Behavioral/Systems/Cognitive

Neural Correlates of Auditory Scene Analysis Based on Inharmonicity in Monkey Primary Auditory Cortex

Yonatan I. Fishman[1] and Mitchell Steinschneider[1,2]

Departments of [1]Neurology and [2]Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461

Segregation of concurrent sounds in complex acoustic environments is a fundamental feature of auditory scene analysis. A powerful cue used by the auditory system to segregate concurrent sounds, such as speakers' voices at a cocktail party, is inharmonicity. This can be demonstrated when a component of a harmonic complex tone is perceived as a separate tone "popping out" from the complex as a whole when it is sufficiently mistuned from its harmonic value. The neural bases of perceptual "pop out" of mistuned harmonics are unclear. We recorded multiunit activity from primary auditory cortex (A1) of behaving monkeys elicited by harmonic complex tones that were either "in tune" or that contained a mistuned third harmonic set at the best frequency of the neural populations. Responses to mistuned sounds were enhanced relative to responses to "in-tune" sounds, thus correlating with the enhanced perceptual salience of the mistuned component. Consistent with human psychophysics of "pop out," response enhancements increased with the degree of mistuning, were maximal for neural populations tuned to the frequency of the mistuned component, and were not observed under comparable stimulus conditions that do not elicit perceptual "pop out." Mistuning was also associated with changes in neuronal temporal response patterns phase locked to "beats" in the stimuli. Intracortical auditory evoked potentials paralleled noninvasive neurophysiological correlates of perceptual "pop out" in humans, further augmenting the translational relevance of the results. Findings suggest two complementary neural mechanisms for "pop out," based on the detection of local differences in activation level or coherence of temporal response patterns across A1.

Introduction

In everyday life, sounds generated by multiple sources impinge upon our ears simultaneously (e.g., speakers' voices at a cocktail party). A fundamental task of the auditory system is auditory scene analysis: to parse the acoustic input and form ecologically meaningful representations of sound sources in the environment (Bregman, 1990; Yost, 1991). Many natural sounds, such as animal vocalizations and human speech, display harmonic structure, containing spectral components that are integer multiples of a common fundamental frequency (F0). As it is unlikely that simultaneous acoustic components produced by two independent sources would be harmonically related, inharmonicity provides a reliable cue for segregating concurrent sounds in the environment. This can be demonstrated by mistuning a single harmonic (or partial) of a harmonic complex tone (HCT), so that it is no longer an integer multiple of the F0. When the mistuning exceeds ~3% of its harmonic frequency, the mistuned component is heard as a separate tone, "popping out" perceptually from

the HCT as a whole (Moore et al., 1986; Hartmann et al., 1990; Alain et al., 2001, 2002). Thus, whereas an "in-tune" HCT is typically perceived as a single sound source with a global pitch at the F0, an HCT with a sufficiently mistuned harmonic is perceived as two separate sound sources, one with a pitch at the F0 and the other with a pitch at the frequency of the mistuned harmonic.

Neural mechanisms underlying concurrent sound segregation based on inharmonicity are poorly understood. Noninvasive studies in humans have compared neuroelectric and neuromagnetic responses evoked by "in-tune" HCTs and by HCTs with a mistuned harmonic that elicit a perceptual "pop out" of the mistuned harmonic and the perception of two sound sources (Alain et al., 2001, 2002, Lipp et al., 2010). Prominent difference-waveform components associated with perceptual "pop out" include a negative deflection, referred to as the "object-related negativity" (ORN), and a subsequent positive deflection (P230), that peak at latencies of approximately 150 and 230 ms after stimulus onset, respectively. Both components are observed under nonattending listening conditions, thus suggesting a preattentive neural basis for the perceptual phenomenon. While these components are thought to reflect neural activity within primary auditory cortex (A1) (Alain et al., 2001, 2002; Lipp et al., 2010), noninvasive studies are limited in their ability to localize and characterize neural processes underlying their generation.

A reasonable hypothesis is that the enhanced perceptual salience of a mistuned harmonic is associated with an increase in the firing rate or a change in the temporal response pattern of local neural populations within A1 that are tuned to the fre-

quency of the mistuned harmonic (Sinex et al., 2002; Fishman and Steinschneider, 2010). Here we test this hypothesis in macaque A1 using population measures that bridge the gap between single-neuron recordings in experimental animals and noninvasive recordings in humans. Our results parallel both psychophysical and noninvasive neurophysiological findings related to perceptual "pop out" in humans, and thus support a role for A1 in concurrent sound segregation based on inharmonicity.

Materials and Methods

Three adult male macaque monkeys (*Macaca fascicularis*) were studied using previously described methods (Fishman et al., 2001a,b; Steinschneider et al., 2003; Fishman and Steinschneider, 2009). Animals were housed in our AAALAC-accredited Animal Institute under daily supervision of laboratory and veterinary staff. All experimental procedures were reviewed and approved by the AAALAC-accredited Animal Institute of Albert Einstein College of Medicine and were conducted in accordance with institutional and federal guidelines governing the experimental use of primates. To minimize the number of monkeys used, other auditory experiments were conducted in the same animals during each recording session. Before surgery, animals were acclimated to the recording environment and trained while sitting in custom-fitted primate chairs.

Surgical procedure. Under pentobarbital anesthesia and using aseptic techniques, holes were drilled bilaterally into the dorsal skull to accommodate matrices composed of 18 gauge stainless steel tubes glued together in parallel. Tubes served to guide electrodes toward A1 for repeated intracortical recordings. Matrices were stereotaxically positioned to target A1. They were oriented at a 30° anterior–posterior angle and with a slight medial–lateral tilt to direct electrode penetrations perpendicular to the superior surface of the superior temporal gyrus, thereby satisfying one of the major technical requirements of one-dimensional current source density analysis (Vaughan and Arezzo, 1988). Matrices and Plexiglas bars, used for painless head fixation during the recordings, were embedded in a pedestal of dental acrylic secured to the skull with inverted bone screws. Perioperative and postoperative antibiotic and anti-inflammatory medications were always administered. Recordings began after a 2 week postoperative recovery period.

Neurophysiological recordings. Recordings were conducted in an electrically shielded, sound-attenuated chamber. During the recordings monkeys successfully performed a simple auditory discrimination task (detection of a randomly presented noise burst interspersed with test stimuli) to ensure attention to the sounds.

Intracortical recordings were performed using linear-array multicontact electrodes containing 16 contacts, evenly spaced at 150 μm ($\pm$10%) intervals (Neurotrack). Individual contacts were maintained at an impedance of $\sim$200 kΩ. An epidural stainless-steel screw placed over the occipital cortex served as a reference electrode. In two monkeys, neural signals were bandpass filtered from 3 Hz to 3 kHz (roll-off 48 dB/octave) and digitized at 12.2 kHz using an RA16 PA Medusa 16-channel preamplifier connected via fiber-optic cables to an RX5 data acquisition system (Tucker-Davis Technologies). Signals were averaged online by computer to yield auditory evoked potentials (AEPs). One-dimensional current source density (CSD) analyses characterized the laminar pattern of net current sources and sinks within A1 generating the AEPs and were used to identify the laminar location of concurrently recorded multiunit activity (MUA). CSD was calculated using a three-point algorithm that approximates the second spatial derivative of voltage recorded at each recording contact (Freeman and Nicholson, 1975; Nicholson and Freeman, 1975). In the monkey tested initially, field potentials were recorded using unity-gain headstage preamplifiers, and subsequently amplified 5000 times by differential amplifiers (Grass) with a frequency response down 6 dB at 3 Hz and at 3 kHz. Signals were digitized at 3.4 kHz and averaged by computer (Neuroscan software and hardware, Neurosoft) to generate AEPs. Testing revealed that no significant aliasing in the AEPs is observed at this digitization rate, as the amplitude of signals at frequencies above 1500 Hz is negligible ($<$0.1%) compared with that of signals that dominate the AEP, which fall below 200 Hz.

To derive MUA, signals were simultaneously high-pass filtered at 500 Hz (roll-off 48 dB/octave), amplified an additional eight times, full-wave rectified, and then low-pass filtered at 520 Hz (roll-off 48 dB/octave) before digitization and averaging (for a methodological review, see Supèr and Roelfsema, 2005). MUA is a measure of the envelope of summed action potential activity of neuronal aggregates within a sphere of $\sim$100 μm in diameter surrounding each recording contact (Legatt et al., 1980; Vaughan and Arezzo, 1988; Brosch et al., 1997; Supèr and Roelfsema, 2005). Using an electrode impedance similar to that of the electrodes used in the present study (100–300 kΩ), MUA and single-unit recordings have been shown to yield similar results in primary visual cortex, with MUA reflecting the average local activity and orientation tuning of closely spaced neurons in the vicinity of the electrode (Supèr and Roelfsema, 2005). Moreover, MUA displays greater response stability than single-unit activity (Nelken et al., 1994; Stark and Abeles, 2007).

Peristimulus time histograms (PSTHs) provided a complementary and more selective measure of action potential activity. PSTHs were derived from unrectified high-pass-filtered data (same filtering parameters as for MUA) using a custom spike detection program implemented in MATLAB and with a Schmitt trigger threshold set at 4 SDs above the mean prestimulus baseline activity. To further enhance selectivity, PSTHs only included negative-going spikes.

Electrodes were positioned within the cortex using a microdrive and were guided by on-line examination of click-evoked potentials. Test stimuli were delivered when the electrode channels bracketed the inversion of early AEP components and when the largest MUA, typically occurring during the first 50 ms after stimulus onset, was situated in the middle channels. Evoked responses to 50 presentations of each stimulus were averaged with an analysis window of 500 ms (including a 100 ms prestimulus baseline interval).

At the end of the recording period, monkeys were deeply anesthetized with sodium pentobarbital and transcardially perfused with 10% buffered formalin. Tissue was sectioned in the coronal plane (80 μm thickness) and stained for Nissl substance to reconstruct the electrode tracks and to identify A1 according to previously published physiological and histological criteria (Merzenich and Brugge, 1973; Morel et al., 1993; Hackett et al., 1998). Based upon these criteria, all electrode penetrations considered in this report were localized to A1. However, the possibility that some sites situated near the lower-frequency anterolateral border of A1 were located in the rostral field R cannot be excluded.

Stimuli. Stimuli were generated and delivered at a sample rate of 48.8 kHz by a PC-based system using RP2 or RX8 modules (Tucker Davis Technologies). Frequency response functions (FRFs) based on pure tone responses characterized the spectral tuning of the cortical sites. Pure tones used to generate the FRFs ranged from 0.15 to 18.0 kHz. Pure tone and complex tone stimuli were 200 and 250 ms in duration, respectively (including 10 ms linear rise/fall ramps), and were presented with a stimulus onset-to-onset interval of 658 and 992 ms, respectively. Resolution of FRFs was 0.25 octaves or finer across the 0.15 to 18.0 kHz frequency range tested. Stimuli were presented via a free-field speaker (Microsatellite; Gallo) mounted at a contralateral azimuth of 60° on a semicircular speaker array located 1 m away from the animal's head (Crist Instrument). Pure tone stimuli were presented at 60 dB SPL. Complex tone stimuli were comprised of 10 equal-amplitude components added in sine phase, and presented at an overall level of 65 dB SPL (55 dB SPL/component). Sound intensity was measured with a sound level meter (type 2236; Bruel and Kjaer) positioned at the location of the animal's ear. The frequency response of the speaker was essentially flat (within $\pm$5 dB SPL) over the frequency range tested.

Complex sounds conformed to stimuli used in human psychoacoustic studies examining the perceptual segregation of partials based on inharmonicity (Moore et al., 1985, 1986; Hartmann et al., 1990; Alain et al., 2001, 2002; Roberts and Brunstrom, 2001; Alain and McDonald, 2007; Lipp et al., 2010). Complex sounds were presented in four separate blocks, each consisting of different sets of stimuli. Blocks and stimuli within blocks were presented in random order. Spectra of the complex stimuli are schematically represented in Figure 1. All four stimulus blocks included the same "in-tune" HCT (referred to as the "harmonic" condition) with the third harmonic set equal to the best frequency (BF; defined
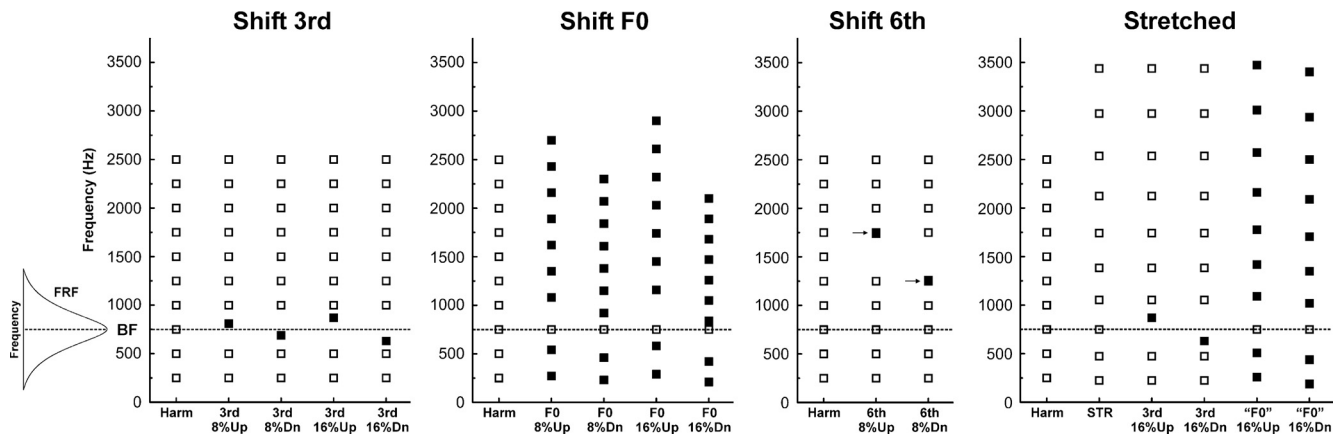
Figure 1. Spectra of complex sound stimuli presented in the study. Frequency and stimulus conditions are represented along the ordinate and abscissa, respectively. Spectral components of the complex sounds are schematically represented by square symbols. Stimuli were presented in four blocks: "shift 3rd," "shift F0," "shift 6th," and "stretched." Each block consisted of an "in-tune" harmonic condition and several mistuned conditions in which stimuli were made inharmonic via various manipulations. Stimuli were designed so that the third component under the harmonic condition (Harm) was set equal to the BF of the site (the peak of its FRF, schematically shown at the left of the figure; in the case depicted, the BF and third harmonic = 750 Hz, indicated by the dashed horizontal line, and the F0 = 250 Hz). Components shifted upward (Up) or downward (Dn) relative to their frequency under the harmonic condition are represented by filled symbols. Arrows in the "shift 6th" panel indicate stimulus components that are visually occluded by the adjacent mistuned components due to their close proximity in frequency.

below) of the recorded neural populations. Other complex stimuli were characterized by modifications to the harmonic condition that, in particular cases, elicit the perceptual "pop out" of a mistuned component and the perception of two distinct sound sources in human listeners (collectively referred to as "mistuned" conditions).

In the "shift 3rd" block, the third component was mistuned by shifting it upward or downward and away from the BF by 8% or 16% of its harmonic value. In the "shift F0" block, the third harmonic was fixed at the BF and mistuned relative to the other partials by jointly shifting the F0 and the other harmonics upward or downward by 8% or 16% of their value under the harmonic condition. Both conditions elicit a perceptual "pop out" of the mistuned component and the corresponding perception of two separate sound sources in human listeners, which is more salient when components are mistuned by 16% than when they are mistuned by 8% (Alain et al., 2001).

Three additional stimulus blocks served as controls. In the "shift 6th" block, the sixth component was mistuned upward or downward by 16% of its harmonic value while the third harmonic was held fixed at the BF of the recording site (Fig. 1). The rationale for this stimulus block was to test whether mistuning-related response enhancements were generalized across A1 or largely restricted to neurons tuned to the frequency of the mistuned component. Only in the latter case would response enhancements potentially qualify as neural correlates of perceptual "pop out" due to inharmonicity. The "stretched" stimulus block (STR) consisted of complex tones made globally inharmonic by applying a 12% cumulative stretch to the components (Roberts and Brunstrom, 2001). As the components of 12% stretched stimuli are not harmonically related to each other, the sounds are perceived as less fused and more fragmented than their harmonic counterparts. As in the "shift 3rd" and "shift F0" stimulus blocks, the third component was either shifted upward or downward in frequency by 16% or the other components were jointly shifted upward or downward in frequency by 16%, while the frequency of the third component remained fixed at the BF. Perceptual "pop out" of a mistuned component in spectrally stretched stimuli depends upon the degree of spectral stretching. While the ability to detect a mistuned component in 3% stretched and in nonstretched harmonic complexes is comparable, under the 12% stretched conditions tested in the present study, listeners generally do not hear a single component "popping out" from the complex sound as a whole, and the ability to detect individual mistuned components is poorer than under the "shift 3rd," "shift F0," and "shift 6th" conditions (Roberts and Brunstrom, 2001). Thus, the rationale for the "stretched" condition was to test whether mistuning-related response modulations reflect inharmonicity in general or correlate specifically with the perceptual "pop out" of a mistuned component. "Shift 3rd" and "shift F0" stimulus blocks were presented to all three animals (46 elec-

trode penetrations total), while "shift 6th" and "stretched" stimulus blocks were presented to only one animal (22 electrode penetrations total), as interim statistical analyses indicated that the inclusion of additional animals would not have significantly altered results obtained under these stimulus conditions.

The third set of control stimuli was presented to one of the three animals (16 electrode penetrations total) to further test the hypothesis that mistuning-related response enhancements are specific to neurons tuned to the frequency of the mistuned partial. As in the "shift 3rd" stimulus blocks, this stimulus set included HCTs with a third harmonic that was either "in tune" or mistuned upward or downward by 16% of its harmonic value. However, in this case the F0s of the HCTs were fixed at 125, 250, 500, 750, and 1000 Hz, regardless of the BF of the recorded neural population (such stimuli are referred to as "fixed F0" stimuli). Thus, the difference in frequency between the third harmonic and the BF of the recorded neural populations varied across stimuli and across recording sites. If mistuning-related response enhancements are tonotopically specific, then they should be maximal when the frequency of the mistuned component is near the BF and minimal when it is far away from the BF (see below for details). All of these "fixed F0" stimuli were presented at each of the 16 sites examined, regardless of whether or not their spectral components overlapped the peak of the FRF.

General data analysis. MUA data are derived primarily from the spiking activity of neural ensembles recorded within lower lamina 3 (LL3), as identified by the presence of large amplitude initial current sinks that are balanced by concurrent superficial sources in upper lamina 3 (Steinschneider et al., 1992; Fishman et al., 2000a,b). Previous studies have localized the initial sinks to thalamorecipient zone layers of A1 (Müller-Preuss and Mitzdorf, 1984; Steinschneider et al., 1992; Sukov and Barth, 1998; Metherate and Cruikshank, 1999). MUA recorded from two superficial electrode channels immediately above the channel located in LL3 was analyzed separately to examine whether correlates of perceptual "pop out" are reflected also in neural activity within more supragranular portions of lamina 3. These two supragranular channels are referred to as SG1 and SG2, respectively. PSTHs were derived from action potential activity recorded in LL3.

The BF of each cortical site was defined as the pure tone frequency eliciting the maximal MUA within LL3 integrated within a time window of 10–75 ms after stimulus onset. For all sites, the BF derived from activity within this time window differed by less than a quarter octave from that based on MUA integrated within a time window of 10–200 ms, which includes responses occurring over the entire duration of the pure tone stimuli. MUA occurring in the two superficial adjacent electrode channels (a 450 μm laminar extent) located above the initial current sink was included in analyses only if it displayed a BF that was within a quarter

octave of the BF based on MUA recorded in LL3. This was the case for all electrode penetrations considered in this report.

As perceptual detection of mistuned components within HCTs tends to improve with increasing stimulus duration (Moore et al., 1986), separate analyses were performed for neural responses occurring within the following earlier and later time windows: 10–75 ms and 75–250 ms, which include the "on" and "sustained" portions of the responses, respectively (see Fishman and Steinschneider, 2009). A third time window extending from 10 to 250 ms, which includes the total response ("total"), was also analyzed.

Analysis of mistuning-related response enhancements. Responses to the complex sounds occurring within each of the three time windows were analyzed as follows. For each electrode penetration, responses were normalized to the amplitude of the maximal response evoked by the complex stimuli within a given stimulus block (i.e., "shift 3rd," "shift F0," "shift 6th," and "stretched"). Normalized response amplitudes were then averaged across electrode penetrations for a given stimulus block. Thus, if responses to mistuned HCTs are enhanced relative to responses to "in-tune" HCTs, then mean normalized response amplitudes under the mistuned conditions should be greater than those under the harmonic condition.

Normalized responses under the different stimulus conditions were compared via planned statistical analyses (one-tailed, unpaired *t* test) to test main predictions of the general hypothesis. Planned comparisons included 8% and 16% shifts versus the harmonic condition, with the prediction that HCTs with a mistuned component will evoke larger responses than the "in-tune" HCT, and 8% versus 16% shifts, with the prediction that 16% shifts will yield larger responses than 8% shifts, thus paralleling the enhanced perceptual salience of the mistuned component with greater degrees of mistuning. Additional planned comparisons included "shift 6th" versus the harmonic condition, with the prediction that shifting the sixth harmonic will produce diminished response enhancements relative to those observed when the mistuned harmonic is set equal to the BF of the recording site, as in the "shift F0" conditions. To evaluate whether mistuning-related response enhancements under the "shift F0" condition were larger than those under "shift 6th" conditions, for each condition we calculated Cohen's *d* (Cohen, 1992; Rosnow and Rosenthal, 1996), a commonly used standardized index of effect size, defined here as the difference between mean normalized responses under mistuned and harmonic conditions divided by their pooled SD. Finally, we predicted that if response enhancement in A1 reflects the perceptual "pop out" of a single mistuned component, rather than inharmonicity in general, then responses to "stretched" stimuli (which do not elicit a clear perceptual "pop out" of a mistuned component) should not be significantly enhanced relative to responses under the harmonic condition.

Responses to "fixed F0" stimuli were analyzed as follows. For each site, and for each "fixed F0" stimulus condition, the MUA amplitude averaged over the "total" response window under the harmonic condition was subtracted from that under each of the mistuned conditions. The result of this subtraction represents the change in MUA amplitude associated with mistuning relative to the harmonic condition. To test whether mistuning-related response enhancements are tonotopically specific, the mean change in MUA amplitude when the frequency of the third harmonic, before mistuning, was far away from the BF of the recorded neural population (greater than half an octave above or below the BF) was compared with that when the frequency of the third harmonic was equal to the BF—i.e., under the "shift F0" condition. If mistuning-related response enhancements are tonotopically specific, then the mean change in amplitude should be significantly larger when the frequency of the third harmonic is equal to the BF (under the "shift F0" condition) than when it is far away from the BF.

Analysis of mistuning-related changes in temporal response patterns. Changes in temporal response patterns associated with mistuning were quantified by computing the Pearson correlation between the waveforms of MUA evoked under the mistuned and harmonic conditions, based on the rationale that a difference in temporal response pattern will yield a lower correlation coefficient than a similar temporal response pattern. The mean Pearson correlation coefficient obtained between responses to mistuned and harmonic stimuli was compared with that obtained be-
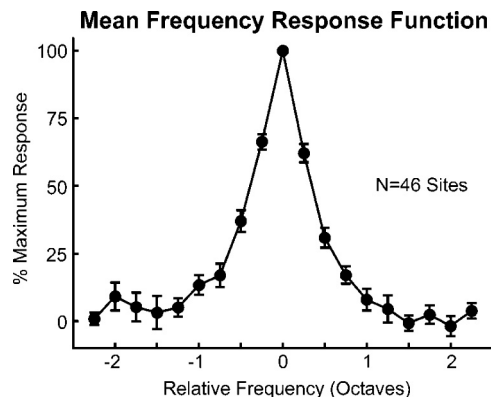

Figure 2. Mean FRF of LL3 MUA averaged across the 46 sites in A1 examined in the study. At each site, MUA evoked by tones was averaged within a time window of 10–250 ms after stimulus onset and then normalized to the amplitude of the largest tone-evoked response. Normalized FRF values were then binned in quarter-octave steps above and below the BF at each site and averaged across sites. Error bars represent SEM.

tween responses to the same harmonic stimulus presented in each of the stimulus blocks. This latter mean correlation coefficient provides a "baseline" measure both of the variability in temporal response patterns evoked by the same stimulus (harmonic) across stimulus blocks and of correlations due to noise. The correlation between responses to the harmonic stimulus presented in the "shift 3rd" and "shift F0" ("shift 6th" and "stretched") stimulus blocks served as the "baseline" in statistical analyses of responses to stimuli comprising the "shift 3rd" and "shift F0" ("shift 6th" and "stretched") stimulus blocks. Correlation coefficients obtained under each stimulus condition were then averaged across electrode penetrations and compared to the "baseline" correlation coefficient via planned, one-tailed paired *t* tests. Separate correlation analyses were performed for each of the three response windows examined. Pearson correlation coefficients were transformed to Fisher's Z_r values before all statistical comparisons (Guilford, 1965).

To test whether mistuning-related changes in temporal response patterns are tonotopically specific, a similar correlation analysis was performed between responses to harmonic and mistuned sounds comprising the "fixed F0" stimulus set (where the distance between the frequency of the third harmonic and the BF varied with F0 and with recording site). The mean correlation coefficient obtained when the frequency of the third harmonic, before mistuning, was equal to the BF (i.e., under the "shift 3rd" condition) was compared with that when the frequency of the third harmonic was near the BF (between 0.25 and 0.5 octaves away) or far away from the BF (between 0.5 and 1 octave away). If mistuning-related changes in temporal response patterns are tonotopically specific, then the mean correlation coefficient should be significantly smaller (indicating a greater difference in temporal response pattern) when the frequency of the third harmonic is equal to or near the BF than when it is far away from the BF. Planned, one-tailed unpaired *t* tests were used to compare (Z-transformed) correlation coefficients (averaged across direction of mistuning) under these three conditions.

Analysis of intracortical AEPs. Potential intracortical homologs of the ORN and P230 recorded noninvasively in humans were examined by subtracting the AEP under the harmonic condition from that evoked under each of the mistuned conditions. Differences between the harmonic and mistuned conditions across sites were evaluated by paired *t* tests performed at each time point in the AEP waveforms. AEP analyses focused on three electrode channels located in lower lamina 3, mid-upper lamina 3, and lamina 1/lamina 2, as physiologically identified by the location of the initial current sink in lower lamina 3 (LL3 sink), the slightly later supragranular sink (SG sink), and the concurrent, more superficial, supragranular source (SG source), respectively, within the CSD profile (see Fig. 3). AEP data were excluded from analysis if one or more of the three aforementioned CSD response components were absent (this usually occurred for electrode penetrations placed too deep within the cortex to record the SG source component). The rationale for
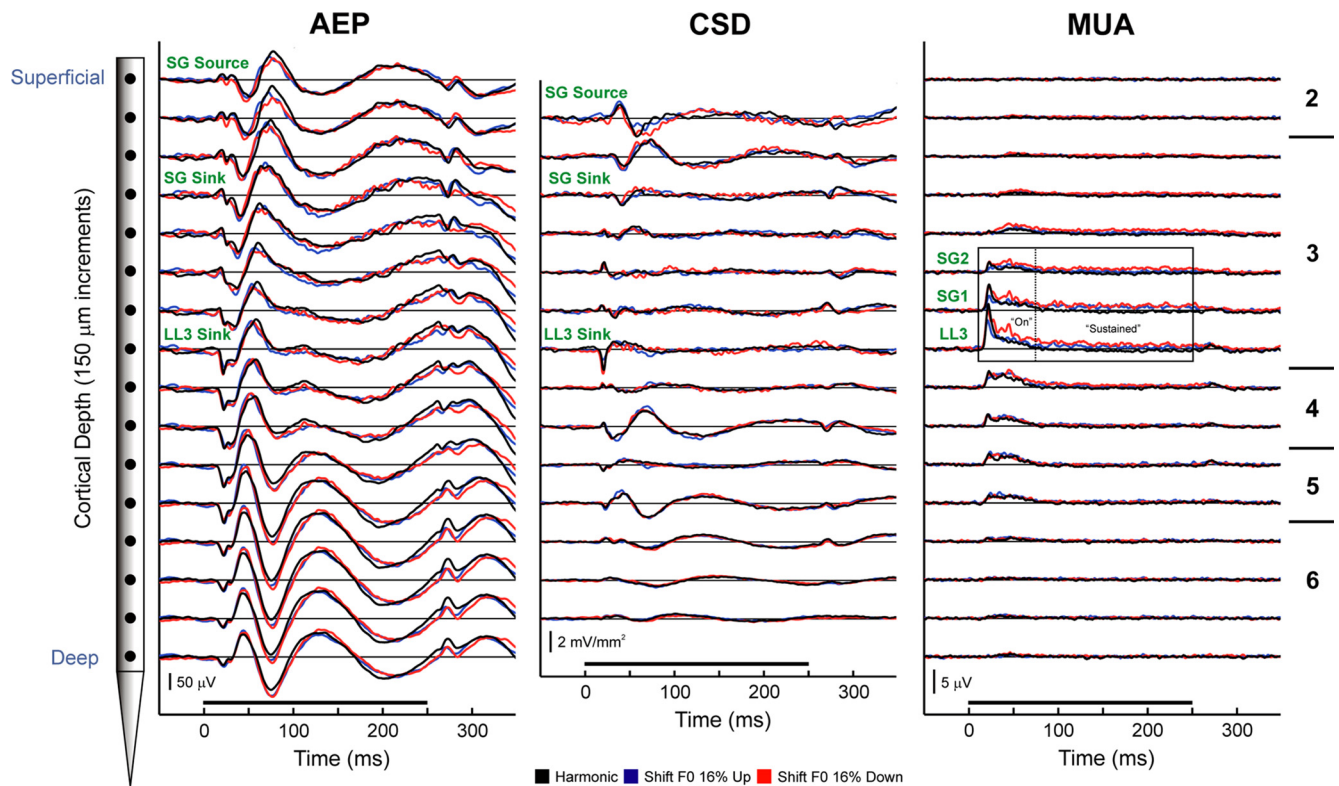
Figure 3. Representative laminar response profiles evoked by harmonic and mistuned stimuli in A1. Responses evoked by harmonic, "shift F0" 16% up, and "shift F0" 16% down stimuli are plotted in black, blue, and red, respectively. AEPs (left column) and MUA (right column) are recorded by a multicontact electrode that enables sampling of activity at 16 laminar depths simultaneously in each electrode penetration (schematic of the electrode is shown on the left; intercontact distance = 150 μm). Approximate laminar boundaries are indicated on the right of the figure. One-dimensional CSD profiles (center column) are derived from the AEP profiles. The frequency of the third harmonic remained fixed at the BF of the site (1000 Hz) under all stimulus conditions. Duration of stimuli is represented by the black horizontal bar above the time axes. Calibration bars indicate response amplitudes. Response components examined in the study are labeled in green. MUA was examined at three depths sampled by electrode contacts positioned within lower lamina 3 (LL3) and at two adjacent supragranular locations (SG1 and SG2). At each of these depths, MUA was analyzed within three time windows, which included the "on" (10–75 ms), "sustained" (75–250 ms), and "total" (10–250 ms) portions of the response, as enclosed by the rectangles superimposed on the MUA waveforms. PSTHs based on multiunit spike activity within LL3 were also analyzed.

examining the AEP at these three depths was to ascertain the laminar origins of differences in AEPs evoked under the harmonic and mistuned stimulus conditions and thereby identify potential homologs of the ORN and P230 recorded in humans. The SG source AEP selected for analysis was recorded in the channel immediately above that corresponding to the location of the SG source in the CSD profile. AEPs recorded at this laminar location are likely to dominate volume-conducted activity seen at the cortical surface (Arezzo et al., 1986; Steinschneider et al., 1992; Yvert et al., 2005) and may therefore contribute to the ORN recorded from the scalp in humans (Alain et al., 2001, 2002).

For all statistical analyses, differences between conditions yielding p values <0.05 were considered significant. Given that t tests were planned, we did not correct for multiple statistical comparisons. However, given the p values obtained, the main conclusions of the study would not have changed appreciably even if such corrections were made. No correction for multiple comparisons was made for paired t tests performed at each time point of the AEP data (the degree of correction in this case would be proportional to the sample rate of the AEP waveforms). However, several additional features of the data were considered (see Results and Discussion) to demonstrate that prominent statistically significant differences between AEPs evoked under the various stimulus conditions were genuine and not simply due to random statistical fluctuations.

Results
Results are based on multiunit responses evoked by "in-tune" and by mistuned HCTs recorded in a total of 46 multicontact electrode penetrations into A1 of three behaving macaque monkeys. A1 sites displaying multipeaked tuning were uncommon (five sites in the three animals) and are not considered in this

report. All sites displayed a clear BF and sharp frequency tuning characteristic of small neural populations in A1 (Fishman and Steinschneider, 2009) (Fig. 2). Mean onset latency and mean 6 dB bandwidths of FRFs of MUA recorded in lower lamina 3 evoked by tones presented at 60 dB SPL were ~16 ms and ~0.5 octaves, respectively. These values are comparable to those reported for single neurons in A1 of awake monkeys (Recanzone et al., 2000). Sites included in the study were limited to lower-frequency regions of A1 and had BFs ranging from 150 to 4000 Hz. These frequencies were deemed appropriate for the testing of the general hypothesis, as they allow the use of test stimuli with F0s lying within the range associated with the perception of complex tone pitch in human listeners (Plack and Oxenham, 2005).

Laminar response profiles
AEP, CSD, and MUA laminar response profiles evoked by harmonic and mistuned stimuli at a representative site are shown in Figure 3 to illustrate characteristic A1 activity patterns across cortical layers and to indicate specific response components analyzed in the study. In this example, the third harmonic was mistuned by fixing it at the BF (1000 Hz) and shifting the F0 and the other components upward or downward by 16% of their value under the harmonic condition ("shift F0" condition). Waveforms of responses evoked by stimuli in which components are shifted upward and downward in frequency relative to the harmonic condition are plotted in blue and red, respectively, whereas waveforms of responses evoked under the har-
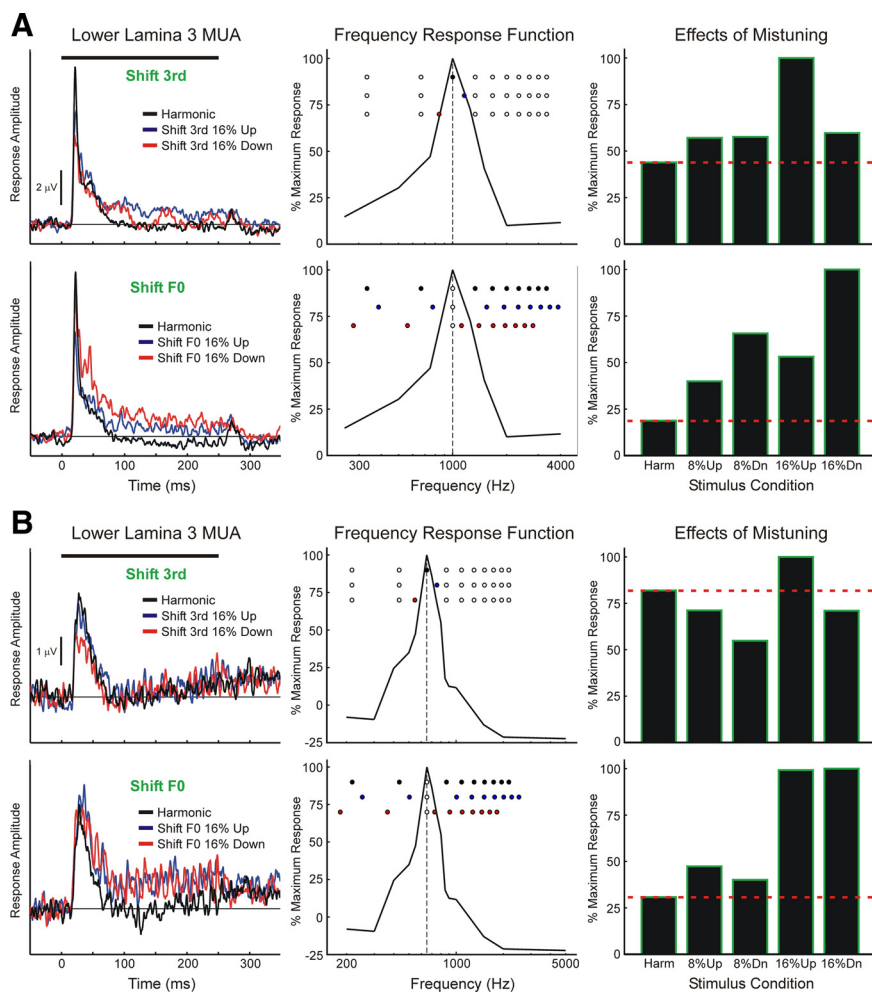
Figure 4. Lower lamina 3 MUA evoked at two sites (**A**, **B**) by harmonic and mistuned stimuli. MUA evoked under "shift 3rd" and "shift F0" conditions is plotted in the top and bottom row of each panel, as indicated. Responses to harmonic stimuli are plotted in black; responses to stimuli mistuned via upward and downward shifts of 16% are plotted in blue and red, respectively (responses to stimuli with 8% shifts are omitted for clarity). Stimulus duration is represented by the black horizontal bar above the waveforms. FRFs of the sites (based on MUA integrated over the "total" 10–250 ms time window) are shown in the center column and spectra of the stimuli are represented by the superimposed round symbols (top-to-bottom order: harmonic, shift 16% up, shift 16% down). Symbols filled blue and red denote components shifted upward and downward in frequency, respectively, relative to the harmonic condition. The dashed vertical line indicates the BF of the site (**A**, 1000 Hz; **B**, 650 Hz). Responses at each site under each stimulus condition are quantified by averaging the MUA within the 10–250 ms time window; average MUA is then normalized to the maximal average MUA evoked within each stimulus block (bar graphs in right column; see Materials and Methods and Fig. 1 for description of stimulus blocks). The red horizontal dashed line superimposed on the graphs indicates the normalized response amplitude under the harmonic condition.

monic condition are plotted in black (a convention followed throughout this report). The AEP recorded in superficial laminae displays a series of positive and negative components that invert in polarity across middle cortical layers. The corresponding CSD profile displays an initial current sink in lower lamina 3 (LL3 sink) that is balanced by deeper and more superficial sources in middle lamina 3. A slightly later current sink in more supragranular portions of lamina 3 (SG sink) is balanced by a more superficial current source (SG source). These two current dipole configurations are characteristic features of sound-evoked activity in A1 (Steinschneider et al., 1992; Fishman et al., 2000a,b; Fishman and Steinschneider, 2006) and are consistent with sequential synaptic activation of pyramidal neuron populations located in thalamorecipient and supragranular layers of A1 (Steinschneider et al., 1992; Metherate and Cruikshank, 1999). Maximal MUA typically occurs in middle cortical laminae, in spatial and temporal correlation with the LL3 sink. Differences among the

responses are observed for all three response measures and are particularly pronounced for "sustained" MUA occurring in LL3 and in adjacent supragranular channels labeled SG1 and SG2. At this site, the amplitude of "sustained" MUA evoked by the mistuned HCTs is greater than that evoked by the "in-tune" HCT.

Responses to mistuned stimuli: individual sites

MUA recorded in LL3 at two sites is shown in Figure 4, **A** and **B**, respectively, to illustrate response patterns commonly observed under "shift 3rd" and "shift F0" conditions (left column; MUA shown in Fig. 4A is from the same electrode penetration as in Fig. 3). FRFs of the sites are shown in the center column, with stimulus components represented by round symbols superimposed to indicate their relationship to the frequency tuning of the neural populations under each condition. Bar graphs (right column) represent mean normalized MUA (10–250 ms) evoked under each stimulus condition. At the site shown in Figure 4A (BF = 1000 Hz), MUA evoked by mistuned stimuli under both "shift 3rd" and "shift F0" conditions is enhanced relative to that evoked under the harmonic condition for mistunings of both 8% and 16%. In contrast, at the site shown in Figure 4B (BF = 650 Hz), MUA is enhanced relative to that evoked under the harmonic condition only under the "shift F0" condition, whereas under the "shift 3rd" condition, responses are generally diminished relative to the response under the harmonic condition. While useful in demonstrating the sensitivity of neural population responses in A1 to small changes in the spectral composition of complex sounds, the "shift 3rd" condition entails a shift of the third harmonic away from the BF of the recorded neural population. This shift confounds the interpretation of response patterns with respect to the hypothesis tested in the study. On the other hand, the "shift F0" condition, wherein the third harmonic remains fixed at the BF while all the other components are jointly shifted relative to it, avoids this potential confound, thus facilitating interpretation of response enhancements correlating with the perceptual "pop out" of a mistuned harmonic.

Importantly, the enhanced responses to mistuned stimuli under the "shift F0" condition cannot be simply explained by shifts of harmonics into the BF of the neural populations for the following two reasons. First, as illustrated in the middle panels of Figure 4, when the F0 is shifted, some stimulus components move toward the BF (e.g., components 4–10 when the F0 is shifted downward), while others move away from the BF (e.g., components 1–2 when the F0 is shifted downward). Second, response enhancements under the "shift F0" condition are observed for both upward and downward directions of mistuning. Thus, if the

response enhancements are simply due to an increase in the concentration of stimulus components near the BF of the recording sites, then, contrary to what is observed, these enhancements should occur only for one direction of mistuning (e.g., the direction that entails a greater number of components shifting toward the BF). Thus, the results suggest that mistuning-related response enhancements under the "shift F0" condition cannot be predicted based solely on the relationship between the spectrum of the stimuli and the FRF of the recording sites.

Responses to mistuned stimuli: mean population data

We examined whether the MUA response patterns displayed at the individual sites shown in Figure 4 are representative of our entire sample of A1 MUA recorded in LL3 and at more superficial cortical depths (SG1 and SG2). Mean waveforms of MUA evoked by harmonic and by 16% mistuned stimuli under "shift 3rd" and "shift F0" conditions averaged across all 46 electrode penetrations are shown superimposed in Figure 5A. Consistent with results shown in Figure 4, mean responses to mistuned stimuli (blue and red waveforms) under the "shift F0" condition, particularly within the "sustained" time window, are larger than the mean response to the harmonic stimulus (black waveform). In contrast, mean responses to mistuned stimuli under the "shift 3rd" condition are generally smaller than the mean response to the harmonic stimulus. This general pattern is displayed also by MUA recorded at more superficial cortical depths (SG1 and SG2) and is quantified in Figure 5B, which indicates statistically significant increases in MUA (integrated over the "total" response window) evoked by mistuned stimuli under the "shift F0" condition relative to that evoked by the harmonic stimulus (planned, one-tailed paired t tests; p values are included in Fig. 5B).

Next, we compared average response patterns obtained under "shift 6th" and "stretched" conditions with those obtained under "shift 3rd" and "shift F0" conditions. Given the smaller number of sites at which "shift 6th," and "stretched" stimuli were presented (22), to reduce variability in response amplitudes across sites and thereby compensate for the smaller sample size, responses under all stimulus conditions were first normalized before averaging across sites as follows. For each stimulus block ("shift 3rd," "shift F0," "shift 6th," and "stretched") and for each response time window examined ("on," "sustained," and "total"), MUA amplitudes were first normalized to the amplitude of the maximal response obtained in the stimulus block. Normalized amplitudes under each stimulus condition were then averaged across electrode penetrations. Results of this analysis are shown in Figure 6. Average normalized responses to HCTs with a mistuned component under the "shift 3rd" condition are generally diminished relative to responses under the harmonic condition (Fig. 6A) (planned one-tailed t test; p values are included in the figure). This finding is consistent with the shift of the mistuned component away from the BF under the "shift 3rd" condition and demonstrates the spectral selectivity of neural population responses in A1 and their sensitivity to relatively small changes in the frequency of a single stimulus component situated near the peak of the FRF. In contrast, under the "shift F0" condition average normalized responses to mistuned stimuli are generally enhanced relative to responses under the harmonic condition (Fig. 6B). Consistent with psychoacoustic findings, response enhancement is observed under both upward and downward mistuning conditions and is significantly greater when components are mistuned by 16% than when they are mistuned by 8% (planned one-tailed t tests comparing "shift F0" 16% up and "shift F0" 8% up and comparing "shift F0" 16%
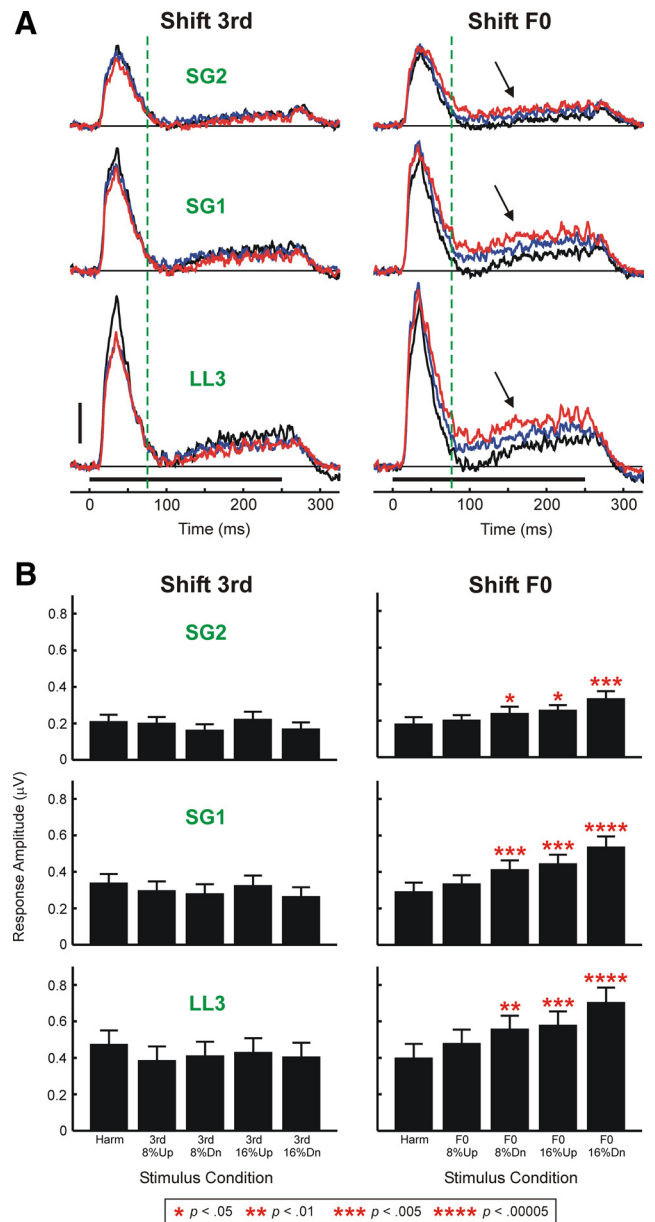

Figure 5. *A*, Waveforms of MUA evoked by harmonic and mistuned stimuli averaged across all 46 electrode penetrations into A1. Mean MUA evoked by harmonic stimuli is plotted in black; responses to stimuli mistuned via upward and downward shifts of 16% are plotted in blue and red, respectively (responses to stimuli with 8% shifts are omitted for clarity). Stimulus duration is represented by the black horizontal bar above the time axes. Mean waveforms of MUA evoked under "shift 3rd" and "shift F0" conditions are plotted in the left and right columns, respectively. Mean MUA recorded in LL3 and at more superficial cortical depths (SG1 and SG2) is plotted in separate rows, as indicated. Vertical dashed green lines mark the boundary between "on" and "sustained" portions of responses analyzed in the study. Black arrows indicate enhanced "sustained" responses to mistuned stimuli under the "shift F0" condition. Height of the vertical calibration bar represents 0.5 μV. *B*, Mean MUA integrated over the "total" response window and averaged across the 46 electrode penetrations. The layout is the same as in *A*. Error bars represent SEM. Statistically significant ($p < 0.05$) increases in MUA amplitude under mistuned conditions relative to the harmonic condition are indicated by red asterisks placed above the bars. The p value associated with each of the planned one-tailed paired t tests comparing MUA amplitudes under mistuned and harmonic conditions is represented by the number of asterisks, as identified in the legend at the bottom of the figure. Significant mistuning-related response enhancements are observed only under the "shift F0" condition.

down and "shift F0" 8% down) (comparisons are indicated by double arrows in Fig. 6B) ($p < 0.05$ for all tests). Statistically significant response enhancements are observed for all response time windows analyzed.
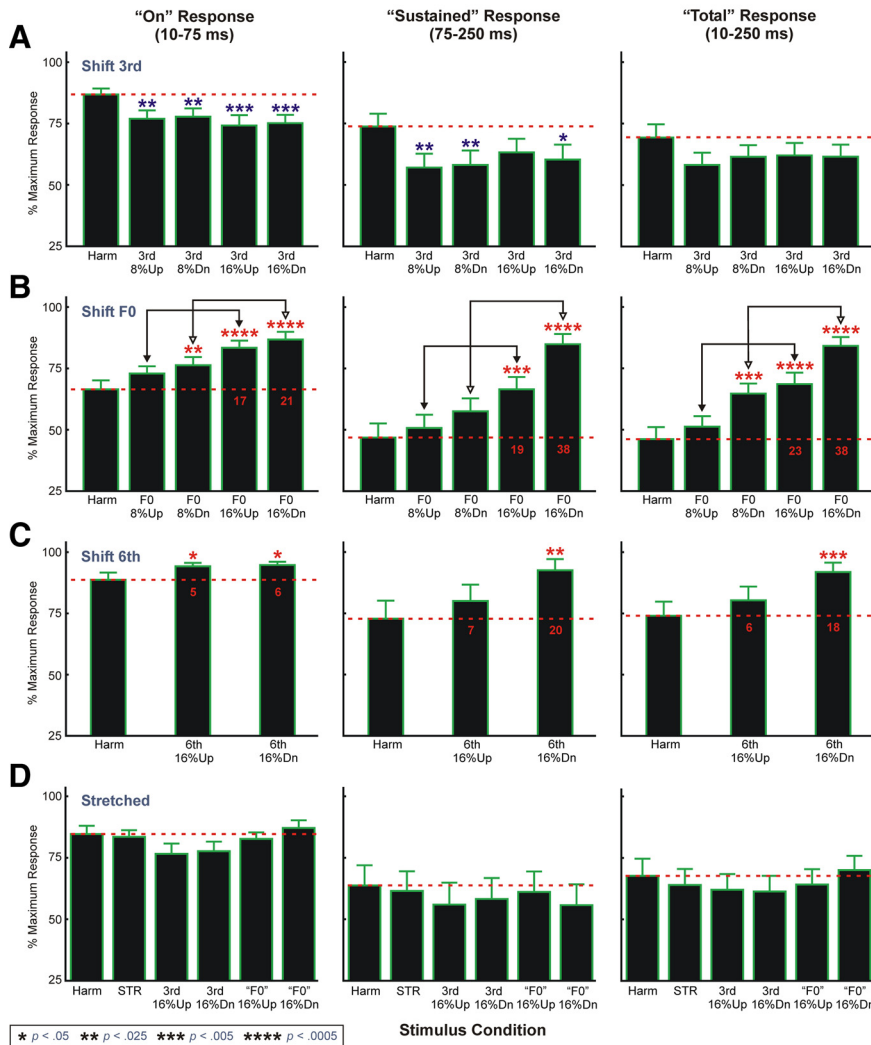
Figure 6. Normalized LL3 MUA data averaged across electrode penetrations. Mean normalized data for "shift 3rd," "shift F0," "shift 6th," and "stretched" stimulus blocks are represented in **A–D**, respectively. Number of electrode penetrations contributing to the mean data for "shift 3rd," "shift F0," "shift 6th," and "stretched" stimulus blocks is 46, 46, 22, and 22, respectively. Data for the three response time windows analyzed are represented in separate columns, as indicated. The red horizontal dashed line superimposed on the graphs indicates the mean normalized response amplitude under the harmonic condition. Error bars represent SEM. Statistically significant ($p < 0.05$) increases and decreases in mean normalized MUA amplitude under mistuned conditions relative to the harmonic condition are indicated by red and blue asterisks, respectively, placed above the bars. The p value associated with each of the planned one-tailed t tests comparing normalized MUA amplitude under mistuned and harmonic conditions is represented by the number of asterisks, as identified in the legend at the bottom of the figure. Additional planned t tests include comparisons (represented by the double arrows) between mean normalized responses under the 8% and 16% mistuning conditions of the "shift F0" stimulus block. For all comparisons, mean normalized responses to 16% mistuned stimuli are significantly larger than those to 8% mistuned stimuli ($p < 0.05$). Red numbers superimposed on the black bars in **B** and **C** indicate the difference in percentage points between mean normalized response amplitudes under mistuned and harmonic conditions.

The greater response enhancement observed for downward shifts in F0 under the "shift F0" condition is consistent with a greater number of stimulus components moving toward the BF of the recorded neural populations. However, as noted earlier, shifts of the stimulus spectrum relative to the FRF cannot fully account for response enhancements under the "shift F0" condition, as significant enhancements are observed also for upward shifts of components, while downward shifts also involve movement of two components away from the BF. Thus, the neural responses to the mistuned stimuli under the "shift F0" condition reflect both genuine enhancements, as well as shifts of the stimulus spectrum in relation to the FRF of the site.

To evaluate the tonotopic specificity of response enhancements observed under the "shift F0" condition, we examined whether they are comparable when the sixth component is mistuned away from its harmonic value by 16% while the third harmonic remains fixed at the BF of the recording site ("shift 6th" condition). While responses under "shift 6th" conditions tend to be enhanced relative to those under the harmonic condition (Fig. 6C), response enhancements are considerably smaller than those observed under the "shift F0" 16% conditions (compare red numbers superimposed on the black bars in Fig. 6B,C indicating the difference in percentage points between mean normalized amplitudes under mistuned and harmonic conditions). These findings suggest that response enhancements are maximal for local neuronal populations tuned to the frequency of the mistuned third harmonic, rather than reflecting a more uniform and nonspecific elevation in activity across A1.

Importantly, responses to globally inharmonic "stretched" stimuli are not enhanced relative to responses under the harmonic condition (Fig. 6D), indicating that response enhancements observed under the "shift F0" condition correlate with the perceptual "pop out" of a mistuned component rather than with inharmonicity per se. Furthermore, no statistically significant enhancements are observed for responses to mistuned stimuli of the "stretched" stimulus block relative to responses under the "STR" condition (one-tailed t tests; $p > 0.05$), consistent with psychoacoustic findings that detection of individual mistuned components within 12% stretched stimuli is poorer than detection of mistuned components within harmonic sounds (Roberts and Brunstrom, 2001).

Qualitatively similar results were obtained for both normalized and unnormalized MUA recorded in the more superficial electrode channels (SG1 and SG2) and for PSTH measures of action potential activity in LL3 [supplemental Figs. 1, 2, and 3 (available at www.jneurosci.org as supplemental material), respectively].

Tonotopic specificity of mistuning-related response enhancements

The tonotopic specificity of mistuned-related response enhancements is suggested by the diminished response enhancements observed under the "shift 6th" condition compared with those observed under the "shift F0" condition. The degree of response enhancement was also quantified by Cohen's d, a standardized index of effect size (see Materials and Methods for details). As shown in Figure 7, effect sizes are consid-
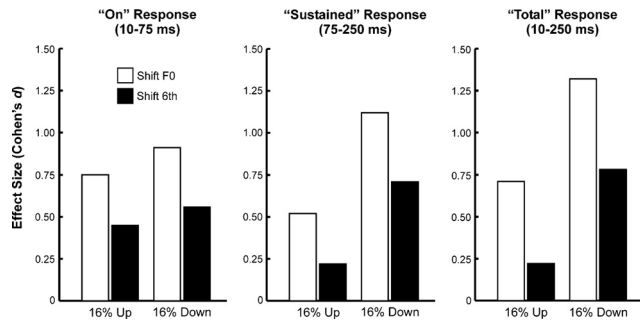
Figure 7. Comparison between effect sizes (as quantified by Cohen's *d*) under "shift F0" (white bars) and "shift 6th" (black bars) conditions for LL3 MUA integrated within the three response windows analyzed, as indicated. See Materials and Methods for details.

erably larger under the "shift F0" condition than under the "shift 6th" condition.

To further test the hypothesis that mistuning-related response enhancements are maximal for neural populations tuned to the frequency of the mistuned component, at a subset of electrode penetrations in one animal, we presented HCTs with F0s fixed at 125, 250, 500, 750, and 1000 Hz (designated as "fixed F0" stimuli; see Materials and Methods for details), regardless of the BF of the recording site. HCTs were either "in tune" or had their third component mistuned upward or downward by 16% of its harmonic value (as in the "shift 3rd" condition). Thus, the difference in frequency between the third harmonic and the BF of the site varied across stimuli and across electrode penetrations.

Consistent with the tonotopic specificity of mistuning-related response enhancements, response enhancement is significantly greater when the frequency of the mistuned harmonic is equal to the BF than when it is far away from the BF. This is demonstrated in Figure 8, which compares the mean difference in the amplitude of LL3 MUA (mean "total" response) evoked by mistuned and by harmonic "fixed F0" stimuli when the frequency of the third harmonic is greater than half an octave above or below the BF (as indicated over the bars plotted in the left and right of the figure) with the mean difference in MUA amplitude when the frequency of the third harmonic is equal to the BF ("shift F0" condition; bars plotted in the center of the figure above zero amplitude). To ensure comparability of results, the data for the "shift F0" condition represented in Figure 8 are derived from the same 16 electrode penetrations in which "fixed F0" stimuli were tested. The mean change in the amplitude of responses to the mistuned stimuli relative to responses to harmonic stimuli is significantly larger when the frequency of the third harmonic is equal to the BF (under the "shift F0" condition), than when it is greater than half an octave above or below the BF (one-tailed *t* tests; for all tests *p* < 0.0001). Consistent with the diminished responses to mistuned stimuli under the "shift 3rd" condition (Fig. 6*A*), the mean change in the amplitude of responses to mistuned stimuli is negative when the third harmonic, set equal to the BF under the harmonic condition, is shifted away from the BF under the mistuned conditions (bars plotted in the center of the figure below zero amplitude). Together with the diminished response enhancements observed under the "shift 6th" condition (Figs. 6*C*, 7), these findings strongly suggest that mistuning-related response enhancements are maximal for neurons tuned to the frequency of the mistuned component.

Mistuning-related changes in temporal response patterns
In addition to increases in neuronal firing rate, responses to mistuned stimuli may exhibit prominent temporal discharges that
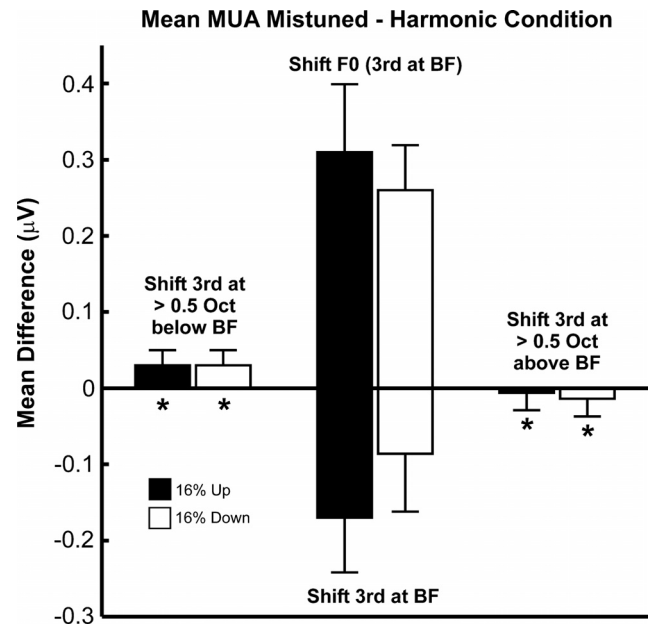

Figure 8. Mistuning-related response enhancements are maximal when the frequency of the mistuned component is equal to the BF. Mean differences between amplitudes of LL3 MUA (averaged within the "total" response window) evoked by mistuned and harmonic stimuli are significantly diminished when the frequency of the mistuned third harmonic is >0.5 octave above or below the BF (asterisks) compared to when the frequency of the mistuned component is equal to the BF (i.e., under the "shift F0" condition; bars in center of the figure above zero amplitude). Bars in center of the figure below zero amplitude represent mean data under the "shift 3rd" condition. Data for upward and downward shifts of stimulus components are represented by black and white bars, respectively. Error bars represent SEM. See Results for details.

are phase locked to low-frequency fundamental and "beat" frequencies similar to those observed in the inferior colliculus (IC) (Sinex et al., 2002). Temporally modulated responses to harmonic and mistuned stimuli ("shift F0" condition) at two representative sites are shown in Figure 9, *A* and *B* (BF = 350 and 250 Hz, respectively). Importantly, temporal response patterns evoked by mistuned stimuli differ from those evoked by harmonic stimuli. Whereas responses evoked by both harmonic and mistuned stimuli are phase locked to the F0, those evoked by mistuned stimuli are phase locked also to lower-frequency "beats" corresponding to the difference in frequency between the mistuned component and adjacent components of the sounds. Phase locking to the F0 and "beat" frequencies is represented by peaks in the associated response spectra (blue curves in Fig. 9; green and red numbers indicate the frequency of spectral peaks corresponding to the F0 and "beat" frequencies, respectively). Differences in temporal response pattern may not only distinguish harmonic from mistuned sounds but may also provide a complementary mechanism for segregating the mistuned component from the other (non-mistuned) components of the complex sounds. Moreover, we propose that the encoding of lower-frequency "beats" as a rate code at the cortical level may explain the enhanced responses to the mistuned stimuli described earlier (see Discussion). Accordingly, at both sites depicted in Figure 9, the mistuned stimuli, which contain lower "beat" frequencies than the harmonic stimuli, evoke greater peak sustained and mean total MUA than the harmonic stimuli (mean total MUA values are shown above the MUA waveforms in Fig. 9).

To examine whether a change in temporal response pattern associated with mistuning is a general phenomenon occurring across the entire sample of A1 sites, we computed at each site the
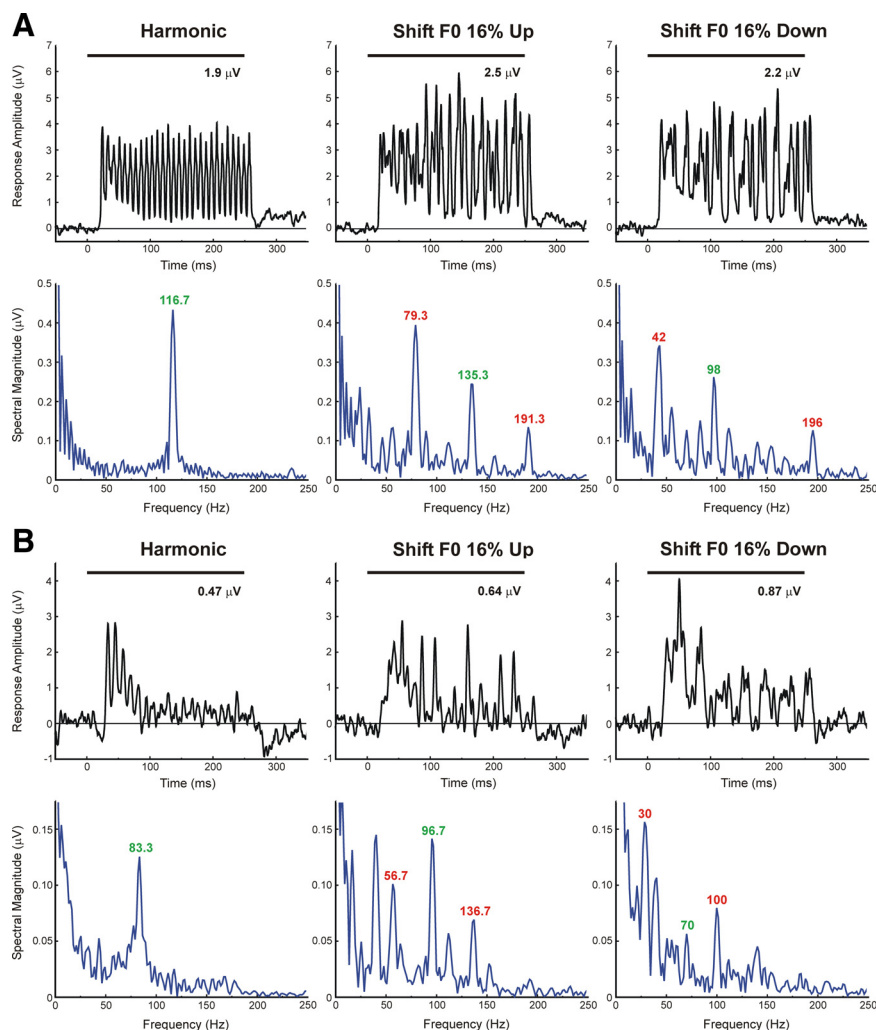
Figure 9. Examples of temporally modulated responses evoked by harmonic and mistuned stimuli. LL3 MUA recorded at two A1 sites (**A**, **B**) displays temporal patterns that are phase locked to the F0 and predicted "beat" frequencies of the stimuli (harmonic, "shift F0" 16% up, and "shift F0" 16% down, as indicated). Plots on the top row (black) and bottom row (blue) of **A** and **B** represent MUA waveforms and corresponding spectra (discrete Fourier transform of MUA waveform from 10 to 250 ms after stimulus onset), respectively. BFs of sites represented in **A** and **B** are 350 Hz and 250 Hz, respectively. Stimulus duration is represented by the black horizontal bar above the waveforms. Green and red numbers in the response spectra indicate the frequency of spectral peaks corresponding to the F0 and "beat" frequencies of the stimuli, respectively. The value of the mean MUA computed over the "total" response window is indicated above the MUA waveforms.

Pearson correlation between the waveforms of MUA evoked under the mistuned and harmonic conditions, based on the rationale that a difference in temporal response pattern will yield a lower correlation coefficient than a similar temporal response pattern. To provide a statistical means of testing whether mistuning results in a significantly lower response correlation, the mean Pearson correlation coefficient obtained between responses to mistuned and harmonic stimuli was compared with that obtained between responses to the identical harmonic stimulus presented in different stimulus blocks (see Materials and Methods). This latter mean correlation coefficient provides a "baseline" measure both of the variability in temporal response patterns evoked by the same stimulus (harmonic) across stimulus blocks, and of correlations due to noise. Thus, a generalized change in temporal response pattern associated with mistuning would be reflected by a statistically significant decrease in the mean (Z-transformed) Pearson correlation between responses under mistuned and harmonic conditions relative to correlations be-

tween responses to identical harmonic stimuli presented in different stimulus blocks. As shown in Figure 10, mistuning is indeed associated with a statistically significant reduction in waveform correlations (one-tailed paired t tests; except where indicated by "ns," all differences are statistically significant, with p values ranging from $<10^{-2}$ to $<10^{-12}$), thus supporting the hypothesis that perceptual "pop out" is related also to a difference in temporal response pattern in A1. Significant differences in temporal response pattern, as quantified by the measures described above, are observed for (nearly) all stimulus conditions, including the "stretched" condition (Fig. 10). Note that under the "shift F0" condition, the change in temporal response pattern relative to that observed under the harmonic condition may reflect not only phase locking to local "beats" associated with the mistuned component, but also phase locking to the different stimulus F0. Differences in temporal response patterns, as quantified by waveform correlations, tend to be greatest at sites with a lower BF (Fig. 11), consistent with "beat" frequencies being lower for stimuli with lower mistuned third harmonics, and therefore more likely to be encoded as phase-locked temporal discharges at the cortical level.

Tonotopic specificity of mistuning-related changes in temporal response patterns

While a change in temporal response pattern associated with mistuning may provide a physiological basis for distinguishing harmonic from inharmonic stimuli, it is insufficient to account for the perceptual "pop out" of the mistuned harmonic, unless it can be shown that the change in temporal response pattern is specific to neurons tuned to the frequency of the mistuned harmonic. Indeed, as shown in Figure 12, the change in temporal response pattern is greatest, as reflected by a significantly lower mean (Z-transformed) Pearson correlation coefficient, when the frequency of the mistuned component (before mistuning) is equal to the BF (under the "shift 3rd" condition) and least when it is far away from the BF (between 0.5 and 1 octave away). Consistent with the tonotopic specificity of response enhancements shown earlier, these findings suggest that perceptual "pop out" is associated also with a local difference in the temporal response pattern of neurons tuned to a frequency equal to or near that of the mistuned component.

Thus, the present findings suggest that perceptual "pop out" of mistuned component may be due to two complementary neural mechanisms in A1: response enhancement or a difference in temporal response pattern that is specific to neurons tuned to the frequency of the mistuned component. Moreover, mistuning-related response enhancements tend to be equally or more pro-

nounced at sites with a higher BF, where the frequency of "beats" is higher, given the higher third harmonic of the stimuli presented at those sites, and therefore where neuronal phase locking to "beats" is less likely to occur (supplemental Fig. 4, available at www.jneurosci.org as supplemental material). This suggests that these two neural mechanisms are nonredundant, with perceptual "pop out" being associated with a local increase in firing rate for mistuned stimuli with higher "beat" frequencies and also with a difference in temporal response pattern for mistuned stimuli with lower "beat" frequencies.

AEP correlates of perceptual "pop out" parallel the human "object-related negativity"

AEPs elicited by "in-tune" and by mistuned HCTs were compared to identify potential monkey homologs of the ORN and P230 components recorded noninvasively in humans that correlate with the perceptual "pop out" of a mistuned harmonic embedded within an otherwise HCT. For each electrode penetration, the AEP under the harmonic condition was subtracted from the AEP evoked under each of the mistuned conditions. The resultant difference waveforms were then averaged across electrode penetrations. Differences between the harmonic and mistuned conditions across penetrations were evaluated by a paired t test performed at each time point in the AEP waveforms. Significant differences in AEPs evoked under mistuned and harmonic conditions were observed under both "shift 3rd" and "shift F0" conditions and for both directions of mistuning (Fig. 13 A, B; upward and downward mistuning conditions represented by blue and red waveforms, respectively). The largest differences in AEPs evoked under the "shift 3rd" and "shift F0" conditions occur at latencies >100 ms after stimulus onset and include negative and positive difference-wave components in superficial cortical layers (labeled as ORN_m and ORP_m, respectively, in Fig. 13 A, B). Amplitudes of the ORN_m and ORP_m are greater when components are mistuned by 16% than when they are mistuned by 8%. In contrast, the ORN_m and ORP_m are markedly diminished or absent in responses evoked by "shift 6th" and "stretched" stimuli (data not shown; available upon request). The polarity of the ORN_m and ORP_m inverts over the laminar extent of A1 (Fig. 13 A, B), indicative of activity within pyramidal neuron populations in auditory cortex. Furthermore, the latencies and polarities of the ORN_m and ORP_m are remarkably similar to those of the human ORN and P230 components (Alain et al., 2001, 2002; Alain and McDonald, 2007; Lipp et al., 2010), thus suggesting species homologies in auditory cortical mechanisms associated with the segregation of concurrent sounds.
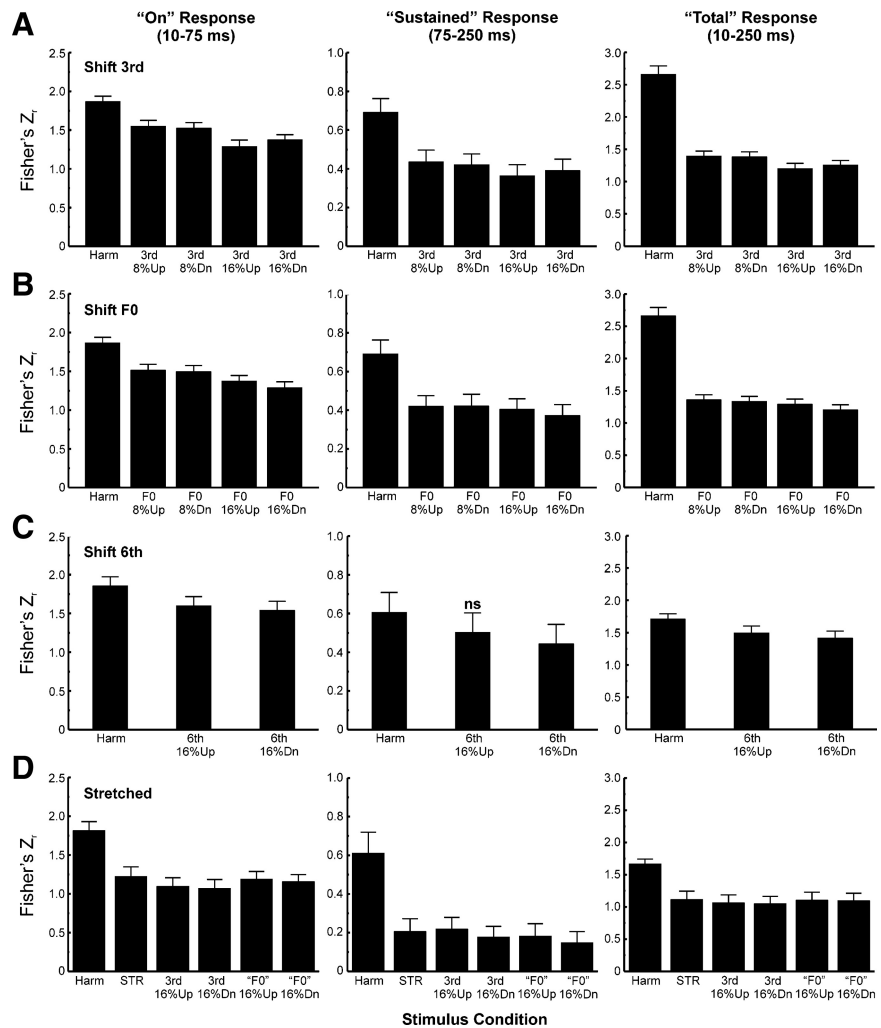

Figure 10. Mean Pearson correlation coefficients (transformed to Fisher's Z_r) quantifying the degree of similarity between temporal response patterns evoked by mistuned and harmonic stimuli. Mean Z-transformed correlation coefficients for responses to stimuli comprising the "shift 3rd," "shift F0," "shift 6th," and "stretched" stimulus blocks are represented in A–D, respectively. Data for the three response time windows analyzed are represented in separate columns, as indicated. Error bars indicate SEM. Note the different ordinate range of the plots for the different response time windows. Except where indicated by "ns," all differences between mean correlations obtained under mistuned and harmonic conditions are statistically significant (planned one-tailed paired t test), with p values ranging from $<10^{-2}$ to $<10^{-12}$. Sample sizes for each stimulus condition are the same as those in Figure 6. See Results for details.

Discussion

Summary of findings

The present study tested the hypothesis that perceptual "pop out" of a mistuned component within an otherwise harmonic complex sound may be explained by two complementary neural processes in A1: a local increase in firing rate or a difference in the temporal response pattern of neural populations that are tuned to the frequency of the mistuned component. The following observations fulfill key predictions of our hypothesis, thereby supporting a role for A1 in concurrent sound segregation based on inharmonicity:

Responses evoked by HCTs containing a mistuned component with a frequency equal to the BF of the recorded neural populations are enhanced relative to those evoked by "in-tune" HCTs, thus correlating with the perceptual "pop out" of the mistuned component. Consistent with psychoacoustic data, response enhancement occurs for both upward and downward directions of mistuning, and is greater when components are mistuned by 16%
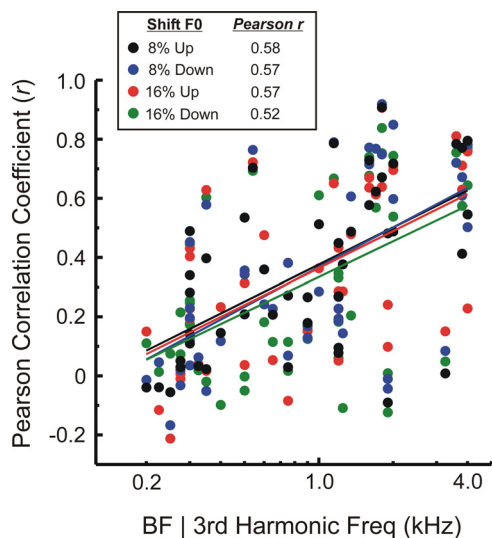
Figure 11. Relationship between BF (third harmonic frequency) and the Pearson correlation coefficient (r) quantifying the similarity between waveforms of "sustained" LL3 MUA evoked by mistuned and harmonic stimuli under the "shift F0" condition. Correlation coefficients for upward and downward directions of mistuning and for degrees of mistuning of 8 and 16% are plotted in different colors, as indicated in the legend. Correlation coefficients tend to be lower for responses evoked by stimuli with lower third harmonic frequencies, indicating greater dissimilarity between responses evoked by mistuned and harmonic stimuli. This trend is quantified by Pearson correlation coefficients included in the legend (computed as r vs log BF) and emphasized by the superimposed color-coded linear regression lines. All four correlations are statistically significant (n = 46; p < 0.0005).
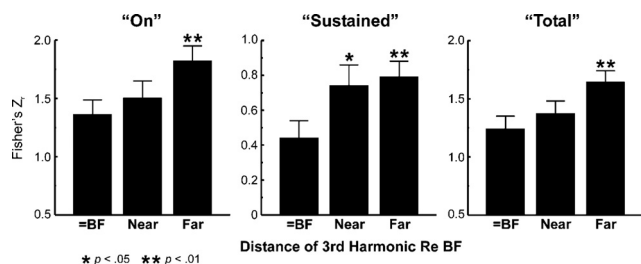

Figure 12. Mean Pearson correlation coefficients (transformed to Fisher's Z_r) quantifying the degree of similarity between waveforms of LL3 MUA evoked by mistuned and harmonic stimuli when the frequency of the mistuned third harmonic, before mistuning, is equal to the BF (under the "shift 3rd" condition; n = 16), near the BF (between 0.25 and 0.5 octave away; n = 14), and far from the BF (between 0.5 and 1 octave away; n = 16). Error bars represent SEM. Correlation coefficients are collapsed across direction of mistuning and across position of the third harmonic above and below the BF. Mean (Z-transformed) correlation coefficients for the three response windows analyzed are represented in separate plots, as indicated. Correlation coefficients are significantly larger (indicating greater similarity between responses) when the frequency of the mistuned component is far from the BF than when it is equal to the BF (planned one-tailed unpaired t test; p values are represented by the number of asterisks, as indicated at the bottom of the figure). Note the different ordinate range of the plot for "sustained" responses. All data are derived from the same 16 electrode penetrations at which "fixed F0" stimuli were presented.

than when they are mistuned by 8% (Alain et al., 2001). Response enhancement is observed also for later "sustained" responses, consistent with the improved detection of a mistuned harmonic when stimulus duration is extended (Moore et al., 1986). Finally, responses to globally inharmonic "stretched" stimuli are not enhanced relative to responses to "in-tune" HCTs, thus indicating that response enhancements correlate with the "pop out" of a mistuned component rather than with inharmonicity per se.

The tonotopic specificity of mistuning-related response enhancements is supported by (1) MUA response enhancement when the mistuned harmonic is fixed at the BF of the recorded neural population ("shift F0" condition) (Figs. 5, 6B), but not when it is shifted away from the BF ("shift 3rd" condition) (Figs. 5, 6A), (2) diminished response enhancement when the third harmonic is set equal to the BF and the sixth harmonic is mistuned (Figs. 6C, 7), and (3) maximal response enhancement when the frequency of the mistuned harmonic is equal to the BF (Fig. 8).

Similarly, mistuning-related changes in temporal response patterns are also specific to neurons tuned to a frequency equal to or near the frequency of the mistuned component (Fig. 12). In contrast, mistuning-related changes in temporal response pattern and firing rate in the IC are tonotopically nonspecific, in that they are observed even when the frequency of the mistuned component does not correspond to the BF of the recorded neurons (Sinex et al., 2002). However, to qualify as a neural correlate of the perceptual "pop out" of a mistuned component (rather than simply the detection of inharmonicity) changes in temporal response pattern or firing rate should be maximal for neurons tuned to the frequency of the mistuned component. In demonstrating the tonotopic specificity of mistuning-related changes, the present findings thus provide a crucial link between neuronal activity and auditory perception.

This link is bolstered further by the observation of two prominent AEP difference-waveform components (designated ORN_m and ORP_m) that parallel the ORN and P230 components recorded noninvasively in humans and that correlate with the perception of two distinct sound sources (see Introduction). The ORN_m and ORP_m are not observed in responses evoked by "stretched" stimuli, again suggesting that they reflect the "pop out" of a mistuned component rather than inharmonicity per se. Amplitudes of the ORN_m and ORP_m are greater when components are mistuned by 16% than when they are mistuned by 8%, consistent with neurophysiological findings in humans (Alain et al., 2001). The polarity and timing of the ORN_m and ORP_m closely match those of the ORN and P230 components recorded in humans, thus suggesting species homologies in neural processes underlying scene analysis based on inharmonicity. The fact that the polarity and timing of the ORN_m and ORP_m are nearly identical for both upward and downward directions of mistuning and for both "shift 3rd" and "shift F0" conditions—stimuli that elicit a similar perceptual "pop out" but that have markedly different acoustic spectra—strongly suggests that the ORN_m and ORP_m are genuine difference-wave components associated with the "pop out" of a mistuned harmonic.

Whereas consistent mistuning-related response enhancements for MUA are observed only under the "shift F0" condition, the observation that the ORN_m and ORP_m occur under both "shift 3rd" and "shift F0" conditions may be explained by the broader frequency tuning of the AEP compared with that of MUA (Fishman et al., 2000b). Consequently, unlike MUA, the AEP may be less sensitive to small shifts of the third component away from the BF under the "shift 3rd" condition.

Mechanisms of response enhancement related to perceptual "pop out"

We propose that mistuning-related response enhancement in A1 is a product of three factors: neuronal phase locking to "beats" in the mistuned stimuli observed in subcortical auditory nuclei (Sinex et al., 2002), the degradation of phase locking at the cortical level, and a transformation of a temporal representation of "beats" into a rate code within A1. Neuronal phase locking to stimulus amplitude modulations is significantly reduced at the
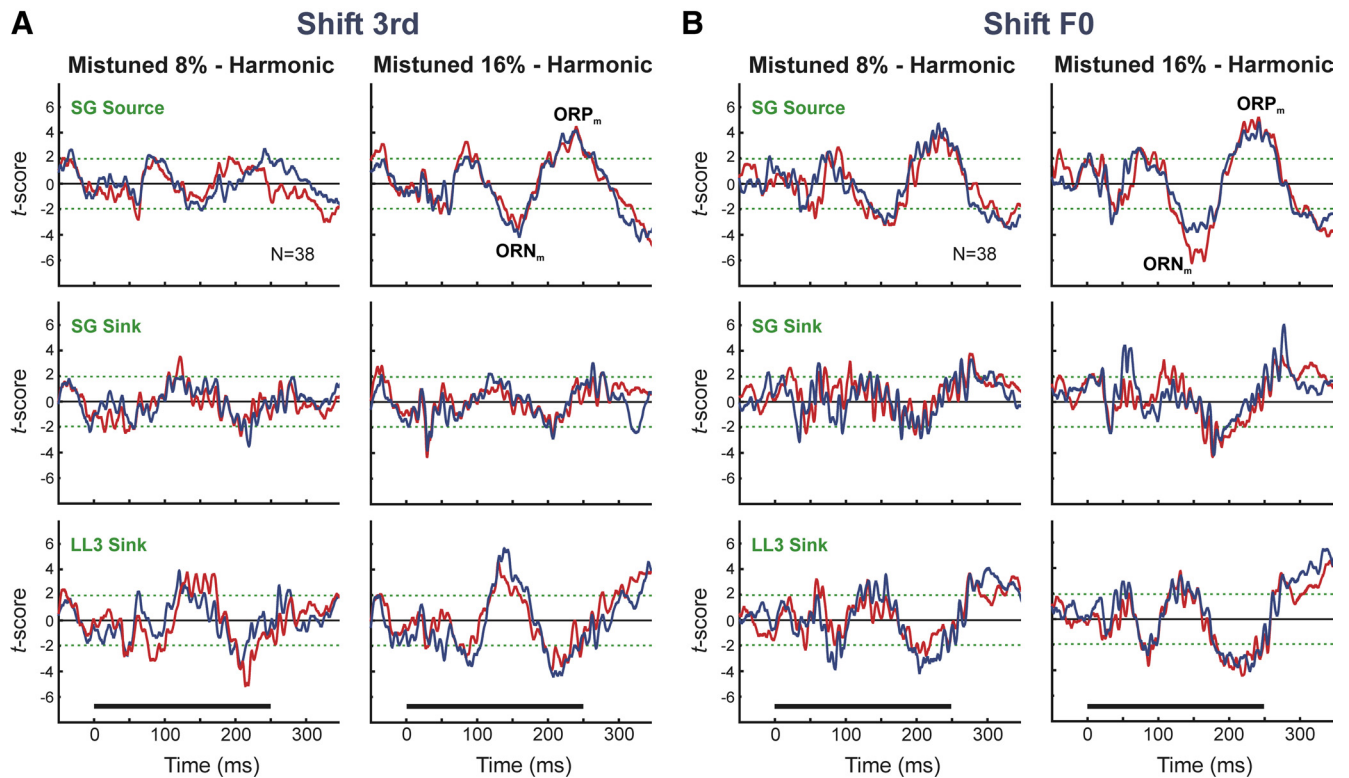
Mean AEP Difference Waveforms

Figure 13. Potential intracortical monkey homologs of the human ORN and P230 difference-waveform components. Mean difference waveforms are obtained by subtracting AEPs evoked by harmonic stimuli from AEPs evoked by mistuned stimuli presented in the "shift 3rd" (**A**) and "shift F0" (**B**) stimulus blocks and averaging across electrode penetrations. The ordinate represents the *t* score obtained for each time-point comparison; the green dashed lines denote *t* scores corresponding to a *p* value of 0.05. Mean difference waveforms under 8% and 16% mistuning conditions are plotted separately, as indicated. Mean difference waveforms for AEPs evoked by stimuli with upward and downward mistunings are plotted in blue and red, respectively. *N* refers to the number of sites contributing to mean difference waveforms. Mean difference waveforms for AEPs recorded at cortical depths corresponding to the location of the LL3 sink, the SG sink, and the SG source are plotted in separate rows, as indicated. Stimulus duration is represented by the horizontal black bar above the time axes. Two prominent difference-waveform components proposed to represent monkey homologs of the human ORN and P230 components are labeled ORN$_m$ and ORP$_m$, respectively, in the plots of SG source data in **A** and **B**. Peak latencies of the two components are ~150 and 230 ms, respectively.

cortical level relative to that observed in subcortical auditory structures, and displays a low-pass characteristic with an upper cutoff of ~150 Hz (Langner, 1992; Bieser and Müller-Preuss, 1996; Steinschneider et al., 1998; Fishman et al., 2000a, 2001b; Lu et al., 2001; Brugge et al., 2009). Thus, temporal encoding of higher "beat" frequencies observed in subcortical structures (Sinex et al., 2002) will typically be unavailable within A1. As mistuned HCTs contain lower "beat" frequencies than "in-tune" HCTs, phase locking to these "beats" will tend to be more robust in responses to mistuned HCTs than in responses to "in-tune" HCTs (as exemplified in Fig. 9 by the larger responses phase locked to lower-frequency "beats" in the mistuned stimuli). If a temporal representation of these "beats" is transformed into a rate code at the level of A1, responses to mistuned HCTs will be larger than those evoked by "in-tune" HCTs. This explanation is consistent with previous findings suggesting a similar temporal-to-rate transformation in the encoding of stimulus periodicities in A1 (Langner, 1992; Bieser and Müller-Preuss, 1996; Lu et al., 2001; Wang et al., 2008).

The tonotopic specificity of mistuning-related response enhancements in A1 is consistent with "beats" arising from local interactions between the mistuned component and adjacent stimulus components along the cochlea. The lack of significant response enhancement under "stretched" conditions may be explained by the fact that, unlike mistuned stimuli under the "shift

F0" condition, stretched stimuli have higher "beat" frequencies than harmonic stimuli. Thus, if the cortical neuron firing rate is inversely proportional to the frequency of the "beats" in the stimuli (as proposed above), neural activity evoked by stretched stimuli will be generally diminished relative to that evoked by harmonic stimuli.

Temporal incoherence as a mechanism of perceptual "pop out"

The finding that mistuning results in a local change in temporal response pattern in A1 suggests a mechanism of perceptual "pop out" whereby the detection of temporally incoherent responses along A1 may facilitate the segregation of a "foreground" mistuned component from a "background" of "in-tune" components. As neurons tuned to the frequency of the mistuned component would display a different temporal response pattern from that of neurons tuned to the frequency of the "in-tune" components, responses across A1 displaying a coherent temporal pattern could be grouped together and segregated from responses displaying a different temporal pattern. Accordingly, the reduced perceptual "pop out" of a mistuned component in 12% stretched stimuli may be explained by a lack of a coherent temporal response pattern (because of the numerous disparate "beat" frequencies in the stretched stimuli) relative to which a "deviant"

temporal response pattern reflecting "beats" associated with the mistuned component may be readily differentiated.

Relation of findings to perceptual "pop out" based on onset asynchrony

The neural mechanisms of perceptual "pop out" based on inharmonicity proposed here may account also for perceptual "pop out" based on onset asynchrony, an additional powerful cue used in auditory scene analysis. For instance, a component that is sufficiently delayed relative to the onset of the other components in an HCT is heard as a separate "auditory object" (Bregman, 1990; Darwin and Carlyon, 1995). An onset asynchrony exceeding 20 ms between two temporally overlapping tones is associated with both a relative response enhancement and a local change in the temporal pattern of activation in A1 (e.g., Steinschneider et al., 2005). Thus, the detection of a local difference in firing rate or temporal response pattern in A1 may represent a generic physiological mechanism underlying the perceptual segregation of temporally overlapping sounds.

Implications and future studies

The present findings bolster the translational relevance of cortical population responses recorded in monkeys to elucidating comparable neural substrates of auditory scene analysis in humans. While speculative, a neural model of "pop out" based on the detection of temporally incoherent responses along A1 may also account for the perceptual segregation of two simultaneous, spectrally overlapping HCTs differing in F0 (Fishman et al., 2001b; Sinex and Li, 2007; Elhilali et al., 2009a; Micheyl and Oxenham, 2010). How neural response patterns in A1 are "read out" by secondary auditory cortical areas and modulated by top-down attentional processes in the analysis of complex auditory scenes (e.g., Elhilali et al., 2009b) remains to be explored in future studies.

References

Alain C, McDonald KL (2007) Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. J Neurosci 27:1308–1314.

Alain C, Arnott SR, Picton TW (2001) Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. J Exp Psychol Hum Percept Perform 27:1072–1089.

Alain C, Schuler BM, McDonald KL (2002) Neural activity associated with distinguishing concurrent auditory objects. J Acoust Soc Am 111:990–995.

Arezzo JC, Vaughan HG Jr, Kraut MA, Steinschneider M, Legatt AD (1986) Intracranial generators of event-related potentials in the monkey. In: Frontiers of clinical neuroscience: evoked potentials, Vol 3 (Cracco RQ, Bodis-Wollner I, eds), pp 174–189. New York: Liss.

Bieser A, Müller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp Brain Res 108:273–284.

Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT Press.

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

Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA 3rd (2009) Coding of repetitive transients by auditory cortex on Heschl's gyrus. J Neurophysiol 102:2358–2374.

Cohen J (1992) A power primer. Psychol Bull 112:155–159.

Darwin CJ, Carlyon RP (1995) Auditory grouping. In: Handbook of perception and cognition: hearing (Moore BCJ, ed), pp 387–424. New York: Academic.

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

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

Fishman YI, Steinschneider M (2006) Spectral resolution of monkey primary auditory cortex (A1) revealed with two-noise masking. J Neurophysiol 96:1105–1115.

Fishman YI, Steinschneider M (2009) Temporally dynamic frequency tuning of population responses in monkey primary auditory cortex. Hear Res 254:64–76.

Fishman YI, Steinschneider M (2010) Formation of auditory streams. In: The oxford handbook of auditory science: the auditory brain (Rees A, Palmer AR, eds), pp 215–245. New York: Oxford UP.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2000a) Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. J Acoust Soc Am 108:235–246.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2000b) Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation. J Acoust Soc Am 108:247–262.

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

Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, Arezzo JC, Howard MA, Steinschneider M (2001b) Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J Neurophysiol 86:2761–2788.

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

Guilford JP (1965) Fundamental statistics in psychology and education. New York: McGraw-Hill.

Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. J Comp Neurol 394:475–495.

Hartmann WM, McAdams S, Smith BK (1990) Hearing a mistuned harmonic in an otherwise periodic complex tone. J Acoust Soc Am 88:1712–1724.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115–142.

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

Lipp R, Kitterick P, Summerfield Q, Bailey PJ, Paul-Jordanov I (2010) Concurrent sound segregation based on inharmonicity and onset asynchrony. Neuropsychologia 48:1417–1425.

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131–1138.

Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res 50:275–296.

Metherate R, Cruikshank SJ (1999) Thalamocortical inputs trigger a propagating envelope of gamma-band activity in auditory cortex in vitro. Exp Brain Res 126:160–174.

Micheyl C, Oxenham AJ (2010) Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. Hear Res 266:36–51.

Moore BC, Peters RW, Glasberg BR (1985) Thresholds for the detection of inharmonicity in complex tones. J Acoust Soc Am 77:1861–1867.

Moore BC, Glasberg BR, Peters RW (1986) Thresholds for hearing mistuned partials as separate tones in harmonic complexes. J Acoust Soc Am 80:479–483.

Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. J Comp Neurol 335:437–459.

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

Nelken I, Prut Y, Vaadia E, Abeles M (1994) Population responses to multifrequency sounds in the cat 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.

Plack CJ, Oxenham AJ (2005) The psychophysics of pitch. In: Pitch: neural coding and perception (Plack CJ, Oxenham AJ, Fay RR, Popper AN, eds), pp 7–55. New York: Springer.

Recanzone GH, Guard DC, Phan ML (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83:2315–2331.

Roberts B, Brunstrom JM (2001) Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. J Acoust Soc Am 110:2479–2490.

Rosnow RL, Rosenthal R (1996) Computing contrasts, effect sizes, and counternulls on other people's published data: general procedures for research consumers. Psychol Methods 1:331–340.

Sinex DG, Li H (2007) Responses of inferior colliculus neurons to double harmonic tones. J Neurophysiol 98:3171–3184.

Sinex DG, Sabes JH, Li H (2002) Responses of inferior colliculus neurons to harmonic and mistuned complex tones. Hear Res 168:150–162.

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 Jr (1992) Cellular generators of the cortical auditory evoked potential initial component. Electroencephalogr Clin Neurophysiol 84:196–200.

Steinschneider M, Reser DH, Fishman YI, Schroeder CE, Arezzo JC (1998) Click train encoding in primary auditory cortex of the awake monkey:

evidence for two mechanisms subserving pitch perception. J Acoust Soc Am 104:2935–2955.

Steinschneider M, Fishman YI, Arezzo JC (2003) Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. J Acoust Soc Am 114:307–321.

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA 3rd (2005) Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb Cortex 15:170–186.

Sukov W, Barth DS (1998) Three-dimensional analysis of spontaneous and thalamically evoked gamma oscillations in auditory cortex. J Neurophysiol 79:2875–2884.

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

Vaughan HG Jr, Arezzo JC (1988) The neural basis of event-related potentials. In: Human event-related potentials, EEG handbook, revised series, Vol 3 (Picton TW, ed), pp 45–96. New York: Elsevier.

Wang X, Lu T, Bendor D, Bartlett E (2008) Neural coding of temporal information in auditory thalamus and cortex. Neuroscience 157:484–494.

Yost WA (1991) Auditory image perception and analysis: the basis for hearing. Hear Res 56:8–18.

Yvert B, Fischer C, Bertrand O, Pernier J (2005) Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. Neuroimage 28:140–153.