Laminar Distribution of stimulus- and task-related variables related to auditory streaming in core and belt auditory cortex

Abbreviated Title: Neural correlates of streaming along the auditory pathway

Taku Banno[1], Harry W. Shirley[1], Yonatan I. Fishman[2], and Yale E. Cohen[1,3,4]

[1]Department of Otorhinolaryngology: Head and Neck Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
[2]Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Bronx NY 10461, USA
[3]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA

Corresponding author:
Yale E. Cohen, PhD
Department of Otorhinolaryngology: Head and Neck Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
E-mail: ycohen@pennmedicine.upenn.edu

Number of pages: 39
Number of figures: 6
Number of tables: 0
Number of multimedia and 3D models: 0
Number of words: 211 (Abstract), 631 (Introduction), and 1488 (Discussion)

Conflict of interest
The authors declare no competing financial interests.

Acknowledgements
This work was supported by the National Institutes of Health (NIDCD). We thank Dr. Jaejin Lee for help with behavioral training.

Abstract

Humans and non-human animals parse the auditory scene into distinct auditory objects or "streams" by grouping together stimuli with common features and segregating those with different features. This process is commonly called "auditory scene analysis". Although previous studies have identified neural mechanisms in the primary (core) auditory cortex that may underlie auditory-stream segregation, we do not have a good understanding of the contribution of cortical regions outside of the core auditory cortex to stream segregation nor do we understand the laminar specificity of these contributions. To examine these issues, we recorded translaminar multiunit activity (MUA) from the core and belt auditory cortex in macaque monkeys while they participated in an auditory streaming task designed to provide an objective measure of auditory-stream segregation. We found that MUA encoded both the stimulus variables and the monkey's behavioral choices related to our streaming task. Overall, core MUA was modulated more by the stimulus variables and the monkey's choices than belt MUA. However, whereas neural correlates of these variables were uniformly distributed in the core auditory cortex, stimulus- and choice-related information was predominantly represented in the superficial and deep layers of belt, respectively. These findings support a differential representation of stimulus- and choice-related variables related to auditory scene analysis along the primate auditory cortical pathway.

Significance Statement

Auditory-stream segregation is a fundamental component of auditory scene analysis. To our knowledge, this is the first study to examine the neural correlates of auditory-stream segregation outside of the core auditory cortex in macaque monkeys and the first to examine the laminar specificity of these neural correlates. We recorded neural responses while monkeys performed a new task for non-human primates designed to objectively test auditory-stream segregation. We found that neural responses reflecting stimulus- and behavior-related variables (i.e., the monkeys' choices) were differentially represented in the core and belt regions of the auditory cortex and across different cortical layers. These findings support a non-uniform representation of stimulus- and behavior-related variables related to auditory scene analysis along the primate auditory cortical pathway.

Introduction

A fundamental goal of the auditory system is to transform complex auditory stimuli into perceptual representations of sound sources in the environment, a process which is commonly called "auditory scene analysis" (Bregman, 1990). One important component of this transformation is "auditory-stream segregation", which involves the grouping of stimuli that have similar spectrotemporal features (e.g., frequency) into one perceptual representation and the concomitant segregation of stimuli that have different features into different perceptual representations. For example, while listening to two interleaved temporal sequences of tone bursts, listeners often report hearing one auditory stream when the frequency separation between the two sequences is small. However, as the frequency separation increases, listeners would be more likely to report hearing two auditory streams. This segregation of multiple overlapping and interleaved sequences of tone bursts into distinct perceptual representations or auditory streams underlies our ability to track a friend's voice in a crowded restaurant or to follow the melody played by a particular instrument during a concert.

Although behavioral aspects of auditory-stream segregation have been studied extensively (see Moore and Gockel, 2012; Denham and Winkler, 2020; Oh et al., 2022 for review), the neural bases of auditory-stream segregation still remain unclear. For example, because neurophysiological studies of auditory streaming have been largely restricted to the analysis of neural responses in the mammalian primary auditory cortex (Fishman et al., 2001a, 2004; Micheyl et al., 2005; Elhilali et al., 2009; Fishman and Steinschneider, 2010; Christison-Lagay and Cohen, 2018; Selezneva et al., 2018) or its avian analogue (Bee and Klump, 2004), we do not have a good understanding of the contribution of non-primary regions of auditory cortex (Kaas and Hackett, 2000) to auditory streaming and whether these different brain regions differentially contribute to streaming. Additionally, as most auditory streaming studies have been conducted in passively listening animals (Fishman et al., 2001a; Bee and Klump, 2004; Fishman et al., 2004; Micheyl et al., 2005; Elhilali et al., 2009; Lu et al., 2017; but see Christison-Lagay et al., 2018 and Selezneva

et al., 2018), the relationship between streaming behavior and simultaneous measures of neural activity has yet to be fully elucidated. Finally, except for important studies of auditory attention (Lakatos et al., 2005, 2008, 2009; Francis et al., 2018), the contributions of different cortical laminae to auditory behavior, especially in non-human primate models of hearing and in regions downstream from primary auditory cortex, have not been thoroughly characterized.

To address these outstanding issues, we examined the neural correlates of auditory streaming in rhesus macaque monkeys at two scales of analysis: cortical area and cortical lamina. We recorded simultaneously, via a multi-channel probe, from different cortical layers in the core or belt regions of the auditory cortex while monkeys performed a target detection task that provided an objective measure of auditory streaming (Sussman and Steinschneider, 2009). In this streaming task, monkeys listened to a temporal sequence of interleaved low- and high-frequency tone bursts and detected a deviantly loud "target" tone burst that was embedded in the low-frequency sequence (Fig. 1a). We found that multiunit activity (MUA) was sensitive to the stimulus variables of the task (i.e., the frequency values of the low- and high-frequency tone bursts). We also found that MUA was modulated by the monkeys' behavioral choices (i.e., hits versus misses). Although MUA was modulated strongly in both the core and belt regions of the auditory cortex, the stimulus variables and the monkeys' choices modulated MUA more in the core auditory cortex than in the belt. Further, the laminar distribution of these effects differed between the core and belt regions of the auditory cortex: whereas stimulus- and behavior-related modulations were widely distributed across cortical layers in the core, MUA in superficial layers of the belt was modulated more by stimulus variables, whereas MUA in deeper layers was modulated more by choice.

Materials and Methods

We conducted multilaminar extracellular recordings in two adult male macaque monkeys (*Macaca mulatta*; monkey D and monkey C) while they performed a task that objectively tested auditory streaming. The Institutional Animal Care and Use Committee of the University of Pennsylvania reviewed and approved all the procedures and protocols. We conducted all the surgeries under general anesthesia with aseptic techniques.

Identification of core and belt auditory cortex

We initially identified the locations of the core and belt regions of the auditory cortex through MRI scans of each monkey's brain and the stereotactic locations of these brain regions (Frey et al., 2004; Saleem and Logothetis, 2012). We functionally verified the anatomical location of each electrode penetration by generating and examining the cortical laminar profiles (as characterized by a current source density [CSD] analysis), calculating the spectrotemporal receptive field (STRF) at each recording site, and constructing tonotopic gradients.

Experimental chamber

We conducted behavioral and recording sessions in a darkened room with echo- and sound-attenuating walls. We seated a monkey in a primate chair with a touch-sensitive lever (monkey D) or joystick (monkey C); the monkey released a lever or moved a joystick to indicate their behavioral report. We stabilized the monkey's head position with a non-invasive head restraint (Drucker et al., 2015). We presented auditory stimuli (RX6 or RX8; Tucker-Davis Technologies) through a calibrated speaker (TR-3 [Anthony Gallo Acoustics] or MSP5 [Yamaha]), which was positioned at eye level ~1 m in front of the monkeys.

Auditory-streaming task

In the auditory-streaming task, a monkey listened to two interleaved sequences of tone bursts (50-ms duration; 25-ms inter-burst interval). One sequence contained "low frequency" (L) tone bursts, and the other sequence contained "high frequency" (H) tone bursts. We presented these tone bursts as a repeating sequence of L-H-H triplets (Fig. 1a), analogous to "ABB" triplets in other streaming studies (van Noorden, 1975; Cusack, 2005; Gutschalk et al., 2005; Micheyl et al., 2005; Sussman and Steinschneider, 2009). The monkey reported (via lever release or joystick movement) a deviantly loud "target" tone burst that was part of the low-frequency tone-burst sequence. The frequency value of the low-frequency tone-burst sequence was held constant across trials, whereas the frequency value of the high-frequency tone-burst sequence varied trial by trial and ranged from 1-24 semitones above the low-frequency value (Fig. 1A). This task is a modified version of a task used to evaluate auditory-stream segregation in human listeners (Sussman and Steinschneider, 2009).

We presented the low-frequency tone bursts at two sound levels: 52 or 68 dB SPL. In each trial, one low-frequency tone burst was presented at 68 dB, whereas all the other low-frequency tone bursts were presented at 52 dB. This 68-dB tone burst was, by definition, the deviant target. In contrast, we presented each of the high-frequency tone bursts at one of four different sound levels, which spanned a range above and below the levels of the low-frequency bursts: 47, 57, 62, and 72 dB SPL. The sound level of each high-frequency tone burst was chosen randomly from these 4 values.

We titrated task difficulty by varying the frequency separation (ΔF) between the low- and high-frequency tone-burst sequences (Fig. 1a). Because the frequency value of the low-frequency tone-burst sequence was held constant across trials, whereas the frequency value of the high-frequency tone-burst sequence varied trial by trial and ranged from 1-24 semitones above the low-frequency value.

Because previous work suggests that monkeys report streaming stimuli in a manner comparable to human listeners (Izumi, 2002; Selezneva et al., 2012; Christison-Lagay and Cohen, 2014), we hypothesized

159 that changes in ΔF would affect the monkeys' performance in a manner similar to human performance
160 (Sussman and Steinschneider, 2009). Specifically, when the ΔF between the two sequences was small, the
161 low- and high-frequency tone-burst sequences would be perceptually integrated into a single auditory
162 stream. As a result, the sound level of the target would be within the variability of the sound levels of the
163 low- and high-frequency tone bursts and would be difficult to detect as being "deviantly loud". In contrast,
164 when the ΔF between the sequences was large, the sequences would perceptually segregate into two
165 auditory streams and the target in the low-frequency tone-burst sequence would be more readily
166 detectable: that is, the louder sound level of the 68-dB target would be more salient relative to the
167 "background" of the softer 52-dB tone bursts.

168 To minimize the possibility that monkeys could guess and anticipate target onset without actually
169 detecting it, we randomized the time of target onset between 675 and 2025 ms, relative to sequence
170 onset. In other words, the target tone burst could appear in any position between the 4^{th} and the 10^{th} L-
171 H-H triplet. The target was never presented prior to the onset of the 4^{th} L-H-H triplet.

172 If the monkey responded within a specified temporal window following target onset (monkey D:
173 window = 650 ms; median response time between target onset and movement: 431 ± 81 ms; monkey C:
174 window = 800 ms; median response time: 517 ± 110 ms), we considered the trial to be a "hit". If the
175 monkey responded after this response window or did not respond at all, we considered the trial to be a
176 "miss". A "false alarm" occurred when the monkey responded before target onset. Conservatively, we
177 also considered very rapid responses (<200 ms for monkey D and <250 ms for monkey C) as false alarms.
178 The target appeared only once in a trial; we did not include catch trials (i.e., trials without a target). We
179 rewarded monkeys only on hit trials. For both miss and false-alarm trials, we added a timeout (1500-2500
180 ms) to the inter-trial interval following these incorrect trials.
181

Neurophysiological Recordings and Recording Strategy

Because we were interested in examining lamina-specific modulations of neural activity related to auditory streaming, we oriented each monkey's recording chamber so that a linear multi-channel electrode (16-channel v-probe, 150-µm spacing between channels; or a 24-channel s-probe, 100-µm spacing between channels; Plexon Inc.) would penetrate the auditory cortex perpendicular to its lamina. This orthogonal orientation further ensured that the spatial sampling of neural activity fulfilled the basic theoretical criteria required for the proper interpretation of one-dimensional CSD analysis (Müller-Preuss and Mitzdorf, 1984; Steinschneider et al., 1992).

At the beginning of each recording session, the electrode was inserted into the brain through a stainless-steel guide tube. The electrode was then advanced gradually with a microdrive (NAN Instruments). While advancing the electrode, we presented an auditory "search" stimulus (100-ms Gaussian noise burst; 10-ms $\cos^2$ rise and fall times; 900 ms inter-burst-interval; sampling rate: 50 kHz) to identify auditory responsive sites. These neural signals were amplified (PZ2 and PZ5, Tucker Davis Technologies), digitized, and stored (RZ2, Tucker Davis Technologies; sampling rate: 24.4 kHz) for online and offline analyses.

Once we encountered sites responsive to the search stimulus, we finely adjusted the electrode's depth so that the largest stimulus-evoked MUA, which typically coincided with a prominent initial current sink in the CSD profile, was positioned in the electrode's middle channels (Fig. 2). MUA measures the summed spiking activity of local neurons within ~50-100 µm of an electrode channel (Legatt et al., 1980; Brosch et al., 1997; Supèr and Roelfsema, 2005). We identified the initial current sink's location through both online and offline CSD analyses; see below. To minimize electrode drift during a recording session and artifactual components in the CSD profile (e.g., due to mechanical compression or distortion of the cortical tissue by the electrode), we retracted the electrode by ~200-450 µm and allowed the tissue to stabilize for $\geq$30 minutes before continuing with data collection. Each recording session lasted about >90

minutes. We confirmed that the electrode position remained relatively stable over the duration of each recording session by comparing the CSD profiles obtained before and after the task.

Next, while the monkey was passively listening, we presented a dynamic moving ripple (DMR) auditory stimulus and simultaneously collected neural activity. We presented this stimulus in order to generate each electrode channel's spectrotemporal receptive field (STRF); see below for more details. From the STRF, we calculated a site's "best frequency" (BF), which was used to determine the frequencies of the tone bursts presented in the auditory-streaming task: we set the value of the low-frequency tone-burst sequence to the BF of the recording site. In rare cases, when the BF was >3 kHz, we set one of the high-frequency values (usually at the largest frequency separation [ΔF]) to this BF value. The monkey then performed the auditory-streaming task. On a trial-by-trial basis, we randomly varied the target onset and the ΔF.

Extraction of local-field potentials (LFPs) and calculation of the CSD

We extracted LFPs by low-pass filtering neural activity with a 4^{th}-order bidirectional Butterworth filter (cutoff frequency: 0.3 kHz). From the LFPs, we derived the one-dimensional laminar CSD profile, as approximated by the second spatial derivative (relative to electrode-channel separation) of the simultaneously recorded LFPs across the electrode channels (Freeman and Nicholson, 1975; Nicholson and Freeman, 1975; Müller-Preuss and Mitzdorf, 1984; Fishman et al., 2001a; Fishman and Steinschneider, 2006). CSD characterizes the laminar pattern of net transmembrane current flow at each electrode channel and indicates whether a channel is near an extracellular current source or sink. A current sink reflects a region of net inward current and is usually generated by net excitatory synaptic activity or passive return from hyperpolarizing currents at adjacent sites. Conversely, a current source indicates net hyperpolarizing current or circuit-completing currents from regions of net depolarization. We also computed the average rectified CSD (AVREC; Mehta et al., 2000a, 2000b; Fishman and

Steinschneider, 2012) by full wave rectifying the CSD waveform at each electrode channel and then averaging these rectified waveforms across electrode channels.

DMR stimulus and STRF analysis

The DMR stimulus is a continuous time-varying broadband noise stimulus that covers the frequency range between 0.1 and 35 kHz (5-min duration; 65 dB spectrum level per ⅓ octave; 96-kHz sampling rate; 24-bit resolution) (Escabí and Schreiner, 2002; Miller et al., 2002). At any instant of time, the stimulus had a sinusoidal spectrum; the spectral modulation frequency (0-4 cycles/octave) determined the density of the spectral peaks. The peak-to-peak amplitude of the ripple was 30 dB. The temporal modulation frequency (0-50 Hz) controlled the stimulus' temporal modulations. Both the spectral and temporal parameters varied randomly and dynamically; the maximum rates of change for these parameters were 0.25 Hz and 1.2 Hz, respectively.

From the multiunit spiking activity, we derived each site's STRF by averaging the spectrotemporal envelope of the DMR relative to time of each spike recorded at each electrode channel. We considered the frequency value corresponding to the STRF peak as the BF of the electrode channel.

Behavioral-data analysis

In each recording session, we calculated the hit, miss, and false-alarm rates. Behavioral d' was defined as the difference between the z-transform of the hit and false-alarm rates. We calculated d' as a function of the frequency separation between the tone-burst sequences.

Neural-data analysis

Our neural-data analyses focused on the time-varying MUA to facilitate comparison between our findings and those of previous studies that examined MUA correlates of auditory streaming (Fishman et al., 2001a,

2004, 2017). MUA and single-unit techniques have been shown to yield similar response properties, whereas MUA is more stable than single-unit activity (Nelken et al., 1994; Supèr and Roelfsema, 2005; Stark and Abeles, 2007).

Extraction of MUA envelope and identification of stimulus-evoked MUA

For each trial and for each electrode channel, we extracted the envelope of the MUA by first bandpass (passband: 0.5-3.0 kHz) filtering the neural signal, full-wave rectifying the filtered signal, and then low-pass (0.6-kHz cutoff frequency) filtering it (Legatt et al., 1980; Steinschneider et al., 1992; Fishman et al., 2001a; Supèr and Roelfsema, 2005; Fishman and Steinschneider, 2006). We then averaged together these trial-by-trial MUA envelopes as a function of the frequency separation between the low- and high-frequency tone bursts and behavioral choice (hits versus misses) and electrode channel. Next, we summed the averaged MUA envelope over a 75-ms window that included each tone burst's 50-ms duration and the 25-ms silent gap that followed the offset of each tone burst. This procedure reduced the time-varying averaged MUA envelope to a single value for each tone burst in the low- and high-frequency tone-burst sequences.

Finally, we z-scored these MUA values relative to an analogous distribution of "baseline" values. This distribution was generated from a random sampling of 75-ms windows during the -1000 ms to -500 ms period that preceded the onset of a tone-burst sequence. If at least one of these z-scored MUA (zMUA) values from the first L-H-H triplet was >1.96 (95% confidence level of z-score value), we considered the zMUA for the channel to be "stimulus evoked". We only report data from recordings with such "stimulus-evoked" zMUAs.

Quantification of stimulus- and task-related neural correlates of auditory streaming

We calculated three indices to test how zMUA was modulated by the frequency values of the low- and high-frequency tone bursts and by the monkeys' choices. These indices were calculated as a function of tone-burst position in a trial and for each electrode channel.

The first index, the "context" index (CI), quantified how zMUA was modulated by a low-frequency tone burst during (1) trials in which the value of the low-frequency tone burst was close to the high-frequency value (i.e., a small ΔF) versus (2) those trials in which the value of the low-frequency tone burst was much lower than the high-frequency value (i.e., a large ΔF). As a reminder, the frequency of the low-frequency tone bursts was held constant across trials, whereas the high-frequency value changed trial-by-trial. Thus, the CI quantified how zMUA elicited by the same low-frequency tone bursts was modulated by different values of the interleaved high-frequency tone bursts. The index was calculated as follows:

$$\mathrm{CI} = \mathrm{abs}(\mathrm{zMUA}^{\mathrm{L}}{}_{\mathrm{sdF,hit}} - \mathrm{zMUA}^{\mathrm{L}}{}_{\mathrm{ldF,hit}}) + \mathrm{abs}(\mathrm{zMUA}^{\mathrm{L}}{}_{\mathrm{sdF,miss}} - \mathrm{zMUA}^{\mathrm{L}}{}_{\mathrm{ldF,miss}}).$$

The superscript L indicates that the zMUA was elicited by a low-frequency tone burst. The subscripts sdF, ldF, hit, and miss indicate the zMUA that was calculated from smallest ΔF trials, from largest ΔF trials, on hit trials, and on miss trials, respectively. We then log-transformed each value to normalize the skewed distribution. Large negative CI values would indicate that the zMUA elicited by the low-frequency tone bursts was relatively invariant to the frequency value of the high-frequency tone bursts. Positive CI values would indicate that the zMUA elicited by the low-frequency tone bursts was modulated by the frequency value of the high-frequency tone bursts. The smallest and largest ΔF values varied across recording sessions (typically 1-8 semitones for monkey D and 4-24 semitones for monkey C).

The second index, the "frequency selectivity" index (FSI), is a more typical assessment of frequency selectivity. The FSI quantified how zMUA was modulated by the smallest and largest frequency values of the high-frequency tone bursts. So, unlike the CI index, the FSI compared zMUA selectivity for different frequency values. The index was calculated as follows:

299 $$\mathrm{FSI} = \mathrm{abs}(\mathrm{zMUA}^{\mathrm{H}}{}_{\mathrm{sdF,hit}} - \mathrm{zMUA}^{\mathrm{H}}{}_{\mathrm{ldF,hit}}) + \mathrm{abs}(\mathrm{zMUA}^{\mathrm{H}}{}_{\mathrm{sdF,miss}} - \mathrm{zMUA}^{\mathrm{H}}{}_{\mathrm{ldF,miss}}).$$

300 The superscript H indicates that the zMUA was elicited by a high-frequency tone burst. The subscripts
301 sdF, ldF, hit, and miss indicate the zMUA that was calculated from smallest ΔF trials, from largest ΔF
302 trials, on hit trials, and on miss trials, respectively. We took the absolute values to eliminate the best-
303 frequency-dependent response difference in FSI. The FSI was then log-transformed each value to
304 normalize the skewed distribution. Large negative FSI values would indicate that the zMUA was not
305 frequency selective, whereas larger positive values would indicate greater frequency selectivity.

306 Finally, to assess the sensitivity of MUA to the behavioral outcome of the trial (i.e., choice), we
307 calculated a "behavioral modulation" index (BMI) in which we compared zMUA selectivity on hit and miss
308 trials with identical ΔF values:

309 $$\mathrm{BMI} = (\mathrm{zMUA}_{\mathrm{sdF,hit}} - \mathrm{zMUA}_{\mathrm{sdF,miss}}) + (\mathrm{zMUA}_{\mathrm{ldF,hit}} - \mathrm{zMUA}_{\mathrm{ldF,miss}}).$$

310 The subscripts are the same as those described above. The BMI values were calculated from zMUA elicited
311 by both low-frequency and high-frequency tone bursts. A BMI value of 0 would indicate that the zMUA
312 was not modulated by choice, whereas higher values would indicate the MUA responses were higher on
313 hit trials than on miss trials.

314 In a second analogous set of analyses, instead of calculating the mean zMUA for each tone burst, we
315 calculated the mean zMUA either over a (225-ms) L-H-H tone-burst triplet or a (150-ms) H-H tone-burst
316 doublet. From these values, we calculated BMI and FSI values, respectively, in order to examine the
317 laminar distribution of response modulations associated with these stimulus- and choice-related
318 variables.

319

320 <u>Statistical analyses of CI, FSI and BMI</u>

321 We used non-parametric statistics and post-hoc comparisons to evaluate null hypotheses. In all statistical
322 tests, we rejected the null hypothesis at $p < 0.05$, false-discovery rate corrected.

Results

While two rhesus macaques (monkey D and monkey C) performed a behavioral task designed to provide an objective measure of auditory streaming (Sussman and Steinschneider, 2009), we recorded laminar profiles of multiunit activity (MUA) and local field potentials (LFPs) in the core and belt regions of the auditory cortex. We collected neural and behavioral data in 38 different recording sessions (16 sessions from monkey D and 22 sessions from monkey C). We report data from the 381 (140 from the core and 241 from the belt) recording sites that had "stimulus-evoked" MUA. We considered MUA to be "stimulus evoked" if there was a significant increase in MUA after the onset of the tone-burst sequence in the streaming task, relative to a "baseline" period that occurred prior to sequence onset; see **Materials and Methods**.

Target detection improved as the frequency separation (ΔF) between the two tone-burst sequences increased

During the streaming task, monkeys detected a "deviantly" loud auditory target. This target stimulus was embedded in a temporal sequence of low-frequency tone bursts (L) that was interleaved with a sequence of high-frequency (H) tone bursts as a repeating L-H-H triplet (Fig. 1a). The sound levels of the low-frequency tone bursts were the same except for the target, which had a higher sound level than the other low-frequency tone bursts. In contrast, the sound level of the high-frequency tone bursts was variable and had sound levels above and below those in the low-frequency tone-burst sequence.

Figure 1b-e plots the monkeys' hit rate and behavioral sensitivity (d'). We found that both hit rate and d' increased as we increased the frequency separation (ΔF) between the tone-burst sequences (Fig. 1b; $\chi^2(4) = 27.53$, $p = 1.56 \times 10^{-5}$, Fig. 1c; $\chi^2(4) = 27.15$, $p = 1.86 \times 10^{-5}$, Kruskal-Wallis test, H_0: medians are the same across all ΔFs; Fig. 1d; $Z = -4.97$, $p = 6.80 \times 10^{-7}$, Fig. 1e; $Z = -4.79$, $p = 1.64 \times 10^{-6}$, two-tailed Wilcoxon signed-rank test, H_0: medians are the same for the smallest and the largest ΔFs). These

behavioral results are consistent with the hypothesis that, like human listeners (Sussman and Steinschneider, 2009), target detection is more reliable when the low-frequency tone-burst sequence (which contained the target stimulus) is perceptually segregated from the high-frequency tone-burst sequence, which occurs at larger values of ΔF.

Core and belt auditory cortex are differentiated by their current source density (CSD) profiles and spectrotemporal receptive fields (STRFs)

We used both CSD profiles and STRFs to differentiate between the core and belt regions of auditory cortex; to date, functional means to differentiate between these cortical regions have not been fully characterized. Whereas CSD profiles in the core auditory cortex have been extensively characterized (e.g., Steinschneider et al., 1992; Fishman et al., 2001b; Lakatos et al., 2005; Szymanski et al., 2009; Fishman and Steinschneider, 2010), less is known regarding the laminar distribution of current flow in non-primary auditory cortex (Fu et al., 2004; Kajikawa et al., 2015), especially in non-human primate models. Similarly, STRFs in the core and belt auditory cortex have not been well characterized in non-human primates.

Using standard analytical techniques, we derived the one-dimensional CSD profiles from the LFPs to compare the laminar distribution of net transmembrane extracellular current flow associated with synaptic activity of neural ensembles in the core and belt auditory cortex (Fig. 2a-f) (Freeman and Nicholson, 1975; Nicholson and Freeman, 1975; Müller-Preuss and Mitzdorf, 1984; Mitzdorf, 1985; Steinschneider et al., 1992). In the core region, the CSD profile displayed a characteristic dipole pattern indicating net current influx (a putative current sink) and efflux (a putative current source) (Fig. 2b). In the middle channels of the electrode array, the CSD showed a sharp negative deflection (current sink) soon after stimulus onset (indicated by an arrow in Fig. 2b). In the same channel, we also observed large increases in MUA (Fig. 2c). This large initial current sink, which typically coincides with the largest increases in MUA, is a characteristic feature of stimulus-evoked laminar response profiles in the core auditory cortex

(Steinschneider et al., 1992; Fishman et al., 2001b; Lakatos et al., 2005; Fishman and Steinschneider, 2010) and is consistent with post-synaptic depolarization of neural populations within putative input (granular) layers of the core auditory cortex (lamina 4 and lower lamina 3).

We found a somewhat different CSD profile in belt auditory cortex (Fig. 2d-f). The CSD profile again showed dipoles evoked by the noise bursts, and we could identify a clear negative deflection immediately after stimulus onset (indicated by the arrow in a middle channel of the electrode array; Fig. 2e). As in core, we observed increases in MUA in the same electrode channel as the initial current sink (Fig. 2f). We operationally defined the channel exhibiting the initial current sink with a concomitant increase in MUA as the input layer of the belt region of the auditory cortex. However, unlike the core, the adjacent channels displayed comparatively weak MUA responses. Further, the current sources and sinks were overall less sharp and temporally less precise in the belt auditory cortex than in the core auditory cortex (Fig. 2e, f).

These differences between the core and belt CSD profiles became clearer when we calculated the average rectified CSD (AVREC; Mehta et al., 2000a, 2000b; Fishman and Steinschneider, 2012). Because the AVREC represents the average (absolute) net extracellular current flow across cortical layers, it is useful tool to identify gross differences in the temporal dynamics of current flow between cortical areas. Figure 2k plots the AVREC for these two core and belt sites. As can be seen, the AVREC in the core auditory cortex has a shorter peak latency, sharper onset, and overall larger amplitude than the AVREC in the belt auditory cortex.

In addition to differences in their CSD profiles, we found differences between the STRFs in the core and belt (Fig. 2g-j). As reported in other studies (Miller et al., 2002; Atencio and Schreiner, 2010), core STRFs had short latencies (typically <20 ms) with small circumscribed excitatory and inhibitory response fields (Fig. 2g, h). Further, the variability of BF across electrode penetrations was larger than

within electrode penetrations (F(9,166) = 7.17, $p = 9.37 \times 10^{-9}$; one-way ANOVA), suggesting that BF is relatively constant across cortical laminae (Atencio and Schreiner, 2010).

Belt STRFs, like core STRFs, also had clearly structured excitatory and inhibitory response fields (Fig. 2i, j). Belt STRFs, however, tended to have longer response latencies (26.7 ms and 48.3 ms in the example STRFs, Fig.2i and 2j, respectively) than core STRFs (16.5 ms and 15.2 ms in the example STRFs, Fig. 2g, and 2h, respectively). Indeed, across our population, the core STRF latency (37.8 ± 21.4 ms, median ± median absolute deviation) was significantly shorter than the belt STRF latency (42.9 ± 17.9 ms) (Z = -2.32, p = 0.02; Wilcoxon rank-sum test; H_0: median latency is the same in the core and belt). Qualitatively, belt excitatory and inhibitory fields tended to be less circumscribed (i.e., they had broader spectral or longer temporal profiles). But, like the core auditory cortex, our estimates of BF were relatively constant across lamina (BF = 208 Hz and 169 Hz, respectively for the STRFs shown in Fig. 2i and 2j). As in the core, the variability of BF across electrode penetrations was larger than within an electrode penetrations (F(19,197) = 6.81, $p = 1.23 \times 10^{-13}$; one-way ANOVA).

MUA encoded a mixture of stimulus and task variables

In both the core and belt regions of auditory cortex, MUA in both the core and belt regions of auditory cortex was modulated by both the stimulus and task variables of the streaming task. A MUA profile from an example site in the core auditory cortex is shown in Figure 3a. In this example, MUA was discretely phase-locked to each tone burst in the L-H-H sequence. When ΔF was small (i.e., the frequency value of the high-frequency tone bursts was near that of the low-frequency tone bursts, which was set to a site's BF), both the low- and high-frequency tone bursts elicited strong bursts of MUA. However, as ΔF increased, the high-frequency tone bursts elicited lower amplitudes of MUA, whereas the MUA elicited by the low-frequency tone bursts (which were set to the site's BF) remained relatively unaffected (Fig. 3a, top). This stimulus-related modulation is further highlighted when we calculated the difference in MUA

amplitude between the small and large ΔF conditions (8 and 24 semitones, respectively; Fig. 3a, bottom). The MUA at this example site was not differentially modulated by the (louder) deviant target (Fig. 3a, top, right) nor was it modulated by the monkey's behavioral choice (hits versus misses; Fig. 3a, right, middle and bottom).

Figure 3b shows a MUA profile from an example site in the belt auditory cortex. Like the example profile from the core, the MUA was phase-locked responses to each tone of the L-H-H sequence and for both small and large values of ΔF (1 and 8 semitones, respectively; Fig. 3b, top). However, these responses were temporally less precise than those seen in the core (Fig. 3a). Furthermore, unlike responses in the core, as ΔF increased, the amplitude of MUA elicited by the low-frequency tone bursts tended to increase (Fig. 3b, bottom). Finally, the target tone burst elicited greater MUA than the non-target tone bursts (Fig. 3b, right). However, the MUA was not appreciably modulated by the monkey's behavioral choices (hits vs misses; Fig. 3b, right, middle and bottom).

Figure 3c shows a second MUA profile from the core auditory cortex. Like the examples in Fig. 3a and 3b, the MUA was phase-locked response to the low-frequency tone burst, which became more distinct in large ΔF trials (Fig. 3c; top). In addition, we found that the MUA was modulated by the monkey's choices: the MUA response was higher on hit trials than on miss trials (Fig. 3c: middle and bottom).

Stimulus- and task-related variables of the auditory streaming task are differentially and dynamically encoded in core and belt auditory cortex

To quantify the stimulus- and task-related variables encoded in the MUA, we calculated three different indices (see **Materials and Methods**). We calculated each index for each tone burst in a triplet (i.e., the low-frequency tone burst [L] and the two high-frequency tone bursts [H1 and H2]) and as a function of time (relative to stimulus onset and target onset). For each of these index values, we collapsed the values across electrode channels to form distributions as a function of time and brain region (core and belt).

The first index, the context index (CI), quantified whether neural responses to the (frequency-fixed) low-frequency tone bursts were modulated by the different values of the high-frequency tone bursts. Large negative CI values would indicate that the zMUA elicited by the low-frequency tone bursts was relatively invariant to the frequency value of the high-frequency tone bursts. Positive CI values would indicate that the zMUA elicited by the low-frequency tone bursts was modulated by the frequency value of the high-frequency tone bursts. The results of this analysis are shown in Figure 4a (left). As can be seen, core and belt CI values at the beginning of the tone-burst sequence were relatively small (CI at the first triplet position; -0.22 ± 0.10 in the core and 0.21 ± 0.06 in the belt, mean ± standard error). However, as the sequence unfolded, the CI increased sharply and remained relatively constant through target presentation (core: $\chi^2(4) = 156.1$, $p < 1.0 \times 10^{-10}$; belt: $\chi^2(4) = 112.9$, $p < 1.0 \times 10^{-10}$; Kruskal-Wallis test, H_0: CI medians are the same across triplet positions). The CI values were consistently larger in the core auditory cortex than in the belt ($\chi^2(1) = 11.57$, $p = 6.70 \times 10^{-4}$; Friedman test, H_0: CI medians are the same in the core and belt).

The second index, the frequency-selectivity index (FSI), quantified the degree to which the MUA elicited by the high-frequency tone bursts was modulated by their different frequency values. The FSI is akin to a traditional measure of frequency selectivity. Large negative FSI values would indicate that the zMUA was not frequency selective, whereas larger positive values would indicate greater frequency selectivity. For both the first and second high-frequency tone bursts in the L-H-H triplet (H1 and H2, respectively), FSI values were generally largest for the first triplet and got smaller over time (Fig. 4a, middle and right; $core_{H1}$: $\chi^2(4) = 5.45$, $p = 0.24$; $core_{H2}$: $\chi^2(4) = 12.56$, $p = 0.01$; $belt_{H1}$: $\chi^2(4) = 10.01$, $p = 0.04$; and $belt_{H2}$: $\chi^2(4) = 10.30$, $p = 0.04$; Kruskal-Wallis test, H_0: FSI medians are the same across triplet positions). In other words, in both the core and belt, MUA frequency selectivity became poorer as the tone-burst sequence unfolded. Although both core and belt MUA had similar temporal dynamics, core FSI values

were consistently larger than those in belt across all time periods (Fig. 4a; H1: $\chi^2(1) = 43.97$, $p < 1.0 \times 10^{-10}$; and H2: $\chi^2(1) = 53.92$, $p < 1.0 \times 10^{-10}$, Friedman test, H_0: FSI medians are the same in the core and belt).

Finally, the behavioral-modulation index (BMI) quantified the degree to which MUA elicited by the tone-burst sequence was modulated by the monkeys' behavioral choices (hits vs misses). A BMI value of 0 would indicate that MUA was not modulated by choice, whereas larger values would indicate increased sensitivity to choice. BMI significantly increased as the tone-burst sequence unfolded over time for all three tone bursts in the triplet both in the core (L: $\chi^2(4) = 52.27$, $p = 1.21 \times 10^{-10}$; H1: $\chi^2(4) = 101.0$, $p < 1.0 \times 10^{-10}$; H2: $\chi^2(4) = 113.3$, $p < 1.0 \times 10^{-10}$; Kruskal-Wallis test, H_0: BMI medians are the same across triplet positions) and in the belt (L: $\chi^2(4) = 53.17$, $p < 1.00 \times 10^{-10}$; H1: $\chi^2(4) = 159.4$, $p < 1.0 \times 10^{-10}$; and H2: $\chi^2(4) = 166.0$, $p < 1.0 \times 10^{-10}$; Kruskal-Wallis test, H_0: BMI medians are the same across triplet positions). Core BMI values tended to be larger than belt BMI values (L: $\chi^2(1) = 23.45$, $p = 1.28 \times 10^{-6}$; H1: $\chi^2(1) = 3.23$, $p = 0.072$; H2: $\chi^2(1) = 35.37$, $p = 2.72 \times 10^{-9}$, Friedman test, H_0: BMI medians are the same in the core and belt).

To further investigate this build-up of BMI values over time, we realigned the BMI values relative to target onset (Fig. 5a). We found that BMI values tended to increase monotonically over time, independent of the frequency value of the tone bursts and whether they were a target (core: $\chi^2(11) = 167.3$, $p < 1.0 \times 10^{-10}$; belt: $\chi^2(11) = 328.7$, $p < 1.0 \times 10^{-10}$; Kruskal-Wallis test, H_0: BMI medians are the same across time). This contrasts with CI and FSI values (Fig. 5b): relative to target onset, CI and FSI values did not change reliably over time (core CI: $\chi^2(3) = 0.53$, $p = 0.91$; belt CI: $\chi^2(3) = 2.15$, $p = 0.54$; core FSI_{H1}: $\chi^2(3) = 1.45$, $p = 0.69$; belt FSI_{H1}: $\chi^2(3) = 0.13$, $p = 0.99$; core FSI_{H2}: $\chi^2(3) = 0.82$, $p = 0.85$; belt FSI_{H2}: $\chi^2(3) = 0.75$, $p = 0.86$; Kruskal-Wallis test, H_0: CI and FSI medians are the same across triplet positions).

This monotonic increase in BMI was further highlighted when we collapsed MUA across the L-H-H triplet and replotted the BMI temporal profile (Fig. 5c; core: $\chi^2(3) = 69.71$, $p < 1.0 \times 10^{-10}$; belt: $\chi^2(3) = 140.1$, $p < 1.0 \times 10^{-10}$; Kruskal-Wallis test, H_0: BMI medians are the same across triplet positions). Even

when we limited this analysis to the triplet containing the target and examined BMI values only for those target triplets that appeared "early" in a tone-burst sequence (i.e., the target that appeared at the earliest triplet position in the recording session; typically 4^{th} and 7^{th} triplet for monkey D and monkey C, respectively) versus those that appeared "late" (i.e., the target that appeared at the latest triplet position in the session; typically 7^{th} and 10^{th} triplet in monkey D and monkey C, respectively), we found that BMI values significantly increased when target onset occurred later in a sequence (Fig. 5d; core: $\chi^2(3) = 14.88$, $p = 0.0019$; belt: $\chi^2(3) = 8.01$, $p = 0.046$, Kruskal-Wallis test, H_0: BMI medians are the same across target positions). Finally, these analyses are consistent with our previous finding (see Fig. 4b) that BMI values were significantly higher in the core auditory cortex than in the belt (Fig. 5c: $\chi^2(1) = 36.64$, $p = 1.42 \times 10^{-9}$; Fig. 5d: $\chi^2(1) = 15.01$, $p = 1.07 \times 10^{-4}$, Friedman test, H_0: BMI medians are the same for core and belt).

Laminar specificity of neural correlates of streaming in core and belt auditory cortex

We reanalyzed our three indices as a function of laminar depth (i.e., supragranular, granular, and infragranular layers), which was operationally identified based on each penetration's CSD profile (Fig. 2), in order to identify the laminar distribution of stimulus- and choice-related modulations of MUA elicited during the streaming task. For our analysis of the laminar distribution of the CI and FSI, we focused on MUA that was elicited by the triplet immediately preceding target onset (i.e., the T-1 triplet position in Fig. 4 and 5) to minimize the possibility that changes in activity were due to an increase in the sound level of the target tone burst. For the BMI, we focused on MUA that was elicited by the triplet containing the target (T triplet position in Fig. 4 and 5).

We found a non-uniform distribution of stimulus- and choice-related representations across different cortical layers (Fig. 6a). A Scheirer-Ray-Hare test (non-parametric two-way [brain region x layer] ANOVA) indicated a significant main effect across layers for all indices (FSI: $H(2) = 18.55$, $p = 9.39 \times 10^{-5}$; CI: $H(2) = 13.83$, $p = 9.93 \times 10^{-4}$; and BMI: $H(2) = 8.52$, $p = 0.014$; H_0: median index values are the same

across layers). However, post-hoc analyses did not reveal differences in the index distributions across the layers of the core auditory cortex ($p > 0.05$; Dunn test). In contrast, we found laminar differences in the distribution of index values in the belt auditory cortex: median FSI and CI values were both lower in the infragranular layers than in the supragranular or granular layers, whereas the median BMI value was higher in the infragranular layers than in the granular layer (Fig. 6a; $p < 0.05$, Dunn test).

Because these indices are not bounded, comparisons based on their absolute values may fail to reveal, or even exaggerate, laminar differences. To address this issue, we z-scored each index and performed pairwise comparisons of the z-scored indices in each layer of each cortical area. Once again, we did not find significant differences for any comparisons in the core auditory cortex (Fig. 6b; $p > 0.05$, Wilcoxon signed-rank test with Bonferroni correction, H_0: median difference is different from zero). In contrast, in the belt auditory cortex, we found that MUA was modulated more by stimulus-related variables (FSI and CI) than by choice (BMI) in the granular layer. In contrast, MUA was modulated more by choice than by stimulus-related variables in the infragranular layer (Fig. 6b; $p < 0.05$, Wilcoxon signed-rank test with Bonferroni correction).

Discussion

In this study, we examined how MUA in the core and belt regions of auditory cortex of rhesus monkeys was modulated during a behavioral task that provided an objective measure of auditory streaming. We found that MUA was modulated by both stimulus- and choice-related variables. Further, we identified two key differences between these variables' representations in the core and belt auditory cortex. First, on average, stimulus- and choice-related MUA modulations were higher in the core than in the belt. Second, whereas these two variables were uniformly distributed across cortical layers in core, we found a layer-dependent representation of these variables in belt, with stimulus- and choice-related modulations predominating in superficial and deep cortical layers, respectively.

Streaming behavior in monkeys is comparable to humans

In our previous study of auditory streaming, we trained monkeys to report whether two interleaved sequences of tone bursts were heard as "one stream" or "two streams" (Christison-Lagay and Cohen, 2014, 2018). As expected from previous work on streaming (van Noorden, 1975; Bregman, 1990; Moore and Gockel, 2002; Carlyon, 2004; Sussman and Steinschneider, 2009), the proportion of the monkeys' reports of "two streams" increased as we increased the frequency separation between the two tone-burst sequences. However, an alternative interpretation is that the monkeys were not reporting their streaming percept per se but rather their categorical judgement of whether the tone-burst sequences had a large or small frequency separation.

Here, we modified the task design so that instead of reports of "one stream" or "two streams", the monkeys reported a deviantly loud target stimulus (Fig. 1). This task change was critical because it enabled a disassociation of the stimulus dimension that the monkeys needed to detect (sound level) from the dimension that affected streaming (i.e., the frequency separation [ΔF] of the tone-burst sequences). In other words, it was the hypothesized segregation of the tone-burst sequences that facilitated target

detection, but ΔF only indirectly affected behavioral performance. Indeed, Sussman and Steinschneider (2009) study, which was conducted in human participants, demonstrated a high correlation between target detection and the participants' subjective reports of stream segregation. Similarly, our monkeys showed a monotonic improvement in performance as we increased the ΔF between the tone-burst sequences (Fig. 1). Together with previous reports (Izumi, 2002; Selezneva et al., 2012; Christison-Lagay and Cohen, 2014), our results add further evidence that non-human primates process auditory streams in a manner like human listeners.

Although qualitatively similar to those in human subjects (Sussman and Steinschneider, 2009), the monkeys' overall hit rates and d' values were lower. These behavioral differences may be attributed, in part, to differences in task design. For example, in the human study, the stimulus sequence was presented for several minutes, whereas our stimulus duration was considerably shorter (<2 s). Because stream segregation "builds up" over time (Bregman, 1978; Carlyon et al., 2001), the longer-duration listening time might have given the human listeners a behavioral advantage. Moreover, during those long stimulus durations, in the Sussman and Steinschneider study (2009), ΔF was held constant with multiple presentations of the same deviant target, whereas, in our task, ΔF varied trial-by-trial with only one target presentation per trial.

Frequency-dependent modulation of MUA during the streaming task

MUA was modulated by the frequency values of the tone-burst sequence (Figs. 3 and 4). For example, we found that MUA sensitivity to differences in the values of the high-frequency tone-burst sequence decreased over time (Fig. 4a). This finding suggests that MUA frequency selectivity to task irrelevant high-frequency tone bursts became poorer as a task trial unfolded. This change in frequency selectivity is consistent with some form of time-sensitive habituation or suppression. In studies in which streaming was examined in passively listening monkeys, such time-dependent habituation or suppression has been

reported as a key neural correlate of auditory streaming in the primary auditory cortex, especially (Fishman et al., 2001a, 2004; Micheyl et al., 2005).

However, our observation that the CI values increased over time (Figs. 3 and 4a) does not readily fit into such a habituation/suppression scheme. Instead, we hypothesize that the dynamics of MUA sensitivity to the low- and high-frequency values of the tone-burst sequence reflect some form of spectrotemporal filter that was engaged when the monkeys were actively participating in our streaming task. Indeed, Lakatos et al. (2013) reported that attention recruits cortical oscillations that can enhance neural responsivity when monkeys are asked to attend to tone bursts in a sequence. Because this filter mechanism is frequency dependent, it could enhance activity in response to tone bursts set to a site's BF (i.e., the low-frequency tone bursts), while simultaneously suppressing responses elicited by non-BF tone bursts (i.e., the high-frequency tone bursts) (Lakatos et al., 2013; O'Connell et al., 2014). Because of these opposing computations, as the tone-burst sequence unfolded, neural populations sensitive to the frequency values of the BF and non-BF tone bursts would become differentially active and thereby provide a means for downstream neurons to readout information about the number of auditory streams in the environment (Fishman et al., 2001a, 2004, 2017). That is, when BF and non-BF neural populations have comparable overlapping levels of activity, downstream neurons could read-out the stimulus as "one stream" but when they have different, non-overlapping activity levels, the read-out would be "two streams". It is worth noting that we cannot comment on whether the present data and this interpretation also support the temporal-coherence model of auditory streaming (Elhilali et al., 2009; Shamma et al., 2011; Rezaeizadeh and Shamma, 2021); this could only be determined if we had presented the tone bursts in the sequences simultaneously and not interleaved as in the present study.

597 *Choice-related modulations in core and belt auditory cortex*

598 MUA in the core and belt auditory cortex was modulated by the monkeys' behavioral choices (i.e., hits
599 versus misses; Figs. 4b and 5). These choice-related modulations were observed in response to both the
600 low-frequency and high-frequency tone bursts. We also found that, as a trial unfolded, choice-related
601 modulations increased, even when we controlled for target tone bursts that occurred "early" in the
602 sequence versus those that occurred "later" (Fig. 5d). This buildup of choice-related activity was also seen
603 in our previous streaming study (Christison-Lagay and Cohen, 2018).

604 Choice-related activity has been reported previously in studies that related behavior to neural
605 activity in the core auditory cortex (Niwa et al., 2012a, 2012b, 2013; Bizley et al., 2013). However, whereas
606 belt activity has been shown to be causally related to an ongoing decision (Tsunada et al., 2016), it is not
607 clear whether core activity is also part of a feedforward process that underlies the ongoing decision.
608 Indeed, our observation that choice-related activity was greater in core than in belt does not fit well into
609 a 'hierarchical' model in which early cortical areas represent sensory attributes of stimulus and later areas
610 convert these representations into a behavioral decision (Gold and Shadlen, 2007; Tsunada et al., 2016).
611 Further, the fact that the choice-related activity was seen prior to the onset of the target and was
612 modulated by the time of target onset within a sequence (Figs. 4 and 5) suggests that a more parsimonious
613 explanation: namely, choice-related activity reflects some form of attention or expectation. Indeed, target
614 anticipation can dynamically recruit top-down attentional processes that modulate neural activity (Ghose
615 and Maunsell, 2002; Nienborg and Cumming, 2009). Further work that compares the time of the actual
616 perceptual decision with changes in neural activity (Tsunada et al., 2016, 2019) as well as causal
617 manipulations would be needed to understand more fully the contribution of this type of activity to
618 behavior.

619

620 *Laminar organization of stimulus- and choice-related MUA*

621 We could not identify a laminar parcellation in the core auditory cortex but were able to identify such
622 laminar differences in the belt (Fig. 6). It is possible, however, that such a distribution does exist in the
623 core but is less robust than in the belt, and our analyses were not sensitive enough to characterize it.

624 We found more choice-related activity in deep (infragranular) layers of the belt auditory cortex,
625 which receive feedback projections from higher-cortical areas (Galaburda and Pandya, 1983; Felleman
626 and Van Essen, 1991; Hackett et al., 2014). These feedback projections may be the source of
627 attentional/expectation signals discussed above. On the other hand, stimulus-related modulations were
628 more pronounced in superficial layers of the belt, consistent with activity transmitted via feedforward
629 projections from earlier sensory areas (Galaburda and Pandya, 1983; Felleman and Van Essen, 1991;
630 Hackett et al., 2014). Although, the laminar distribution of spectrotemporal tuning, tonotopy, and other
631 sensory features have been studied in other model systems (Atencio and Schreiner, 2010; Kanold et al.,
632 2014; O'Connell et al., 2014), this is the first identification of the differential representation of task-related
633 variables in different layers of the non-human primate auditory cortex. Overall, the concordance of these
634 previously described anatomical projection patterns and the laminar distributions of task variables
635 suggests a differential representation of stimulus- and choice-related variables along the ventral cortical
636 auditory pathway during our auditory streaming task.

637

638

References

Atencio CA, Schreiner CE (2010) Columnar Connectivity and Laminar Processing in Cat Primary Auditory Cortex. PLOS ONE 5:e9521.

Bee MA, Klump GM (2004) Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain. J Neurophysiol 92:1088–1104.

Bizley JK, Walker KMM, Nodal FR, King AJ, Schnupp JWH (2013) Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. Curr Biol 23:620–625.

Bregman AS (1978) Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform 4:380–387.

Bregman AS (1990) Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, MA: MIT.

Brosch M, Bauer R, Eckhorn R (1997) Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex. Cereb Cortex 7:70–76.

Carlyon RP (2004) How the brain separates sounds. Trends Cogn Sci 8:465–471.

Carlyon RP, Cusack R, Foxton JM, Robertson IH (2001) Effects of attention and unilateral neglect on auditory stream segregation. J Exp Psychol Hum Percept Perform 27:115–127.

Christison-Lagay KL, Cohen YE (2014) Behavioral correlates of auditory streaming in rhesus macaques. Hear Res 309:17–25.

Christison-Lagay KL, Cohen YE (2018) The Contribution of Primary Auditory Cortex to Auditory Categorization in Behaving Monkeys. Front Neurosci 12:601.

Cusack R (2005) The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci 17:641–651.

Denham SL, Winkler I (2020) Predictive coding in auditory perception: challenges and unresolved questions. Eur J Neurosci 51:1151–1160.

Drucker CB, Carlson ML, Toda K, DeWind NK, Platt ML (2015) Non-invasive primate head restraint using thermoplastic masks. J Neurosci Methods 253:90–100.

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA (2009) Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. Neuron 61:317–329.

Escabí MA, Schreiner CE (2002) Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. J Neurosci 22:4114–4131.

Felleman DJ, Van Essen DC (1991) Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb Cortex 1:1–47.

Fishman YI, Arezzo JC, Steinschneider M (2004) Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116:1656–1670.

Fishman YI, Kim M, Steinschneider M (2017) A Crucial Test of the Population Separation Model of Auditory Stream Segregation in Macaque Primary Auditory Cortex. J Neurosci 37:10645–10655.

672 Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001a) Neural correlates of auditory stream
673 segregation in primary auditory cortex of the awake monkey. Hear Res 151:167–187.

674 Fishman YI, Steinschneider M (2006) Spectral Resolution of Monkey Primary Auditory Cortex (A1)
675 Revealed With Two-Noise Masking. J Neurophysiol 96:1105–1115.

676 Fishman YI, Steinschneider M (2010) Neural Correlates of Auditory Scene Analysis Based on Inharmonicity
677 in Monkey Primary Auditory Cortex. J Neurosci 30:12480–12494.

678 Fishman YI, Steinschneider M (2012) Searching for the Mismatch Negativity in Primary Auditory Cortex of
679 the Awake Monkey: Deviance Detection or Stimulus Specific Adaptation? J Neurosci 32:15747–15758.

680 Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, Arezzo JC, Howard MA, Steinschneider M (2001b)
681 Consonance and Dissonance of Musical Chords: Neural Correlates in Auditory Cortex of Monkeys and
682 Humans. J Neurophysiol 86:2761–2788.

683 Francis NA, Elgueda D, Englitz B, Fritz JB, Shamma SA (2018) Laminar profile of task-related plasticity in
684 ferret primary auditory cortex. Sci Rep 8:16375.

685 Freeman JA, Nicholson C (1975) Experimental optimization of current source-density technique for anuran
686 cerebellum. J Neurophysiol 38:369–382.

687 Frey S, Comeau R, Hynes B, Mackey S, Petrides M (2004) Frameless stereotaxy in the nonhuman primate.
688 NeuroImage 23:1226–1234.

689 Fu K-MG, Shah AS, O'Connell MN, McGinnis T, Eckholdt H, Lakatos P, Smiley J, Schroeder CE (2004) Timing
690 and Laminar Profile of Eye-Position Effects on Auditory Responses in Primate Auditory Cortex. J
691 Neurophysiol 92:3522–3531.

692 Galaburda AM, Pandya DN (1983) The intrinsic architectonic and connectional organization of the superior
693 temporal region of the rhesus monkey. J Comp Neurol 221:169–184.

694 Ghose GM, Maunsell JHR (2002) Attentional modulation in visual cortex depends on task timing. Nature
695 419:616.

696 Gold JI, Shadlen MN (2007) The Neural Basis of Decision Making. Annu Rev Neurosci 30:535–574.

697 Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic Correlates of
698 Streaming in Human Auditory Cortex. J Neurosci 25:5382–5388.

699 Hackett TA, de la Mothe LA, Camalier CR, Falchier A, Lakatos P, Kajikawa Y, Schroeder CE (2014)
700 Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the
701 hierarchical model. Front Neurosci 8:72.

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

703 Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proc Natl
704 Acad Sci 97:11793–11799.

705 Kajikawa Y, Frey S, Ross D, Falchier A, Hackett TA, Schroeder CE (2015) Auditory Properties in the Parabelt
706 Regions of the Superior Temporal Gyrus in the Awake Macaque Monkey: An Initial Survey. J Neurosci
707 35:4140–4150.

708 Kanold PO, Nelken I, Polley DB (2014) Local versus global scales of organization in auditory cortex. Trends
709 Neurosci 37:502–510.

710 Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of Neuronal Oscillations as a
711 Mechanism of Attentional Selection. Science 320:110–113.

712 Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The Spectrotemporal
713 Filter Mechanism of Auditory Selective Attention. Neuron 77:750–761.

714 Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The Leading Sense:
715 Supramodal Control of Neurophysiological Context by Attention. Neuron 64:419–430.

716 Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An Oscillatory Hierarchy Controlling
717 Neuronal Excitability and Stimulus Processing in the Auditory Cortex. J Neurophysiol 94:1904–1911.

718 Legatt AD, Arezzo J, Vaughan HG (1980) Averaged multiple unit activity as an estimate of phasic changes
719 in local neuronal activity: effects of volume-conducted potentials. J Neurosci Methods 2:203–217.

720 Lu K, Xu Y, Yin P, Oxenham AJ, Fritz JB, Shamma SA (2017) Temporal coherence structure rapidly shapes
721 neuronal interactions. Nat Commun 8:13900.

722 Mehta AD, Ulbert I, Schroeder CE (2000a) Intermodal Selective Attention in Monkeys. I: Distribution and
723 Timing of Effects across Visual Areas. Cereb Cortex 10:343–358.

724 Mehta AD, Ulbert I, Schroeder CE (2000b) Intermodal Selective Attention in Monkeys. II: Physiological
725 Mechanisms of Modulation. Cereb Cortex 10:359–370.

726 Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual Organization of Tone Sequences in the
727 Auditory Cortex of Awake Macaques. Neuron 48:139–148.

728 Miller LM, Escabí MA, Read HL, Schreiner CE (2002) Spectrotemporal Receptive Fields in the Lemniscal
729 Auditory Thalamus and Cortex. J Neurophysiol 87:516–527.

730 Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of
731 evoked potentials and EEG phenomena. Physiol Rev 65:37–100.

732 Moore BCJ, Gockel H (2002) Factors Influencing Sequential Stream Segregation. Acta Acust United Acust
733 88:320–333.

734 Moore BCJ, Gockel HE (2012) Properties of auditory stream formation. Philos Trans R Soc B Biol Sci
735 367:919–931.

736 Müller-Preuss P, Mitzdorf U (1984) Functional anatomy of the inferior colliculus and the auditory cortex:
737 current source density analyses of click-evoked potentials. Hear Res 16:133–142.

738 Nelken I, Prut Y, Vaadia E, Abeles M (1994) Population responses to multifrequency sounds in the cat
739 auditory cortex: One- and two-parameter families of sounds. Hear Res 72:206–222.

Nicholson C, Freeman JA (1975) Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. J Neurophysiol 38:356–368.

Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature 459:89–92.

Niwa M, Johnson JS, O'Connor KN, Sutter ML (2012a) Active Engagement Improves Primary Auditory Cortical Neurons' Ability to Discriminate Temporal Modulation. J Neurosci 32:9323–9334.

Niwa M, Johnson JS, O'Connor KN, Sutter ML (2012b) Activity Related to Perceptual Judgment and Action in Primary Auditory Cortex. J Neurosci 32:3193–3210.

Niwa M, Johnson JS, O'Connor KN, Sutter ML (2013) Differences between Primary Auditory Cortex and Auditory Belt Related to Encoding and Choice for AM Sounds. J Neurosci 33:8378–8395.

O'Connell MN, Barczak A, Schroeder CE, Lakatos P (2014) Layer Specific Sharpening of Frequency Tuning by Selective Attention in Primary Auditory Cortex. J Neurosci 34:16496–16508.

Oh Y, Zuwala JC, Salvagno CM, Tilbrook GA (2022) The Impact of Pitch and Timbre Cues on Auditory Grouping and Stream Segregation. Front Neurosci 15:725093.

Rezaeizadeh M, Shamma S (2021) Binding the Acoustic Features of an Auditory Source through Temporal Coherence. Cereb Cortex Commun 2(4):tgab060.

Saleem KS, Logothetis NK (2012) A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates, 2nd edition. Amsterdam: Academic Press.

Selezneva E, Gorkin A, Budinger E, Brosch M (2018) Neuronal correlates of auditory streaming in the auditory cortex of behaving monkeys. Eur J Neurosci 48:3234–3245.

Selezneva E, Gorkin A, Mylius J, Noesselt T, Scheich H, Brosch M (2012) Reaction times reflect subjective auditory perception of tone sequences in macaque monkeys. Hear Res 294:133–142.

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

Stark E, Abeles M (2007) Predicting Movement from Multiunit Activity. J Neurosci 27:8387–8394.

Steinschneider M, Tenke CE, Schroeder CE, Javitt DC, Simpson GV, Arezzo JC, Vaughan HG (1992) Cellular generators of the cortical auditory evoked potential initial component. Electroencephalogr Clin Neurophysiol Potentials Sect 84:196–200.

Supèr H, Roelfsema PR (2005) Chronic multiunit recordings in behaving animals: advantages and limitations. In: Progress in Brain Research, 147:263–282.

Sussman E, Steinschneider M (2009) Attention effects on auditory scene analysis in children. Neuropsychologia 47:771–785.

Szymanski FD, Garcia-Lazaro JA, Schnupp JWH (2009) Current Source Density Profiles of Stimulus-Specific Adaptation in Rat Auditory Cortex. J Neurophysiol 102:1483–1490.

774 Tsunada J, Cohen Y, Gold JI (2019) Post-decision processing in primate prefrontal cortex influences
775 subsequent choices on an auditory decision-making task. eLife 8:e46770.

776 Tsunada J, Liu ASK, Gold JI, Cohen YE (2016) Causal contribution of primate auditory cortex to auditory
777 perceptual decision-making. Nat Neurosci 19:135–142.

778 van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. Eindhoven, The
779 Netherlands: Institute for Perceptual Research.

780

781

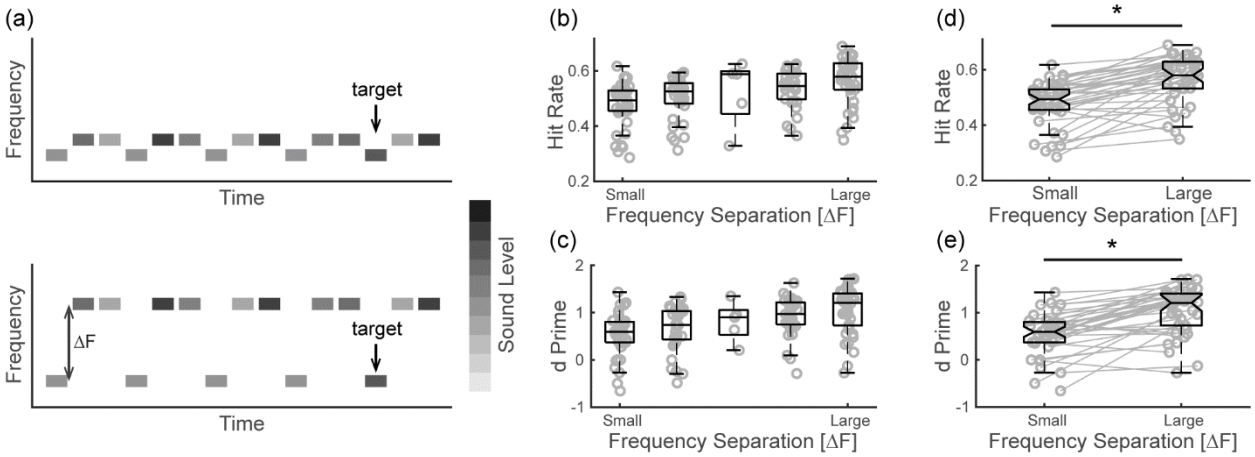

Figure 1: Auditory streaming task and behavioral performance to the task. (a) Schematic illustration of the auditory streaming task. Each rectangle indicates a low (L) and high (H) frequency tone burst presented as repeating L-H-H triplets. The grey shading of each rectangle represents the relative sound level of the tone bursts. The monkeys detected a relatively louder "deviant" target. We varied the frequency separation (ΔF) between the L and H tone bursts to titrate the task difficulty; target detection is relatively more difficult in the small ΔF trials (top) than in large ΔF trials (bottom). **(b)** Hit rate as a function of ΔF. The gray circles indicate the hit rate from each recording session, and the box plot indicates the median and upper and lower quantiles of the hit rate across sessions. **(c)** d' as a function of ΔF. The gray circles indicate the d' from each recording session, and the box plot indicates the median and upper and lower quantiles of the d' across sessions. **(d)** Comparison of hit rate between small (typically, 1 and 4 semitone difference for monkey D and C, respectively) and large ΔF (typically, 8 and 24 semitone difference for monkey D and C, respectively) trials. **(e)** Comparison of d' between small and large ΔF trials. In **(d)** and **(e)**, the asterisk indicates significant difference in hit rate and d', respectively between the small and large ΔF trials ($p < 0.05$, Wilcoxon signed-rank test).

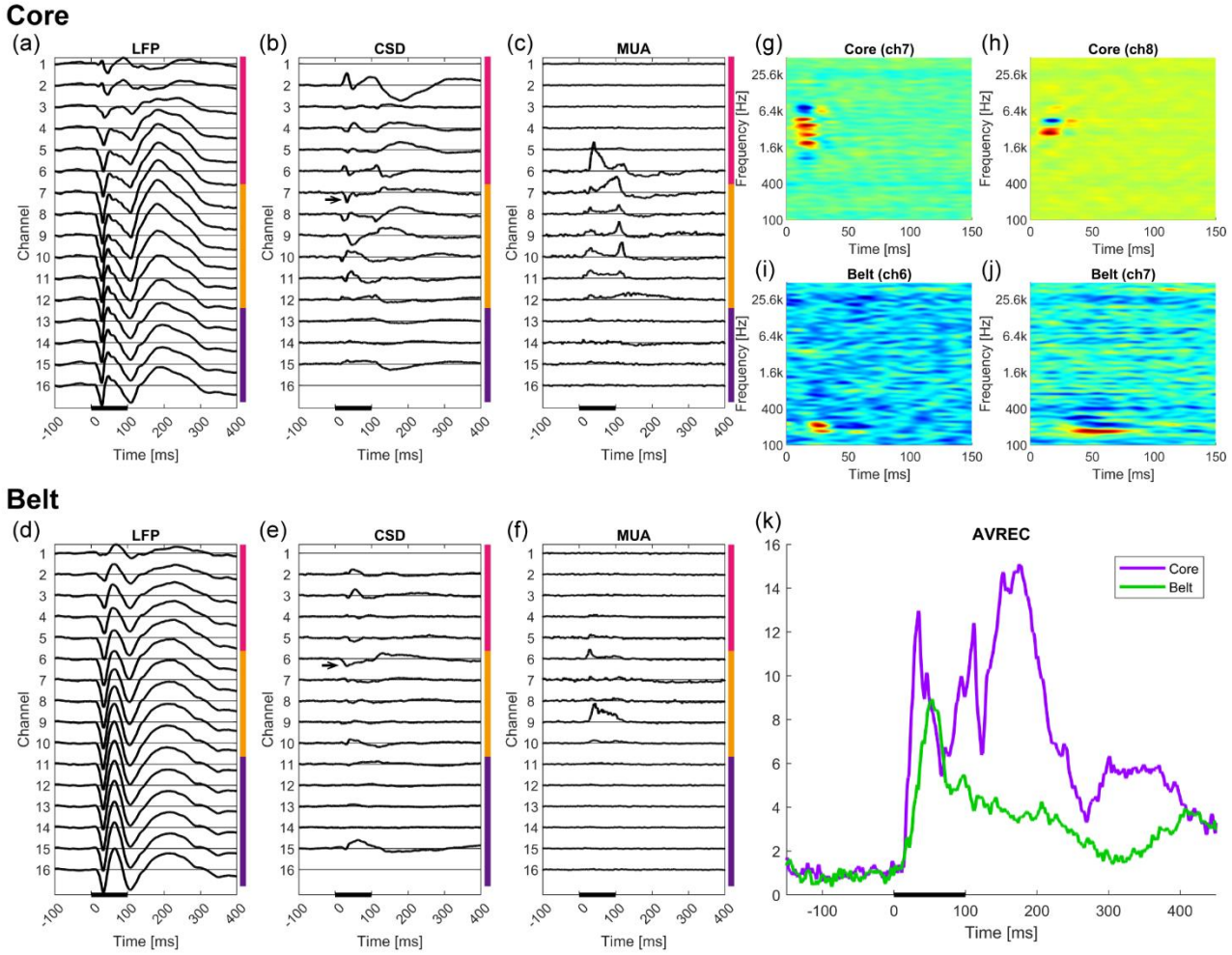

Figure 2: Examples of laminar recording from core and belt auditory cortex. (a-c) Representative laminar response profiles evoked by a Gaussian noise burst in core auditory cortex: **(a)** local field potential (LFP), **(b)** the one-dimensional current source density (CSD), and **(c)** multiunit activity (MUA) are shown as a function of cortical depth from superficial (top) to deep (bottom). The spacing between channels = 150 µm. The arrow in **(b)** indicates the initial current sink corresponding to the bottom of layer 3. The vertical color bar indicates the approximate locations of the supragranular (pink), granular (yellow), and infragranular (purple) layers that were operationally identified from the CSD profile. The thick black bar at the bottom of each plot indicates the 100-ms duration of a Gaussian noise burst. **(d-f)** Representative laminar response profiles from belt auditory cortex presented in the same format as in **(a-c)**. **(g-j)** Representative spectrotemporal receptive fields (STRFs) obtained from the middle channels of the representative electrodes shown in **(a-f)**. The x-axis is aligned relative to stimulus onset. Hotter/cooler colors indicate increased/decreased firing rates, respectively. **(k)** The average rectified CSD (AVREC) from the representative core and belt sites shown in **(a-f)**.

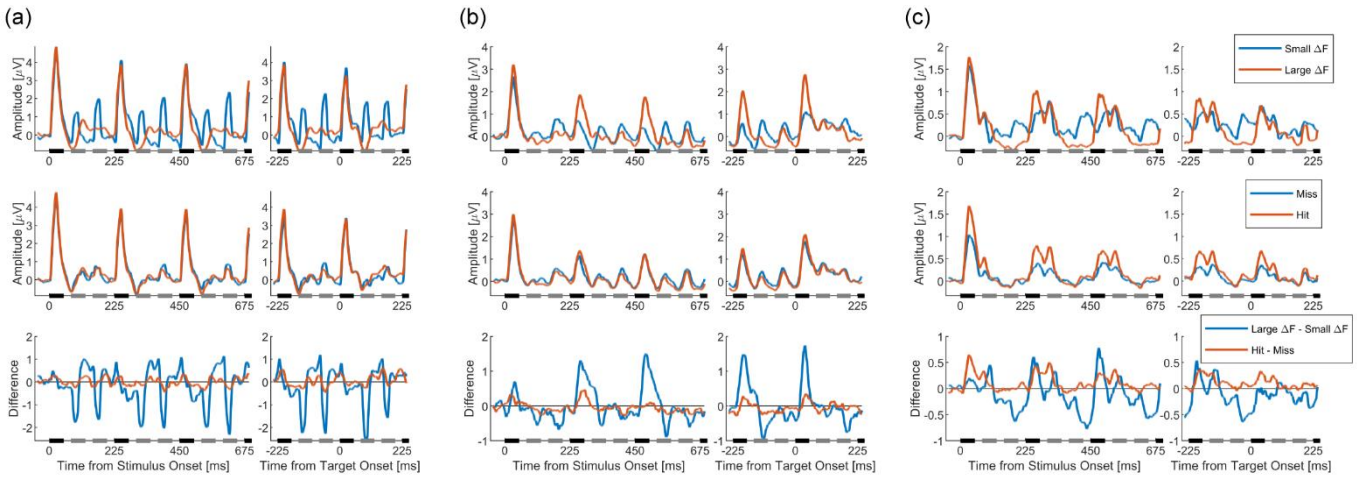

Figure 3: Examples of MUA during the auditory streaming task. (a) A MUA response profile from the core auditory cortex. **(top)** The mean MUA from trials with the smallest (8 semitone difference) and largest (24 semitone difference) frequency separation. The rectangles on the horizontal axis indicate the presentation periods of low- (black) and high- (gray) frequency tone bursts. **(middle)** The mean MUA for hit (red) and miss (blue) trials. **(bottom)** The difference in mean MUA between different stimulus (small ΔF versus large ΔF) and choice (hit versus miss) conditions. **(b)** A MUA response profile from the belt auditory cortex, organized in the same way as in **(a)**. The small and large ΔF are 1 and 8 semitone differences, respectively. **(c)** Another MUA response profile from the core auditory cortex, organized in the same way as in **(a)**. The small and large ΔF are 4 and 24 semitone differences, respectively. The MUA is smoothed for the purpose of display.

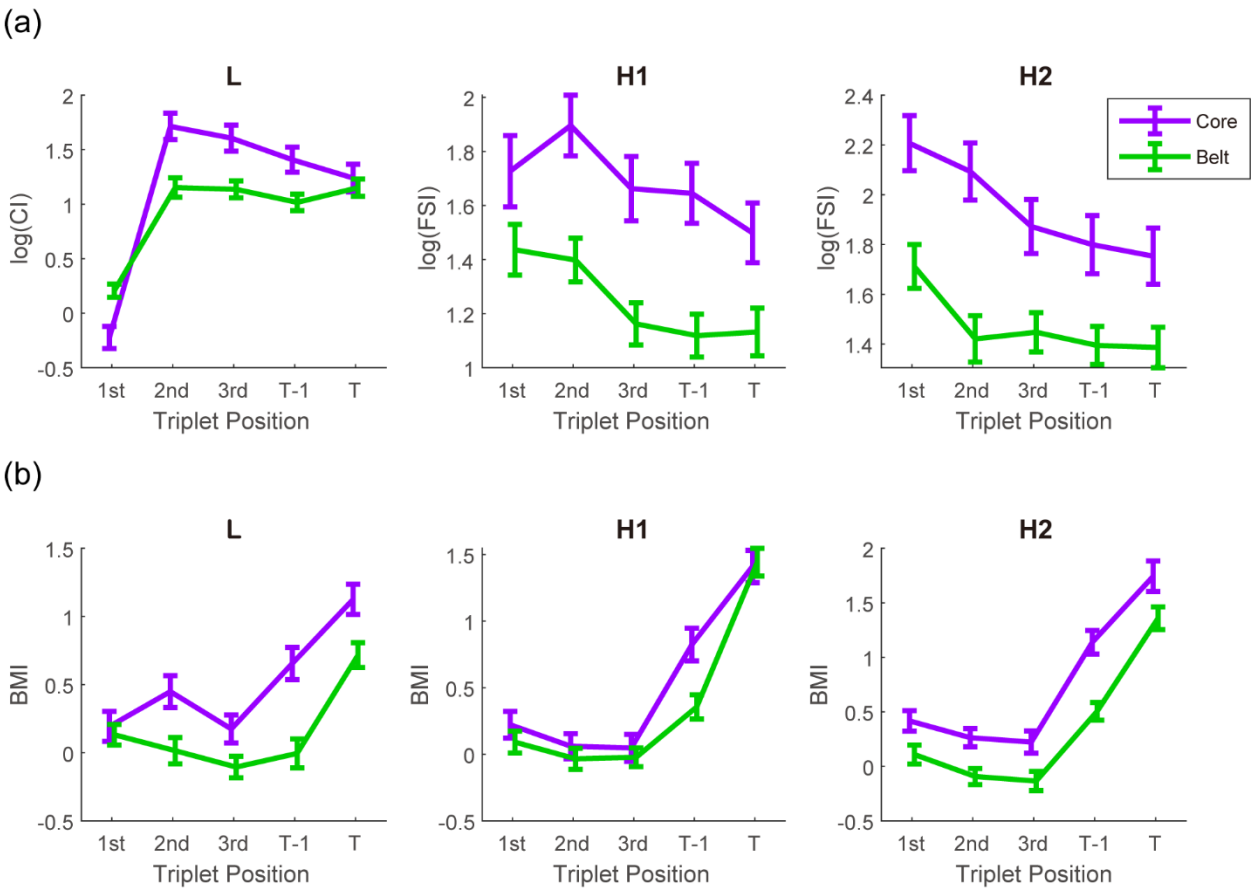

Figure 4: Population time course of stimulus- and choice-related MUA modulation, relative to the onset of the tone-burst sequence. (a) CI values as a function of time (left; only for the low-frequency [L] tone burst) and FSI values as a function of time (middle and right; for the two high-frequency tone bursts in the triplet [H1 and H2, respectively]). Data points indicate the population mean and standard error for core (purple) and belt (green) auditory cortex for the first three ("1st", "2nd", and "3rd") triplet positions and for the triplet immediately preceding the target (i.e., "T-1") and the triplet that included the target ("T"). **(b)** BMI values as a function of time for L, H1, and H2. Population data are plotted analogous to that shown in **(a)**.

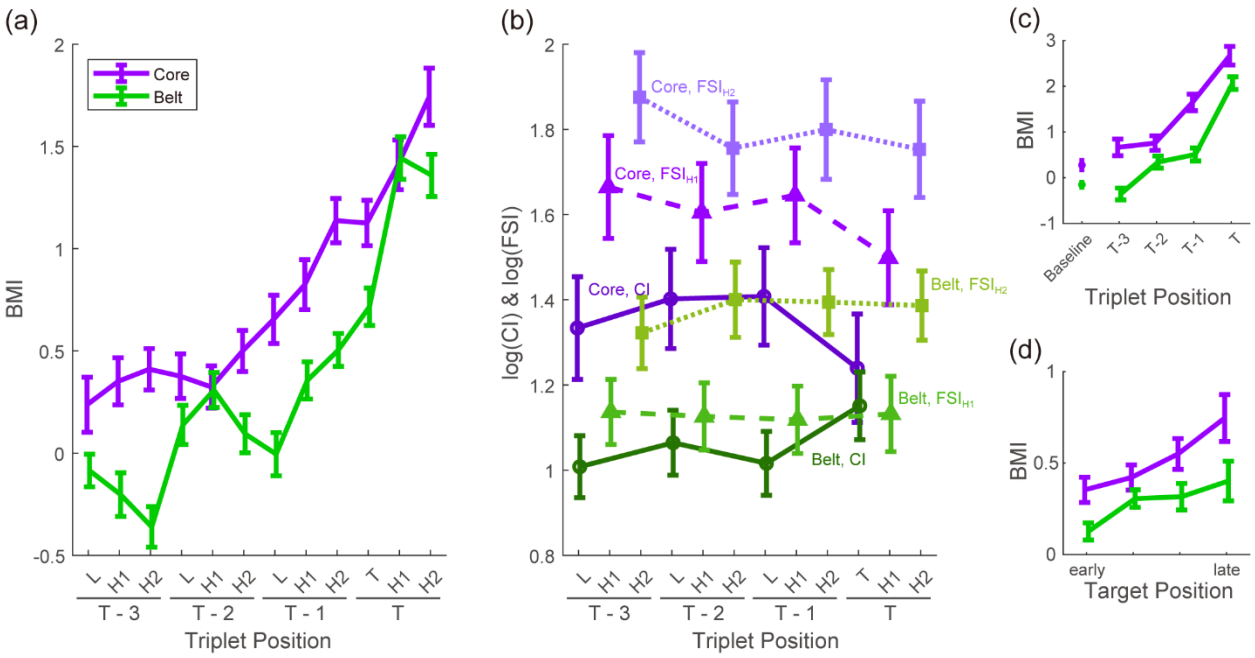

Figure 5: Stimulus- and behavior-related modulation, relative to target onset. (a) BMI values replotted as a function of target onset, starting with the three triplets preceding target onset (respectively, "T-3", "T-2", and "T-1") and the target triplet ("T"). **(b)** CI and FSI values replotted as a function of target onset in a manner analogous to that done in **(a)**. **(c)** BMI values calculated across an entire 225-ms triplet window and plotted as a function of target onset. For comparison, we plot the BMI values generated during a baseline period (3rd triplet position) as well. **(d)** BMI values for target times that occurred "early" (typically 4th and 7th triplet for monkey D and monkey C, respectively) in a tone-burst sequence versus those that occurred "late" (typically 7th and 10th triplet in monkey D and monkey C, respectively). In all panels, data points indicate the population mean and standard error.

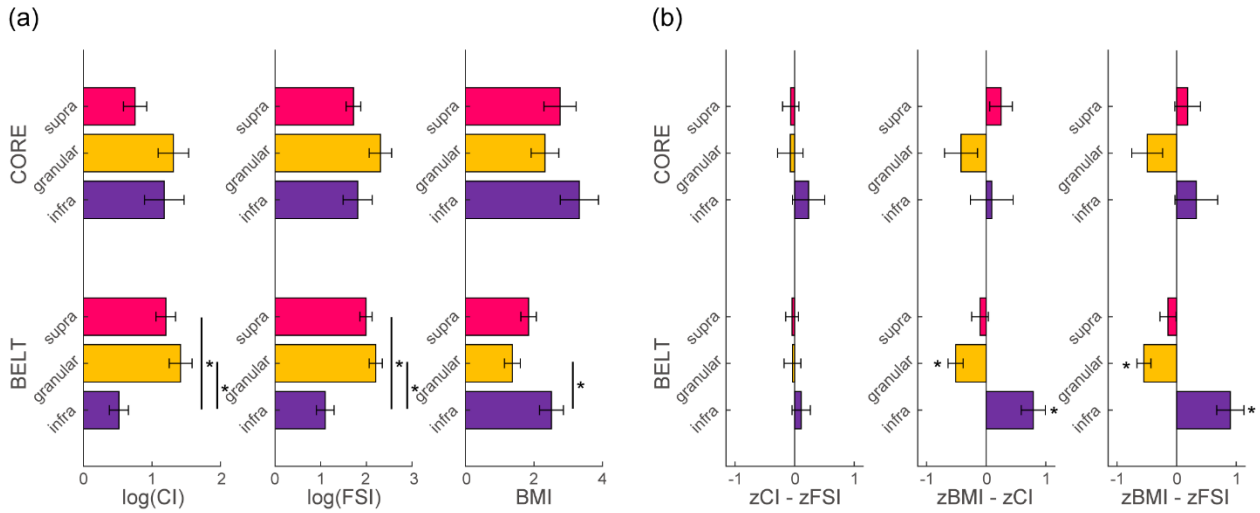

Figure 6: Comparisons of stimulus and behavior-related modulations across layers. (a) Laminar distributions of CI, FSI, and BMI values in core and belt. The mean and standard error of these indices are plotted separately for the supragranular (pink), granular (yellow) and infragranular (purple) layers. Asterisks indicate the laminar differences with statistically significant ($p < 0.05$; Dunn test) median values. **(b)** Laminar distributions of z-scored CI, FSI and BMI values in core and belt. Bar graphs indicate mean and standard error. The asterisks indicate values that are significantly different from zero ($p < 0.05$; Wilcoxon signed-rank test with Bonferroni correction). For both panels, the CI and FSI were calculated from the triplet immediately preceding the target, whereas the BMI was calculated from the response in the triplet that included the target.