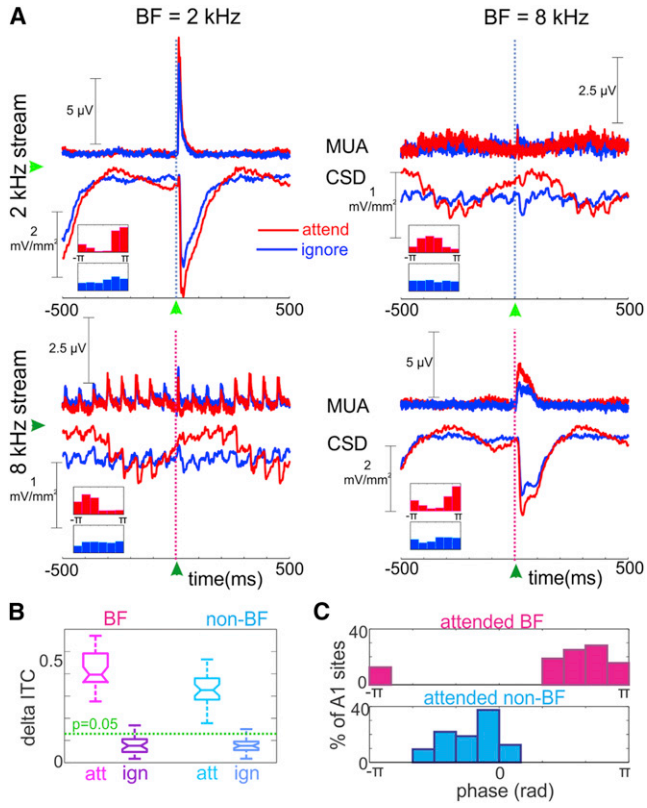

Figure 5. MUA and CSD Responses to Attended and Ignored Auditory Stimuli in a Selective Auditory Attention Task

(A) Traces show laminar MUA (upper) and supragranular CSD responses (lower) to attended (red) and ignored (blue) stimulus streams. The data were recorded simultaneously in two A1 sites tuned to 2 kHz (left) and 8 kHz (right). The panels in the figure are arranged the same way as in Figure 1, but the data are from a different experiment and are recorded in selective attention as opposed to cueing trial blocks. Histogram insets show the distribution of δ oscillatory phases at stimulus onset (0 ms, marked by green arrows), across trials aligned to the timing of attended (red) and ignored (blue) stimuli. Similar to when presented alone, attended BF stimulus streams (top left and bottom right red) entrain δ oscillations so that high excitability phases coincide with stimuli (negative peak in CSD, increased prestimulus MUA), whereas non-BF stimulus streams entrain oscillatory activity so that low excitability phases coincide with stimuli (positive peak in CSD, suppressed prestimulus MUA). Concurrently, responses to attended BF stimuli are amplified, whereas responses to attended non-BF stimuli are attenuated compared to the ignore condition. The "sawtooth" pattern in responses aligned to the onset of stimuli in non-BF stimulus streams (bottom left and top right) is due to responses to simultaneously presented BF stimuli. The amplitude change of these "background" responses across the time frame illustrates that it is related to the phase of entrained oscillations: responses are smallest around the time when attended non-BF stimuli occur (0 ms), because entrained δ is in its low excitability phase (positive peak).

(B) Boxplots of pooled δ ITC during attended and ignored BF and non-BF stimulus streams (n = 32). Although δ ITC was significant in all experiments in the case of attended streams, it was mostly nonsignificant related to ignored ones. ITC values related to attended streams did not significantly differ from ITC values in the single stream condition, displayed in Figure 2A (Wilcoxon rank sum, $p < 0.01$).

(C) Histograms show the opposite distribution of mean δ phases at stimulus onset in response to attended BF and non-BF stimulus streams across experiments in selective attention trial blocks for all experiments (n = 32), similar to when stimulus streams were presented alone (Figure 2B).

selective auditory attention condition only reflects critical temporal and spectral features of the attended stimulus stream. Similar to the single auditory stream condition, the distribution of mean δ phases at stimulus onset across all experiments (Figure 5C) was significantly different for attended BF versus non-BF stimulus streams (Fisher's nonparametric test for the equality of circular means, $p < 0.01$).

Parallel to the phase effects, as described above, the effect of attention on response amplitudes was also opposite for BF versus non-BF stimulus streams (Figure 5A). We found that although MUA response amplitudes across all experiments were significantly larger for attended compared to ignored BF stimuli (on average 24% response amplitude increase), response amplitudes were significantly lower when non-BF stimuli were attended (Wilcoxon signed rank, $p < 0.01$; Figure S2). Because a wealth of recent studies have demonstrated the local modulatory effects of oscillatory phase on event-related responses and perception, it is feasible to think that the observed opposite sign amplitude differences in attended versus ignored conditions are due to the entrained oscillations modulating local excitability across A1 in opposite directions depending on whether the attended frequency content matches the BF of an A1 neuronal ensemble (BF region) or not (non-BF regions). If true, we should be able to verify two testable predictions in our data.

First, the response amplitude variance should decrease in cases when a stimulus stream is attended because, as demonstrated above, entrainment results in a relatively constant phase—and thus excitability—at times when attended rhythmic inputs are predicted to arrive. Figure 6 shows that although the direction of attention's effect on MUA response amplitude depends on the match of inputs with local neuronal tuning properties, response amplitude variability clearly decreases when stimuli are attended, in the representative case (Figure 6A) and across the data set (Figure 6B; Wilcoxon signed rank test, $p < 0.01$).

Second, the amplitude of responses to ignored stimuli should depend on their temporal relationship to stimuli in the attended stream, because excitability at any time point is determined by the phase of oscillations that are entrained by inputs related to attended tones. To test this prediction, we analyzed the relationship between MUA response amplitudes to ignored, preferentially processed (BF) stimuli, and their timing relative to attended non-BF tones. The entrainment of δ oscillations to their low excitability phases with attended non-BF streams predicts that responses to ignored BF tones should be suppressed around the time when attended non-BF tones occur (at short relative SOAs). As an example and pooled data (Figures 6C and 6D) illustrate, this is indeed the case: there is a characteristic relative SOA-dependent response amplitude variation, with attenuated responses to ignored (background) BF tones around the time when attended non-BF stimuli occur (relative SOA = 0). This result highlights an important aspect of the subthreshold response modulation: just like the enhancement of the attended frequency content due to increased excitability occurs at time points when attended stimuli are predicted to occur, the suppression of responses related to nonattended frequency content is also maximal at these time points due to predictively decreased excitability. The same effect can be observed in the

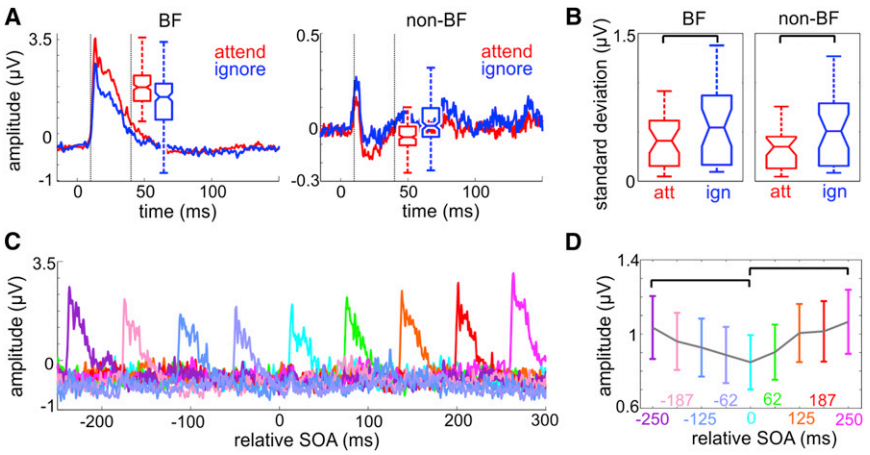

Figure 6. The Effect of Attention on MUA Response Amplitudes in Single Trials

(A) Attended (red) and ignored (blue) averaged MUA responses to BF (left) and non-BF (right) stimuli from a representative experiment. The boxplots show the distribution of response amplitudes (10–40 ms poststimulus) across all single trials. They indicate that enhanced and suppressed averaged responses to attended stimuli compared to ignored ones are both mainly due to a decrease in response amplitude variation across trials.

(B) Standard deviation of single trial response amplitudes pooled across all experiments (n = 32). Brackets indicate significant differences (Wilcoxon signed rank test, $p < 0.01$).

(C) Averaged responses to ignored pooled BF stimuli based on their timing to the onset of attended non-BF stimuli (relative SOA) from a representative experiment. Response amplitudes are smallest around the time when attended non-BF stimuli occur.

(D) Response amplitudes to ignored BF stimuli sorted based on their SOA relative to attended non-BF stimuli averaged across all experiments. The error bars denote SE, whereas brackets indicate that responses to BF stimuli were significantly attenuated when they co-occurred with attended non-BF stimuli compared to when they occurred in between attended stimuli (Wilcoxon signed rank test, $p < 0.01$).

CSD and MUA waveforms averaged to the presentation of non-BF stimuli in Figure 4A directly. Because the two SOAs used in our auditory streams were set to have a common multiple ($9 \times 624.5 = 10 \times 562.05 = 5620.5$), the temporal relationship between stimuli in the two concurrent streams was not completely random: every ninth stimulus in the slower stream had the same relative SOA to stimuli in the faster stream. As a result, "background" BF tone-related responses do not "average out," and directly illustrate the relation of response amplitude and the phase of oscillation entrained to attended non-BF stimuli.

Taken together, these data strongly suggest that auditory selective attention-related response amplitude modulations are a consequence of low frequency oscillations entrained with opposing excitability phases across A1 neuronal ensembles processing attended versus nonattended frequency content. As a result, when attended stimuli are predicted to occur, the A1 region tuned to attended frequency content is in a high excitability state, whereas other A1 regions are in a low excitability state, which serves as a spectral filter at time points when attended stimuli are predicted to occur. Consequently, if an ignored stimulus co-occurs with an attended one, the response to it will be maximally suppressed, minimizing the influence of ignored information on the processing of attended information at key time points. In other words, there are ideal regions both in the spectral and temporal feature dimensions, and at the crossing of these is the frequency and timing of the attended inputs that thus get amplified, whereas all other inputs outside this crossing will be more or less dampened, depending on their "position" in the spectrotemporal plane.

DISCUSSION

The central finding of this study is that when attended auditory stimuli form rhythmic streams, auditory selective attention can be implemented through the entrainment of ongoing neuronal oscillations, providing concerted modulation of neuronal excitability across neuronal ensembles of the primary auditory cortex. The rhythm and attended frequency content-dependent modulation of ongoing neuronal oscillations via entrainment enables the brain to form an adaptive representation of the task-relevant event stream in A1 in the form of subthreshold neuronal activity. This consists of topographically organized excitability fluctuations which in turn modulate auditory responses across tonotopically organized neuronal ensembles in A1, resulting in amplification, sharpening, and stabilization of the attended sensory information. Specifically, our results indicate that the match between attended stimulus frequency and frequency tuning at any A1 site determines the phase of oscillatory entrainment and thus, whether attention will have a facilitative or suppressive effect on neuronal activity at time points when attended stimuli are predicted to occur. This two-dimensional, spectrotemporal filter provides a mechanism to segregate attended stimulus streams from both temporally and spectrally overlapping ones: if stimuli are temporally overlapping, neuronal activity related to ignored frequency content will be suppressed, and if stimuli are spectrally overlapping, neuronal activity at nonattended time points will be suppressed. Because ignored stimulus streams do not seem to significantly affect ongoing neuronal activity in A1, it appears that attended auditory stimulus streams are represented selectively by subthreshold neuronal activity.

Our findings substantiate the notion that subthreshold excitability fluctuations in distributed neuronal ensembles form the context for the processing of specific sensory content (Buzsáki and Chrobak, 1995). Our data suggest that by utilizing phase reset and entrainment, attention generates a dynamically-evolving, spectrotemporal "model" of the attended auditory stimulus stream in the form of modulatory subthreshold oscillations distributed across the neuronal ensembles comprising the tonotopic map in A1. Guided by this context, information-bearing suprathreshold responses related to the content of the attended auditory stream that fit this model are enhanced and stabilized at the expense of responses to unattended or "background" auditory stimuli. This multidimensional spectrotemporal

mechanism is ideally suited for the figure ground segregation of auditory streams, as suggested by the spatiotemporal model of Elhilali et al. (2009a), because it enhances the representation of attended auditory streams along the two most fundamental organizing dimensions of auditory processing: frequency and time (Kubovy and Van Valkenburg, 2001).

Albeit not traditionally considered as the most important speech rhythm, it is well established that the rate of words and phrasal units corresponds to the δ frequency range of the electroencephalogram (EEG) spectrum (1–2 Hz). The close correspondence between the hierarchically nested structure of low to high frequency neuronal oscillations (Bragin et al., 1995; Lakatos et al., 2005b; Canolty et al., 2006) and the temporal structure of speech also suggests an important role for δ oscillatory activity in modulating and parsing inputs structured at finer temporal scales (Schroeder et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012). A recent study provided electrophysiological evidence that δ oscillations indeed play a significant role in speech processing, by showing that attended speech is represented more accurately by δ versus θ band filtered neuronal responses (Ding and Simon, 2012). Our data indicate that the mechanistic role of δ (and θ) oscillations in primary auditory cortex is a stabilization and enhancement of the attended auditory stream at key time points and frequencies. We propose that in the case of speech, this results in an enhanced representation of the "frame" of the relevant speech stream, which facilitates the selective synchronization of the internal electrophysiological context to the temporal regularities in the attended speech stream on multiple timescales (the context of speech). This in turn results in the segregation and efficient, predictive processing of the relevant content that is represented by speech units modulated on a faster temporal scale (syllables, formants, etc.). It is likely, that the alignment of the internal rhythm to specific time points results in the subjective perception of a beat in speech or speech rhythm (Lehiste, 1977; Schmidt-Kassow and Kotz, 2009; Cason and Schön, 2012), and that not only amplitude, but spectral cues and top down influences play an important role in the establishment and maintenance of the synchrony between external and internal context (Obleser et al., 2012).

As an important aside, our results demonstrate that reducing the variability of responses to attended sensory stimuli does not always require "desynchronization," or suppression of low frequency oscillatory activity. On the contrary, when the timing of relevant inputs is predictable, stabilization is achieved by entrainment (enhancement of oscillatory synchrony), which can have added benefits. For example, it is likely that neuronal oscillations in higher order auditory regions also entrain to the timing of attended stimuli, and that this facilitates the central transmission of relevant information via "communication through coherence" (Fries, 2005). In fact, because the tuning of auditory cortical regions outside of A1 is broader (Rauschecker et al., 1995; Tian and Rauschecker, 2004), this mechanism probably plays a substantial role in segregating and enhancing information related to attended stimuli as it is passed through higher levels of the auditory processing hierarchy. Nonetheless, we speculate that in cases when the timing of attended auditory stimuli is unpredictable (like waiting for a starter's gun to go off in a footrace), low frequency oscillations would be suppressed across all of A1 (Schroeder and Lakatos, 2009). It is up to future studies to determine whether in this case excitability is selectively increased in neuronal ensembles processing attended frequency content as, by analogy with attended space, visual studies would predict (Fries et al., 2001; Womelsdorf et al., 2006). Regardless, because the temporal filter mechanism of entrained oscillatory activity could not be utilized without stimulus timing predictability, behavioral performance would probably be degraded.

Regarding the mechanisms of selective attention in general, our results outline novel roles for ongoing neuronal oscillations modulated via entrainment, because we show that in addition to timing (Large and Jones, 1999; Jones et al., 2002; Nobre et al., 2007; Lakatos et al., 2008; Schroeder and Lakatos, 2009; Schroeder et al., 2010), oscillations topographically organized across neuronal ensembles are capable of encoding and predicting other rudimentary stimulus features, like frequency in the auditory domain. Thus the entrainment of ongoing oscillatory activity in A1 is not only predicting when, but also what types of stimuli are anticipated to occur. This begs the question what other feature dimensions can be encoded by the phase of reset/entrained ongoing oscillatory activity across neuronal ensembles. An obvious candidate is spatial location, which is a fundamental feature dimension in vision, and—similar to frequency in A1—is mapped across neuronal ensembles in low-level representation regions in a topographic fashion.

The results of our study provide evidence that "evoked type" and "modulatory" responses are differentially affected by attention. The attentional modulation of evoked responses in A1 is "graded," meaning that although responses to ignored stimuli are suppressed, they still convey information about the stimuli. In contrast to this, we found that entrainment related to ignored stimuli is mostly absent even in A1, which is in line with previous studies (Lakatos et al., 2009; Elhilali et al., 2009b). A recent human study indicates that the representation of ignored stimulus features is largely degraded in higher order auditory cortical regions like the parabelt (Mesgarani and Chang, 2012). We speculate that a progressive bias toward the representation of attended (over ignored) auditory objects is the result of a cascade of downstream "subthreshold filters" (consisting of both local excitability and functional connectivity modulations) across nodes of the auditory processing hierarchy, that enhance attended stimulus features and chisel away the activity related to nonrelevant inputs along multiple feature dimensions.

The fundamental difference in the attention-related modulation of evoked type and modulatory neuronal activity supports distinct sets of driving and modulatory type inputs that generate/orchestrate them, as suggested earlier (Sherman and Guillery, 1998; Jones, 1998a, 1998b; Lakatos et al., 2009; Viaene et al., 2011). Although it is clear that evoked type responses are driven by lemniscal feed-forward thalamocortical afferents, the inputs modulating ongoing oscillatory activity in the supragranular layers are not known. Previous direct anatomical (Hashikawa et al., 1991; Molinari et al., 1995; Jones, 1998a, 1998b; Huang and Winer, 2000) and indirect physiological evidence (Lakatos et al., 2007, 2009; O'Connell et al., 2011) suggests that likely candidates for mediating modulatory effects

are direct thalamocortical inputs from nonlemniscal thalamic nuclei. If true, these inputs have to be under strong top-down control either at their thalamic origins or their cortical targets, given that they are completely absent if stimuli are ignored. Because primary auditory cortex does not receive direct projections from prefrontal cortical areas (Romanski et al., 1999; Kaas and Hackett, 2000) that are thought to play a key role in suppressing irrelevant information (Bartus and Levere, 1977; Knight et al., 1989, 1999), the thalamocortical circuitry responsible for the attentional modulation of phase reset and entrainment most likely involves the thalamic reticular nucleus or TRN (Zikopoulos and Barbas, 2006, 2007; McAlonan et al., 2008). Based on these considerations, we propose that attention projects the prefrontal representation of the attended auditory object onto the topographically organized neuronal ensembles of A1 in the form of subthreshold oscillatory phase modulation anchored to the timing of attended stimuli by fine-tuning nonlemniscal auditory thalamocortical afferents via the TRN.

CONCLUSIONS

We found that attended rhythmic auditory streams entrain ongoing oscillatory activity across A1 regions tuned to different frequencies, and that ignored stimulus streams do not. Although neuronal ensembles processing the attended frequency content are entrained with their high excitability phases, neuronal ensembles tuned to two octaves higher or lower are entrained with their opposite, low excitability phases to the timing of attended events. The coherent but opposite phase oscillations simultaneously amplify responses in A1 regions that process attended frequency content and suppress responses outside this region, resulting in an enhanced representation of the attended frequency content at time points when attended stimuli occur. These results suggest that the mechanism that enables attentive auditory perception to segregate and preferentially process relevant rhythmic auditory streams is the spatiotemporal pattern of entrained subthreshold neuronal oscillations across A1, that models and predicts both spectral (what) and temporal (when) properties of attended auditory streams internally via phase and frequency adjustment, thereby enabling the predictive stimulus-specific modulation of driving inputs.

EXPERIMENTAL PROCEDURES

In the present study, we analyzed electrophysiological data recorded during 32 penetrations of area A1 of two female macaques and one male macaque, who had been prepared surgically for chronic awake electrophysiological recordings. All procedures were approved in advance by the Animal Care and Use Committee of the Nathan Kline Institute. During the experiments, animals sat in a primate chair in a dark, isolated, electrically shielded, sound-attenuated chamber with head fixed in position, and were monitored with infrared cameras. Laminar profiles of field potentials (EEG) and concomitant population action potentials (multiunit activity or MUA) were obtained using linear array multicontact electrodes (23 contacts, 100 μm intercontact spacing). Multielectrodes were inserted acutely through guide tube grid inserts, lowered through the dura into the brain, and positioned such that the electrode channels would span all layers of the cortex (Figure 1), which was determined by inspecting the laminar response profile to binaural broadband noise bursts. The neuroelectric signal recorded was split into the field potential (0.1–300 Hz) and MUA (300–5,000 Hz) range by zero phase shift digital filtering. One-dimensional CSD profiles were calculated as the second spatial derivatives of field potential profiles (Freeman and Nicholson, 1975). The advantage of CSD profiles is that they are not affected by volume conduction like the local field potentials, and they also provide a more direct index of the location, direction, and density of the net transmembrane current flow. After refining the electrode position in auditory cortex, we established the best frequency (BF) of the recording site(s) by determining the maximum MUA in response to a serious of pure tones with frequencies varying from 353.5 Hz to 32 kHz in half octave steps.

The goal of the present set of experiments was to examine the mechanisms of auditory selective attention, not stream segregation per se (although the two appear related), thus our selective attention paradigm is different from the ones commonly used to study stream segregation in human studies. We presented the subjects streams of pure tone beeps at 40 dB SPL with constant stimulus onset asynchronies (SOAs) of either 624.5 or 562.05 ms. The small difference in presentation rate was mainly meant to eliminate any differences in the degree of entrainment that might occur as a consequence of intrinsic resonant properties of the neurons and neuronal circuitry generating ongoing oscillatory activity. The rhythmic stimulus streams consisted of standard, frequently repeating tones whose frequency was set to one of two values determined based on the BF of the recording site: one of the frequency values corresponded to the BF, whereas the other was either two octaves higher (if the site's BF was $\leq$8 kHz) or lower (if the site's BF was >8 kHz). These settings were based on previous results showing that sites tuned to low frequencies have high frequency inhibitory sidebands, whereas sites tuned to frequencies higher than 8 kHz have a low frequency inhibitory sideband (O'Connell et al., 2011). Frequency deviants occurred in the stream of standard tones every 3–9 s randomly. The monkeys had to stick out their tongue in order to get the juice reward associated with deviant tones. To determine that they were attending to the tones and actively discriminating the deviants, we omitted the reward on 20% of the deviants. We only analyzed trials in segments where subjects were reliably licking on juiceless deviants. Two of the subjects performed above 90% correct, whereas one monkey only around 60% correct, which remained stable throughout the course of all experiments.

In the auditory selective attention condition, we presented two stimulus streams concurrently, which differed in their repetition rate (1.6 versus 1.8 Hz corresponding to 624.5 versus 562.05 ms SOA, respectively) and the frequency of tones constituting the two streams. Monkeys were cued to attend to one of the streams by the preceding cueing stream that matched the properties (rhythm and frequency content) of the stream to be attended to. The subjects always responded to deviants in only one of the streams, never to deviants in both streams. The frequency separation between standard tones in the two streams was two octaves with one exception, where it was 1.5 octaves to match the frequency separation of the two concurrently recorded A1 sites.

Utilizing the BF-tone-related laminar CSD profile, the functional identification of the supragranular, granular, and infragranular cortical layers in area A1 (see Figures 1 and S1) is straightforward based on our earlier studies (Schroeder et al., 2001; Lakatos et al., 2005b, 2007). In the present study, we focused the analyses of ongoing and event-related neuronal activity on the supragranular CSD with largest BF tone-related activation (sink), and the MUA averaged across all layers. The reason for this selection is that both ongoing and entrained oscillatory activity are most prominent in the supragranular layer (Lakatos et al., 2005b, 2007, 2008), and they appear to reflect synchronous excitability fluctuations of the local neuronal ensembles across all layers, as evidenced by synchronous MUA amplitude fluctuation across the layers (O'Connell et al., 2011; Figure S1). Also, dominant δ frequency neuronal activity in all cortical layers is largely coherent with supragranular δ oscillatory activity (Lakatos et al., 2005b; O'Connell et al., 2011), albeit actual phase values signaling high or low excitability are different at different laminar locations (see Figure S1).

For the analysis of ongoing and event-related (entrained) δ oscillatory activity, instantaneous power and phase in single trials were extracted by wavelet decomposition. To characterize δ phase distribution across trials, the wavelet transformed single trial data was normalized (unit vectors), the trials were averaged, and the length (modulus) of the resulting vector was computed. The value of the mean resultant length, also called ITC, ranges

from 0 to 1; higher values indicate that the observations (oscillatory phase at a given time point across trials) are clustered more closely around the mean than lower values (phase distribution is biased).

Independent of their frequency composition, cyclically occurring events like the suprathreshold, "evoked type" response waveforms can artificially bias phase measures at the frequency that corresponds to the stimulus presentation rate (see Figure S3 for examples and further explanation). Thus after analyzing the "raw data," we repeated all analyses after a linear interpolation was applied to the single trials in the time interval that in the case of most BF tones contained evoked type activation (5–150 ms). We found that as expected, in the case of non-BF stimulus streams with no significant evoked type responses related to stimuli (O'Connell et al., 2011), this manipulation did not change the distribution of δ phases across trials, and thus the ITC value significantly (Wilcoxon signed rank, $p > 0.01$). However, in the case of BF stimulus streams, ITC values were significantly lower after eliminating the transient evoked type sink in response to the BF tones. A visual inspection of the averaged waveforms confirmed that the δ ITC values we got after the elimination of the transient response reflected the phase distribution of entrained subthreshold oscillatory activity better: in cases where the ITC was not significant with this method, the baseline appeared flat, whereas in cases where ITC was significant, there appeared to be a rhythmically fluctuating baseline (Figure 5A). Additionally we confirmed with data simulations that linearly interpolating a 145 ms segment of sinusoidal oscillations with wavelengths that correspond to the SOAs used in our experiments does not change phase measures derived by wavelet analysis (Figure S3C1), unless the segment contains an "added waveform" (Figure S3C2). Thus the mean phase and ITC values we report in the Results section were calculated from trials where the transient "evoked type" responses were eliminated by linear interpolation.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2012.11.034.

ACKNOWLEDGMENTS

Support for this work was provided by NIH grant DC010415 and DC011490 from the NIDCD, and grants, MH060358, MH086385 and MH49334 from the NIMH.

Accepted: November 29, 2012
Published: February 20, 2013

REFERENCES

Atiani, S., Elhilali, M., David, S.V., Fritz, J.B., and 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.

Bartus, R.T., and Levere, T.E. (1977). Frontal decortication in rhesus monkeys: a test of the interference hypothesis. Brain Res. *119*, 233–248.

Besle, J., Schevon, C.A., Mehta, A.D., Lakatos, P., Goodman, R.R., McKhann, G.M., Emerson, R.G., and Schroeder, C.E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. J. Neurosci. *31*, 3176–3185.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.E., Giard, M.H., and Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. J. Neurosci. *27*, 9252–9261.

Bosman, C.A., Womelsdorf, T., Desimone, R., and Fries, P. (2009). A microsaccadic rhythm modulates gamma-band synchronization and behavior. J. Neurosci. *29*, 9471–9480.

Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., and Buzsáki, G. (1995). Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. J. Neurosci. *15*, 47–60.

Broadbent, D.E. (1958). Perception and communication (London: Pergamon Press).

Buzsáki, G., and Chrobak, J.J. (1995). Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. Curr. Opin. Neurobiol. *5*, 504–510.

Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., and Knight, R.T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. Science *313*, 1626–1628.

Cason, N., and Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. Neuropsychologia *50*, 2652–2658.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.

Ding, N., and Simon, J.Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. USA *109*, 11854–11859.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., and Shamma, S.A. (2009a). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron *61*, 317–329.

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

Freeman, J.A., and Nicholson, C. (1975). Experimental optimization of current source-density technique for anuran cerebellum. J. Neurophysiol. *38*, 369–382.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. *9*, 474–480.

Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science *291*, 1560–1563.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat. Neurosci. *6*, 1216–1223.

Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hear. Res. *206*, 159–176.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front. Psychol. *2*, 130.

Giraud, A.L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. *15*, 511–517.

Hartmann, W.M., and Johnson, D. (1991). Stream segregation and peripheral channeling. Music Percept. *9*, 155–184.

Hashikawa, T., Rausell, E., Molinari, M., and Jones, E.G. (1991). Parvalbumin- and calbindin-containing neurons in the monkey medial geniculate complex: differential distribution and cortical layer specific projections. Brain Res. *544*, 335–341.

Huang, C.L., and Winer, J.A. (2000). Auditory thalamocortical projections in the cat: laminar and areal patterns of input. J. Comp. Neurol. *427*, 302–331.

Izumi, A. (2002). Auditory stream segregation in Japanese monkeys. Cognition *82*, B113–B122.

Jaramillo, S., and Zador, A.M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat. Neurosci. *14*, 246–251.

Jones, E.G. (1998a). A new view of specific and nonspecific thalamocortical connections. Adv. Neurol. *77*, 49–71, discussion 72–73.

Jones, E.G. (1998b). Viewpoint: the core and matrix of thalamic organization. Neuroscience *85*, 331–345.

Jones, M.R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. Psychol. Sci. *13*, 313–319.

Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. USA *97*, 11793–11799.

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

Knight, R.T., Scabini, D., and Woods, D.L. (1989). Prefrontal cortex gating of auditory transmission in humans. Brain Res. *504*, 338–342.

Knight, R.T., Staines, W.R., Swick, D., and Chao, L.L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. Acta Psychol. (Amst.) *101*, 159–178.

Kosaki, H., Hashikawa, T., He, J., and Jones, E.G. (1997). Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. J. Comp. Neurol. *386*, 304–316.

Kubovy, M., and Van Valkenburg, D. (2001). Auditory and visual objects. Cognition *80*, 97–126.

Lakatos, P., Pincze, Z., Fu, K.M., Javitt, D.C., Karmos, G., and Schroeder, C.E. (2005a). Timing of pure tone and noise-evoked responses in macaque auditory cortex. Neuroreport *16*, 933–937.

Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., and Schroeder, C.E. (2005b). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. *94*, 1904–1911.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., and Schroeder, C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron *53*, 279–292.

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

Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., and Schroeder, C.E. (2009). The leading sense: supramodal control of neurophysiological context by attention. Neuron *64*, 419–430.

Large, E.W., and Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. Psychol. Rev. *106*, 119–159.

Lehiste, I. (1977). Isochrony reconsidered. J. Phonetics *5*, 253–263.

Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. *8*, e1000445.

Mathewson, K.E., Fabiani, M., Gratton, G., Beck, D.M., and Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. Cognition *115*, 186–191.

McAlonan, K., Cavanaugh, J., and Wurtz, R.H. (2008). Guarding the gateway to cortex with attention in visual thalamus. Nature *456*, 391–394.

Merzenich, M.M., and Brugge, J.F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res. *50*, 275–296.

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

Molinari, M., Dell'Anna, M.E., Rausell, E., Leggio, M.G., Hashikawa, T., and Jones, E.G. (1995). Auditory thalamocortical pathways defined in monkeys by calcium-binding protein immunoreactivity. J. Comp. Neurol. *362*, 171–194.

Nobre, A., Correa, A., and Coull, J. (2007). The hazards of time. Curr. Opin. Neurobiol. *17*, 465–470.

O'Connell, M.N., Falchier, A., McGinnis, T., Schroeder, C.E., and Lakatos, P. (2011). Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron *69*, 805–817.

Obleser, J., Herrmann, B., and Henry, M.J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. Front. Hum. Neurosci. *6*, 250.

Rauschecker, J.P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. Science *268*, 111–114.

Romanski, L.M., Bates, J.F., and Goldman-Rakic, P.S. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J. Comp. Neurol. *403*, 141–157.

Saleh, M., Reimer, J., Penn, R., Ojakangas, C.L., and Hatsopoulos, N.G. (2010). Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues. Neuron *65*, 461–471.

Schmidt-Kassow, M., and Kotz, S.A. (2009). Attention and perceptual regularity in speech. Neuroreport *20*, 1643–1647.

Schroeder, C.E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. *32*, 9–18.

Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F., and Javitt, D.C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. J. Neurophysiol. *85*, 1322–1327.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. *12*, 106–113.

Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. Curr. Opin. Neurobiol. *20*, 172–176.

Shamma, S.A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. Trends Neurosci. *34*, 114–123.

Sherman, S.M., and Guillery, R.W. (1998). On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators". Proc. Natl. Acad. Sci. USA *95*, 7121–7126.

Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., and Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. J. Neurosci. *30*, 13578–13585.

Tan, A.Y., Zhang, L.I., Merzenich, M.M., and Schreiner, C.E. (2004). Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. J. Neurophysiol. *92*, 630–643.

Tian, B., and Rauschecker, J.P. (2004). Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. *92*, 2993–3013.

Treisman, A.M. (1969). Strategies and models of selective attention. Psychol. Rev. *76*, 282–299.

Van Noorden, L.P.A.S. (1975). Temporal coherence in the perception of tone sequences. PhD thesis, Technische Hogeschoel Eindhoven., Eindhoven, the Netherlands.

Viaene, A.N., Petrof, I., and Sherman, S.M. (2011). Synaptic properties of thalamic input to layers 2/3 and 4 of primary somatosensory and auditory cortices. J. Neurophysiol. *105*, 279–292.

Wehr, M., and Zador, A.M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature *426*, 442–446.

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

Womelsdorf, T., Fries, P., Mitra, P.P., and Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. Nature *439*, 733–736.

Zhang, L.I., Tan, A.Y., Schreiner, C.E., and Merzenich, M.M. (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. Nature *424*, 201–205.

Zikopoulos, B., and Barbas, H. (2006). Prefrontal projections to the thalamic reticular nucleus form a unique circuit for attentional mechanisms. J. Neurosci. *26*, 7348–7361.

Zikopoulos, B., and Barbas, H. (2007). Circuits for multisensory integration and attentional modulation through the prefrontal cortex and the thalamic reticular nucleus in primates. Rev. Neurosci. *18*, 417–438.