Behavioral/Systems/Cognitive

Dual-Pitch Processing Mechanisms in Primate Auditory Cortex

Daniel Bendor,* Michael S. Osmanski,* and Xiaoqin Wang
Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21025

Pitch, our perception of how high or low a sound is on a musical scale, is a fundamental perceptual attribute of sounds and is important for both music and speech. After more than a century of research, the exact mechanisms used by the auditory system to extract pitch are still being debated. Theoretically, pitch can be computed using either spectral or temporal acoustic features of a sound. We have investigated how cues derived from the temporal envelope and spectrum of an acoustic signal are used for pitch extraction in the common marmoset (*Callithrix jacchus*), a vocal primate species, by measuring pitch discrimination behaviorally and examining pitch-selective neuronal responses in auditory cortex. We find that pitch is extracted by marmosets using temporal envelope cues for lower pitch sounds composed of higher-order harmonics, whereas spectral cues are used for higher pitch sounds with lower-order harmonics. Our data support dual-pitch processing mechanisms, originally proposed by psychophysicists based on human studies, whereby pitch is extracted using a combination of temporal envelope and spectral cues.

Introduction

Pitch is our perception of fundamental frequency (f_0) (Plack et al., 2005; Oxenham, 2012). Because pitch is not explicitly represented in the cochlea, it must be computed by the central auditory system. Periodicity information from either spectral or temporal acoustic features could potentially be used as a cue by the auditory system to extract the f_0 of a sound (Goldstein, 1973; Shackleton and Carlyon, 1994; Shamma and Klein, 2000).

Complex sounds containing components at integer multiples of a common f_0 are spectrally periodic, and this spectral periodicity can theoretically be measured using a harmonic template (a form of spectral pattern matching) (Goldstein, 1973). Any harmonic template-based mechanism must be sensitive to both the frequency spacing between harmonics and the absolute frequency of each harmonic component. A harmonic template mechanism can only function with resolved harmonics. In humans, the first five to eight harmonics of a complex tone are resolved (Plomp, 1964; Plomp and Mimpen, 1968; but see Bernstein and Oxenham, 2003). Evidence for the existence of harmonic template neurons suitable for extracting pitch is still lacking in the central auditory system (including auditory cortex).

A complex sound composed of more than one harmonic of a common f_0 also has a temporal periodicity in the envelope, with a repetition rate equal to the f_0 of a sound. Temporal envelope periodicity can be calculated by performing an autocorrelation on the phase-locked auditory nerve firing, which only provides a suitable estimate of pitch for f_0 below 1300 Hz (Cariani and Delgutte, 1996a,b). At higher frequencies, the fidelity of the temporal envelope information needed for this analysis is compromised by the degraded phase locking of auditory nerve fibers. The most common all-order interspike interval (analogous to the peak of the autocorrelation) provides a good estimate of the f_0 of the acoustic signal (Cariani and Delgutte, 1996a,b; Cedolin and Delgutte, 2005), whereas pitch salience is related to how well a periodic template can be fitted to all peaks in the autocorrelation function (Bidelman and Heinz, 2011). However, whether downstream neurons use this information for computing pitch has not been determined.

In both humans and monkeys, a putative pitch-processing center has been reported in the low-frequency region of auditory cortex (Patterson et al., 2002; Penagos et al., 2004; Bendor and Wang, 2005, 2010; Schneider et al., 2005; Schönwiesner and Zatorre, 2008; Puschmann et al., 2010). In the common marmoset (*Callithrix jacchus*), pitch-selective neurons in this region can encode the pitch of missing fundamental (MF) sounds that are spectrally outside their excitatory frequency response area (Bendor and Wang, 2005). In addition, pitch-selective neurons are sensitive to pitch salience and temporal envelope regularity (Bendor and Wang, 2005, 2010). A crucial question that has yet been

Received May 28, 2012; revised Aug. 25, 2012; accepted Sept. 12, 2012.

Author contributions: D.B., M.S.O., and X.W. designed research; D.B. and M.S.O. performed research; D.B. and M.S.O. analyzed data; D.B., M.S.O., and X.W. wrote the paper.

*D.B. and M.S.O. contributed equally to this work.

This work was supported by National Institutes of Health Grants DC 03180 (X.W.), F31 DC 006528 (D.B.), K99-DC012321-01 (D.B.), and T32 DC000023 (M.S.O.), a Merck Award/Helen Hay Whitney Postdoctoral Fellowship (D.B.), and a Charles A. King Trust Postdoctoral Fellowship (D.B.). We thank A. Pistorio, J. Estes, E. Issa, E. Bartlett, and Y. Zhou for assistance with animal care. We are grateful to C. Cummings, R. Desideri, and M. Maguire for help running the behavioral experiments. We also thank E. Issa, L. Johnson, and two anonymous reviewers for comments and suggestions related to this manuscript.

Correspondence should be addressed to Dr. Xiaoqin Wang, Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Traylor 410, Baltimore, MD 21025. E-mail: xiaoqin.wang@jhu.edu.

D. Bendor's present address: Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Building 46, Room 5233, 77 Massachusetts Avenue, Cambridge, MA 02139.

DOI:10.1523/JNEUROSCI.2563-12.2012

Copyright © 2012 the authors 0270-6474/12/3216149-13$15.00/0

addressed in these previous studies is what acoustic cues are used by pitch-selective neurons to encode pitch. In the present study, using both behavioral and electrophysiological techniques, we provide evidence of dual-pitch-processing mechanisms in marmoset auditory cortex. These findings further advance our understanding of the neural basis of pitch processing in the primate brain.

Materials and Methods

Detailed procedures for conducting single-unit recording from awake marmosets can be found in previous publications from our laboratory (Lu et al., 2001a; Wang et al., 2005). We recorded from four marmosets in the present study [M36N (right hemisphere), M2P (left hemisphere), M41O (left hemisphere), and M32Q (left hemisphere)]. In each experiment, the marmoset was sitting quietly in a semi-restraint device with its head immobilized. Experiments were performed within a double-walled soundproof chamber (Industrial Acoustics) with an interior covered by 3-inch acoustic absorption foam (Sonex; Pinta Acoustic). High-impedance tungsten microelectrodes (2–5 MΩ; AM Systems) were mounted on a micromanipulator (Narishige) and advanced by a manual hydraulic microdrive (Trent Wells). Typically 5–15 electrode penetrations were made within a miniature recording hole (~1 mm diameter), after which the hole was sealed with dental cement and another hole was opened for new electrode penetrations. Action potentials were detected online using a template-based spike sorter (MSD; Alpha Omega), which was continuously monitored during the experiment. Neurons were recorded from all cortical layers but most commonly from supragranular layers. All experimental procedures were approved by the Johns Hopkins University Animal Use and Care Committee.

Generation of acoustic stimuli

Digitally generated acoustic stimuli were played from a free-field speaker located 1 m in front of the animal. All sound stimuli were generated at a 100 kHz sampling rate and low-pass filtered at 50 kHz. Harmonic artifacts were at least 43 dB lower than the fundamental at 80 dB SPL. This difference grew as the sound level of the fundamental decreased. The maximum sound level of individual frequency components used in this study was 80 dB SPL. The animal was awake and semi-restrained in a custom-made primate chair but was not performing a task during these experiments.

After each neuron was isolated, we measured its basic response properties, such as best frequency (BF) and sound level threshold. Neurons that were responsive to pure tones were tested with harmonic complex tones and/or acoustic pulse trains. Complex tone and acoustic pulse train stimuli were 500 ms in duration. All intertrial intervals for these stimuli were at least 1 s long.

Harmonic complex tones. The most commonly used harmonic complex tone had three components in cosine phase, and each component was played at the sound level threshold of the neuron measured at its BF. Sound levels 10 dB or more below BF threshold were also used in approximately one-third of pitch-selective neurons (25 of 74) to evoke significant MF responses. In a few cases, we used harmonic complex tones composed of more than three components with harmonics most commonly in Schroeder phase (Schroeder, 1970). Neurons failed the criteria of pitch selectivity if they did not respond to MF sounds with the individual harmonics presented at 10 dB above the sound level threshold of the neuron at its BF (or lower sound levels). Components of the MF harmonic complex tone were considered outside the excitatory frequency response area of the neuron if each harmonic component, when played individually at 0 and +10 dB relative to its sound level within the harmonic complex, did not evoke a significant excitatory response. Also for neurons tested with harmonic complex tones composed of greater than three components, no response to individual components +20 dB above threshold was also required. Sound levels were varied in 10 dB steps.

Pitch-shifting experiment. Three component complex tones were frequency shifted by an amount proportional to the fundamental frequency. The frequency spacing between components remained constant and equaled the fundamental frequency. Complex tones were shifted from 0% (harmonics 1–3) to 600% in steps of 25% (relative to f_0).

Phase manipulation experiment. Nine tone harmonic complex tones were used, with either all components in cosine phase (COS) or alternating phase (ALT). ALT phase stimuli contained odd harmonics in sine phase and even harmonics in cosine phase. Harmonic complex tones composed of harmonic components 1–9, 2–10, 3–11, 4–12, 6–14, 8–16, and 12–20 were used. Several neurons were also tested using Schroeder phase stimuli (Schroeder, 1970).

Acoustic pulse trains. Each pulse was generated by windowing the preferred carrier signal (pure tones) by a Gaussian envelope. In this report, we refer to this type of acoustic pulse train as a Gaussian click train. In a limited number of cases, we used other types of acoustic pulse trains, including rectangular clicks and Gaussian windowed acoustic pulses with a broadband noise carrier. Pulse widths ranged from 0.1 to 1 ms for rectangular clicks, and Gaussian pulses had a σ ranging from 0.89 to 2.53 ms.

Regular acoustic pulse trains had envelope repetition rates near the preferred fundamental frequency and had interpulse intervals (IPIs) equal to the inverse of the fundamental frequency. Irregular acoustic pulse trains were created by randomly jittering each IPI by a random number generated from a uniform distribution between (J_a, J_b), where J_a = IPI − IPI × (maximum jitter) and J_b = IPI + IPI × (maximum jitter). The maximum jitter was varied between 5 and 50% in 5% steps. Thus, for a mean IPI of 10 ms and a maximum jitter of 10%, each IPI would be chosen from a uniform distribution spanning values between 9 and 11 ms. For a given jitter amount, the temporal pattern of aperiodic acoustic pulses was the same across all trials.

Typically, 10 repetitions of each acoustic pulse train, complex tone, and rate level stimulus set were presented, but data with a minimum of five repetitions per stimulus were included in the analysis. Frequency tuning curves and rate-level functions were typically generated using 200-ms-long pure tone stimuli and interstimulus intervals >500 ms.

Behavioral experiments

The subjects in the behavioral experiments were four common marmosets (three male, one female) maintained at ~90% of their free-feeding weight on a diet consisting of a combination of monkey chow, fruit, and yogurt. This was a different subject group than was used for the electrophysiological experiments. Reward food consisted of a mixture of Similac baby formula, Gerber single-grain rice cereal, and strawberry-flavored Nesquik and was delivered via a syringe pump (model NE-500; New Era Pump Systems) mounted to the base of a custom restraint chair. Marmosets were tested while seated in the center of a single-walled sound-isolation chamber [model 400A (101 × 124 × 230 cm interior dimensions); Industrial Acoustic Company] lined with 3-inch acoustic absorption foam (Sonex; Pinta Acoustic). An infrared (IR) photo beam was positioned at the end of a feeding tube placed in front of the animal, and subject behavior was recorded when the animal licked at the feeding tube, which caused the IR beam to break.

All sound stimuli were generated offline using MATLAB software (MathWorks) and delivered at a nominal sampling rate of 100 kHz through a digital signal processor and programmable attenuator (models RX6 and PA5; Tucker-Davis Technologies), followed by an audio amplifier (model D-75; Crown Instruments). Stimuli were played from a loudspeaker (Arena series; frequency response from 80 Hz to 54 kHz; Tannoy) mounted 40 cm directly in front of the animal and were calibrated before the experiment using a ½-inch free-field microphone (type 4191; Brüel and Kjær) positioned in the chamber at the same location as the animal's head.

All animals were initially trained using operant conditioning techniques on a basic tone detection task in which they had to respond to a 7 kHz tone against a silent background (for a detailed description of this procedure, see Osmanski and Wang, 2011) before beginning the current experiment. This initial training took ~4–6 weeks for each animal, and, once trained, animals were moved on to a discrimination training task in which they had to detect a change in f_0 between a target and a repeating background tone (see below). Once trained on this discrimination task (~1 week), animals were moved on to the test sessions.

In test sessions, marmosets had to discriminate between same phase harmonic complex tones that differed by 1 octave in their f_0. Additionally, alternating phase complex tones were used to probe whether marmosets perceived these stimuli as similar to same phase harmonic complex tones with a fundamental frequency equal to f_0 or $2f_0$. Each stimulus consisted of the harmonics 4–12 of a common f_0 and was constructed so that the components either all began in cosine phase (COS stimulus) or were alternated such that the odd harmonics began in sine phase and the even harmonics began in cosine phase (ALT stimulus). These MF harmonic stimuli were presented with a 3-octave-wide band of noise (10 dB signal-to-noise ratio) centered at f_0 to minimize the effects of potential harmonic distortion products.

Each behavior session comprised 100 trials. A trial consisted of a variable duration (5–15 s) waiting period in which a background sound was repeatedly presented to the animal. This waiting period was followed by a 5 s response window within which a target sound was alternated with the background sound four times, which provided the animal with ample opportunity to detect and respond to the presentation of the target sound. All stimuli were 500 ms in duration (20 ms rise/fall time) with a 200 ms interstimulus interval and were presented at an average level of 70 ± 5 dB SPL. The intensity of each individual stimulus was roved to remove potential amplitude cues between targets and backgrounds (see below) and between stimuli containing different phase relationships among their components.

Stimuli were presented in blocks of 10 trials that contained seven target trials [five $2f_0$ COS targets and two f_0 ALT targets (ALT stimuli essentially functioned as low-probability probe sounds to see whether they were perceived in the same way as the higher-probability COS stimuli)] and three sham trials in which the target sound was identical to the background (f_0 COS). Licking the feeding tube during a sham trial was recorded as a "false alarm," which resulted in the chamber lights being extinguished for 5 s (a "blackout"). Licking the feeding tube during a target presentation was recorded as a "hit," while failing to respond to a target presentation during the 5 s response window was recorded as a "miss." At the end of each session, a corrected hit rate was calculated for each target based on high-threshold theory [$Pc^* = (Pc - \mathrm{FA})/(1 - \mathrm{FA})$, where FA is false alarm] (Gescheider, 1985). Final hit rate values for each target are based on a minimum of 300 trials total. Hit rates were then averaged across the two target types (i.e., ALT and COS stimuli). Sessions with a false alarm rate higher than 25% were discarded (<10% of all sessions).

Marmosets were tested using background COS stimuli with f_0 values of 150, 300, 450, 600, and 900 Hz ($2f_0$ COS targets thus had f_0 values ranging from 300 to 1800 Hz, respectively). Each f_0 was tested until a minimum of 300 trials was completed. Marmosets were then moved on to the next f_0. The order of f_0 values tested was randomly assigned to each subject.

Data analyses

Discharge rates >2 SDs above the mean spontaneous rate and more than one spike for 50% of the trials were considered significant. The mean spontaneous rate was subtracted from the discharge rate. A discharge rate less than zero was therefore below the mean spontaneous rate.

Criteria for pitch-selective neurons. Pitch selectivity was defined as a neuron that (1) responded to pure tones, (2) responded to MF sounds with a fundamental frequency near its BF, (3) did not respond significantly to components of the MF sound when they were played individually, and (4) the sound level of the MF sound (measured relative to the individual components) did not need to be >10 dB above the BF sound level threshold of the neuron to drive the neuron (Bendor and Wang, 2005). Firing rates were calculated using the entire stimulus duration. It is important to note that sensitivity to temporal envelope regularity was not used as part of the criteria for pitch selectivity.

The spectral sensitivity index is defined as $2 \times (R_{100\%} + R_{200\%})/(R_{50\%} + R_{100\%} + R_{150\%} + R_{200}) - 1$, where R is the discharge rate (spontaneous rate is not subtracted), and n% for $R_{n\%}$ was the percentage frequency shift of the complex tone relative to the fundamental frequency. The value was scaled by multiplying by two and subtracting one, so that values were 0 for no spectral cue preference (equal response for harmonic complex tones with a fundamental frequency equal to the best fundamental frequency (Bf_0) and odd-harmonic complex tones with a fundamental frequency equal to Bf_0/2) and 1 for a maximum spectral cue preference (neuron only responded to harmonic complex tones with a fundamental frequency equal to the Bf_o of the neuron.

The phase sensitivity index (PSI) is defined as $\mathrm{PSI} = (R_{\mathrm{COS}} - R_{\mathrm{ALT}})/(R_{\mathrm{COS}} + R_{\mathrm{ALT}})$, where R_{COS} is the firing rate of the neuron in response to the cosine phase complex tone, and R_{ALT} is the firing rate of the neuron in response to the alternating phase complex tone (spontaneous rate is not subtracted). Both complex tones had the same fundamental frequency. A PSI of 0 indicated that the firing rate for the alternating phase complex tone was equal to the firing rate of the cosine phase complex tone. If the neuron only responded to the cosine phase complex tone and had no response to the alternating phase complex tone, the PSI was 1. The spontaneous firing rate was not subtracted from the firing rate so that the values of the PSI were between −1 and 1. If a neuron did not respond to a complex tone of a particular harmonic composition, for either the cosine or alternating phase condition, this response was not included in the population analysis in Figure 4.

Normalized responses for the pitch-shifting experiment were calculated by dividing the firing rate for each frequency shifted complex tone by the response to a 0% shifted complex tone (harmonics 1–3).

Results

To investigate whether the encoding of pitch relies on temporal envelope or spectral cues, we examined whether marmosets perceived similar pitches for harmonic complex tones with two different phase relationships, an acoustic signal manipulation allowing us to delineate the contributions of these two types of cues. The fundamental frequency of a harmonic complex tone with all of its harmonics in cosine phase (referred hereafter as a COS stimulus) is the same whether it is measured using the temporal envelope periodicity or spectral periodicity of the acoustic signal (frequency spacing between harmonics) (Fig. 1*A*,*B*). However, a harmonic complex tone with its odd harmonics in sine phase and its even harmonics in cosine phase (referred hereafter as an ALT stimulus) has an envelope repetition rate that is double that of a COS harmonic tone (i.e., the fundamental frequency based on temporal envelope cues is one octave higher) (Fig. 1*A*,*C*). This phase manipulation does not change the spacing between harmonics, and so the f_0 measured using spectral cues is the same between these two stimuli. Additionally, a COS harmonic complex tone with a fundamental frequency equal to f_0 has the same envelope repetition rate (and pitch derived from temporal envelope cues) as an ALT complex tone with a fundamental frequency equal to $f_0/2$ (Fig. 1*B*,*F*). Because this acoustic stimulus manipulation allows us to independently vary the frequency spacing between harmonics and the repetition rate of the envelope, we can verify whether marmosets extract pitch using temporal envelope and/or spectral cues (Fig. 1*A*–*F*). In humans, an alternating phase harmonic complex tone has a pitch 1 octave higher than a same phase complex tone with an identical f_0 when all the harmonics are unresolved (Shackleton and Carlyon, 1994). The perceived pitch of a harmonic complex tone composed of only resolved harmonics does not depend on temporal envelope cues and is therefore identical for same and alternating phase complex tones.

Behavioral measurements of pitch discrimination in marmosets

Marmosets were tested in two daily sessions of 100 trials each. A trial consisted of a repeating "background" sound (i.e., an f_0 COS stimulus) that was alternated with one of two "target" sounds: (1) a $2f_0$ COS stimulus or (2) an f_0 ALT stimulus. The animal's task was to detect the presentation of the target sound and respond by licking a feeding tube for access to food reward. If marmosets

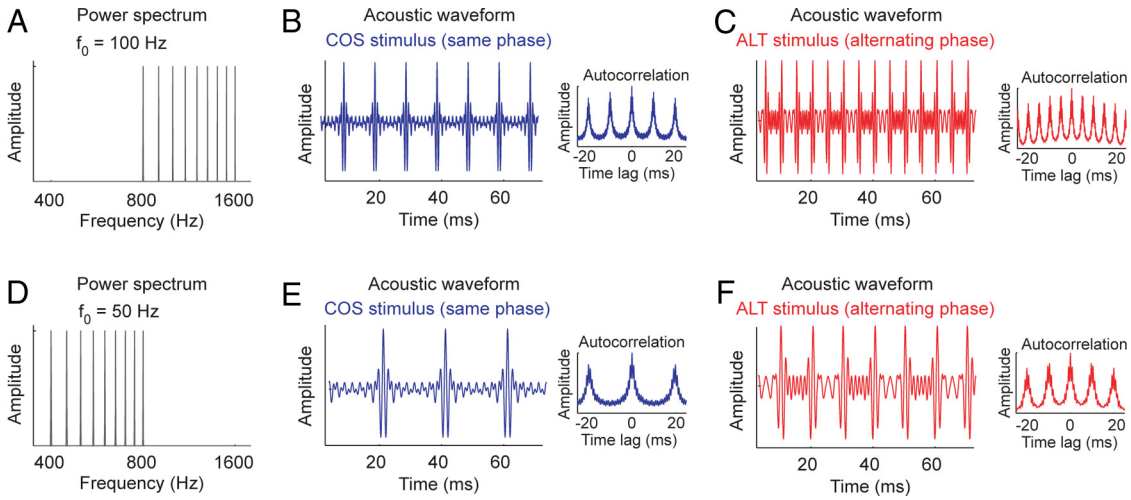

Figure 1. Harmonic complex tones with same or alternating phase. Harmonic complex tones with fundamental frequencies 1 octave apart ($f_0 = 100$ Hz and $f_0 = 50$ Hz in this example, harmonics 8–16). For each fundamental frequency, harmonics are either in same (COS) or alternating (ALT) phase (odd harmonics in sine phase, even harmonics in cosine phase). Alternating phase complex tones have twice the envelope repetition rate as their same phase counterpart, and the autocorrelation of the half-wave-rectified acoustic signal indicates a periodicity at twice the fundamental frequency. ***A***, Amplitude spectrum of a harmonic complex tone with $f_0 = 100$ Hz. ***B***, Acoustic waveform and autocorrelation plots (inset) for a same phase harmonic complex (COS stimulus) with $f_0 = 100$ Hz. ***C***, Acoustic waveform and autocorrelation plots (inset) for an alternating phase harmonic complex (ALT stimulus) with $f_0 = 100$ Hz. ***D***, Amplitude spectrum of a harmonic complex tone with $f_0 = 50$ Hz. ***E***, Acoustic waveform and autocorrelation plots (inset) for a COS stimulus with $f_0 = 50$ Hz. ***F***, Acoustic waveform and autocorrelation plots (inset) for an ALT stimulus with $f_0 = 50$ Hz.

used only spectral information to derive fundamental frequency, they would detect a pitch change for the $2f_0$ COS stimulus but not the f_0 ALT stimulus in this task. However, if marmosets used temporal envelope information to derive fundamental frequency, they would detect a pitch change for both stimuli.

Marmosets were able to easily discriminate $2f_0$ COS targets from f_0 COS background stimuli, showing highly consistent hit rates (>95% correct) across all f_0 values tested ($F_{(4,12)} = 0.894$, $p = 0.50$) (Fig. 2*A*). Performance with f_0 ALT targets, however, showed a significant decrease in discrimination ability as f_0 increased ($F_{(4,12)} = 34.484$, $p < 0.001$) (Fig. 2*A*). Discrimination was statistically equivalent for both f_0 ALT and $2f_0$ COS targets when the background f_0 was 150 Hz ($t_{(3)} = -2.74$, $p = 0.07$, Bonferroni's corrected) (Fig. 2*A*), but discrimination of f_0 ALT targets became increasingly different from $2f_0$ COS targets as f_0 increased (300 Hz, $t_{(3)} = -4.37$, $p < 0.05$; 450 Hz, $t_{(3)} = -3.89$, $p < 0.05$; 600 Hz, $t_{(3)} = -12.29$, $p < 0.01$; 900 Hz, $t_{(3)} = -17.25$, $p < 0.001$, Bonferroni's corrected). In general, performance dropped below 50% correct at f_0 values above 450 Hz (Fig. 2*A*), suggesting that the marmosets relied more strongly on spectral cues compared with temporal envelope cues as we increased the f_0 of the complex tone above 450 Hz.

To look at the influence of harmonic order on these behavioral results, we also tested the same four subjects on a second set of harmonic complex tones with f_0 values of either 150 or 900 Hz and the harmonic components confined to one of three distinct spectral regions: low (harmonics 1–4), medium (harmonics 5–8), or high (harmonics 9–12). All medium and high harmonics stimuli were presented with a 3-octave-wide band of noise (10 dB signal-to-noise ratio) centered at f_0 to minimize the influence of cochlear distortion products at the fundamental frequency. However, other non-f_0 distortion products (e.g., $2f_1$–f_2) were above the frequency range of this noise masker and not blocked. Therefore, it is possible that, although a complex tone was composed entirely of unresolved components, additional resolved harmonic components were produced by distortion products and may have influenced our results (Oxenham et al., 2009).

Each marmoset was pseudorandomly assigned to one f_0 and one spectral region. After completing 300 trials at one spectral region, a subject was moved to the next spectral region. Marmosets were moved on to the second f_0 only after all three spectral regions at the first f_0 were tested.

Results for the 150 and 900 Hz f_0 conditions are shown in Figure 2, *B* and *C*, respectively. As expected from the previous experiment, marmosets on this task easily discriminated $2f_0$ COS targets from f_0 COS backgrounds (>90% correct) in both 150 and 900 Hz f_0 conditions (Fig. 1*B*,*C*, blue curves). Hit rates did not significantly change across any of the spectral regions tested at either 150 Hz f_0 ($F_{(2,6)} = 1.52$, $p = 0.29$) or 900 Hz f_0 ($F_{(2,6)} = 0.952$, $p = 0.45$). However, the performance of the marmosets decreased when the targets changed to f_0 ALT stimuli. At the 150 Hz f_0 condition, the hit rate was well below 50% in the low spectral region (Fig. 2*B*), suggesting that the marmosets did not use the envelope repetition rate when accomplishing this task when the f_0 component was present in the target. The average hit rate increased as spectral region moved from low to medium and high ($F_{(2,6)} = 21.54$, $p < 0.01$). Data in Figure 2*B* show that the f_0 ALT targets were much harder to discriminate compared with $2f_0$ COS targets at both low ($t_{(3)} = -9.27$, $p < 0.01$, Bonferroni's corrected) and medium ($t_{(3)} = -8.54$, $p < 0.01$, Bonferroni's corrected) spectral region conditions, although the two target types were indistinguishable in the high spectral region condition ($t_{(3)} = -1.96$, $p = 0.15$, Bonferroni's corrected). Conversely, in the 900 Hz f_0 condition, performance for f_0 ALT targets was significantly worse than $2f_0$ COS targets at all three spectral regions (low, $t_{(3)} = -8.21$, $p < 0.01$; medium, $t_{(3)} = -16.06$, $p < 0.01$; high, $t_{(3)} = -6.79$, $p < 0.01$, Bonferroni's corrected) (Fig. 2*C*). The average hit rates for f_0 ALT targets never rose above 30% and showed little variation across spectral region ($F_{(2,6)} = 1.23$, $p = 0.36$), suggesting that the marmosets did not use the temporal envelope cues to accomplish this task for high f_0 (900 Hz).

According to the behavioral data shown in Figure 2, the transition point in f_0 for detecting a pitch change in an ALT harmonic complex tone (harmonics 4–12) relative to a COS harmonic

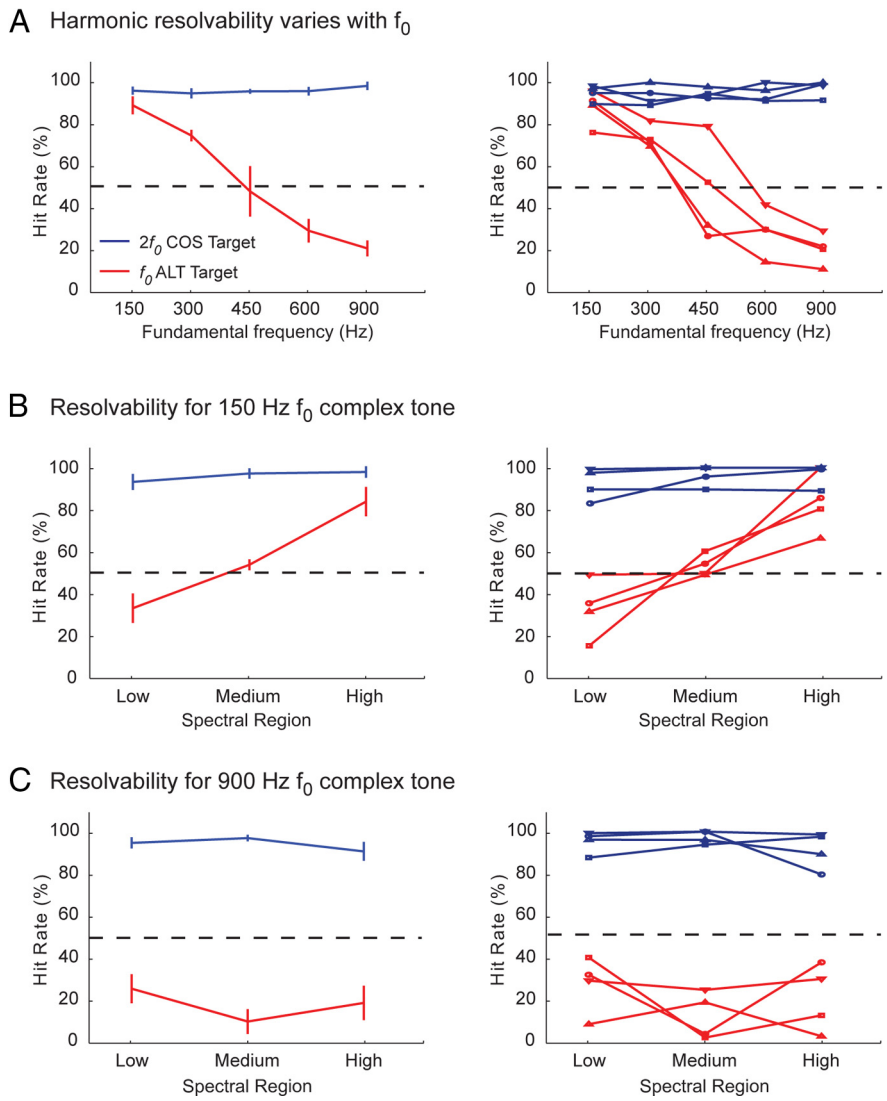

Figure 2. Harmonic complex discrimination performance in marmoset. The left column shows the group average data, and the right column shows data from individual marmosets (each line represents the performance of 1 individual marmoset). ***A***, Results of a behavioral discrimination task using nine-component harmonic complex tone stimuli in which the first three harmonics were removed (harmonics 4–12). Marmosets were trained to discriminate both an f_0 ALT (red lines) and $2f_0$ COS (blue lines) harmonic complex tone stimulus from a repeating f_0 COS stimulus. Sounds were presented with a 3-octave-wide noise band (signal-to-noise ratio of 10 dB) centered at f_0 to minimize harmonic distortion effects. We measured the percentage of times marmosets correctly identified presentation of a target sound (100% = subjects respond every time a target or probe is presented; 0% = subjects never respond). Results show that marmosets can use the envelope repetition rate to correctly discriminate an f_0 ALT stimulus against f_0 COS stimulus at low f_0 values. Discrimination performance falls off at higher f_0 values with a 50% crossover ~450 Hz. ***B***, Results of a behavioral task examining the role of resolved versus unresolved harmonics on marmoset pitch perception. Animals had to discriminate four-component harmonic complex tone stimuli with an f_0 of 150 Hz at three different spectral regions. Discrimination was poor for ALT stimuli (red lines) but not COS stimuli (blue lines) at the lowest spectral region. Performance improved for ALT phase stimuli at higher spectral regions. ***C***, A second behavioral test of the role of resolvability on marmoset pitch perception. Animals had to discriminate stimuli with an f_0 of 900 Hz at the same three spectral regions. Discrimination was poor for ALT stimuli (red lines) but not COS stimuli (blue lines) at all spectral regions. Different markers indicate individual animals.

complex tone was ~450 Hz (Fig. 2*A*). This finding suggests that, in marmosets, the spectral resolvability of harmonic components in a complex tone depends on both harmonic order and f_0. In humans, harmonic order is directly linked to spectral resolvability; the relationship with f_0 is less clear. The first five to eight harmonics of a complex tone are resolved, and harmonic resolvability is only weakly dependent on frequency (Plomp, 1964). However, these psychophysical measurements were based on the ability to hear out individual components rather than measuring the interactions between components (temporal envelope cues). At lower f_0 values (below 100 Hz), temporal envelope cues dominate the perceived pitch of alternating polarity click trains (Flanagan and Guttman, 1960). Although previous measurements of human cochlear tuning reported a relatively constant bandwidth across frequency channels (Glasberg and Moore, 1990), more recent measurements suggest a sharpening of tuning as frequency increases (from a Q_{ERB} of 10 at 1 kHz to a Q_{ERB} of 20 at 8 kHz) (Oxenham and Shera, 2003).

Temporal envelope information processing by pitch-selective neurons

Pitch-selective neurons were observed previously in a low-frequency region of marmoset auditory cortex, bordering primary auditory cortex, the rostral field, and lateral belt (Fig. 3*A*), (Bendor and Wang, 2005, 2010). Pitch-selective neurons were identified using three criteria: (1) they had pure-tone responses (Fig. 3*B*), (2) they responded to MF sounds at sound levels in which estimated combination tone responses were below threshold (Fig. 3*C*), and (3) they had no response to the individual harmonics of the MF sound, indicating that the MF sound is outside the excitatory frequency response area of the neuron (Fig. 3*D*). MF responses have not been observed outside of this pitch region, in neighboring primary auditory cortex (Schwarz and Tomlinson, 1990; Fishman et al., 1998, 2000). In addition to MF responses, pitch-selective neurons have also been reported previously to be sensitive to temporal envelope regularity, have similar pitch tuning for spectrally different sounds (Fig. 3*B*,*E*), and show sensitivity to pitch salience (Fig. 3*F*,*G*) (Bendor and Wang, 2005, 2010).

Using high-impedance tungsten electrodes (see Materials and Methods), we recorded 74 well-isolated pitch-selective neurons in four marmosets, with 68 pitch-selective neurons localized within the putative pitch region. Four of six of the remaining pitch neurons were found near the border of the putative pitch region. A total of 203 neurons were recorded in the pitch center, of which 157 were tone responsive. Thus, 33.5% of the total neurons encountered were classified as pitch selective (43% of tone-responsive neurons). Here, we present data from 28 pitch-selective neurons that were further studied in the four subjects to examine the sensitivity of pitch responses to temporal envelope and spectral cues. Because of the limited recording time in each neuron, we were not able to examine every pitch-selected neuron for their temporal envelope and spectral cue sensitivity.

Given the quantitative differences between marmoset and human pitch perception in terms of how the temporal envelope and

spectral information is used, we next examined the responses of pitch-selective neurons to the COS and ALT harmonic complex tone stimuli used in the marmoset behavioral experiments (Fig. 2), to examine whether these differences were also reflected in responses of pitch-selective neurons.

We observed that pitch-selective neurons differed in their responses to ALT and COS stimuli, shifting their f_0 tuning for ALT stimuli down by 1 octave compared with COS stimuli (Fig. 4*A*,*B*). We tested pitch-selective neurons with both COS and ALT stimuli in several harmonic compositions. The fundamental frequency of these harmonic complex tones was selected to be at the preferred fundamental frequency of the pitch-selective neuron (Bendor and Wang, 2005), referred to hereafter as the Bf_0. An example pitch-selective neuron is shown in Figure 4*C*. Firing rates were higher for COS stimuli with f_0 values equal to the Bf_0 of the pitch-selective neuron (152 Hz) than for their ALT counterpart, except for the highest-order harmonic composition (8–16) that did not elicit a significant response for either stimuli (Fig. 4*C*, COS, solid blue curve; ALT, solid red curve). This neuron did not respond to COS stimuli with f_0 values set at an octave below its Bf_0 (Fig. 4*C*, dashed blue curve), but when tested by ALT stimuli with f_0 values set at Bf_0/2, this pitch-selective neuron showed higher firing rates than its COS counterpart for higher-order harmonic compositions (4–12 and 8–16) (Fig. 4*C*, dashed red curve). Thus, temporal envelope cues (preference for the envelope repetition rate to equal to the Bf_0) affected the pitch responses in this neuron when higher-order harmonics were present in the complex tones, whereas when lower-order harmonics were present (1–9), spectral cues (preference for the f_0 to equal the Bf_0) dominated the pitch response. Large variations in firing rate between COS and ALT stimuli were more common in pitch-selective neurons with low Bf_0 values (Fig. 4*C*, Bf_0 = 152 Hz) compared with pitch-selective neurons with high Bf_0 values (Fig. 4*D*, Bf_0 = 1.45 kHz).

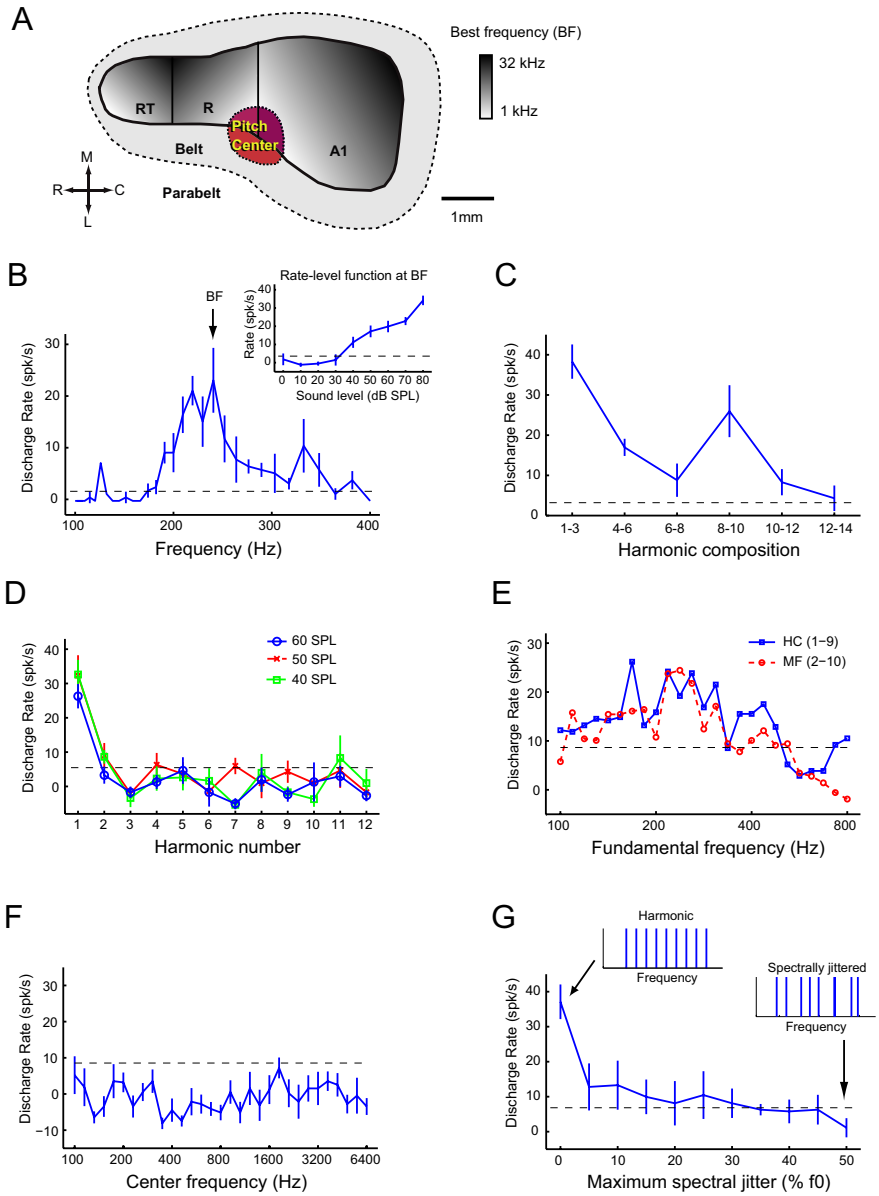

Figure 3. Criteria for pitch-selective neurons. ***A***, Schematic diagram of marmoset auditory cortex with the location of the pitch-selective region indicated in low-frequency auditory cortex (red). A1, Primary auditory cortex; R (on diagram), rostral field; RT, rostrotemporal field; C, caudal; L, lateral; M, medial; R (in direction scale), rostral. ***B–G***, Responses from pitch-selective neuron M2P-U201.1. The dashed line indicates a significant firing rate (2 SDs above mean spontaneous rate). Mean spontaneous rate is subtracted from all displayed discharge rates. ***B***, Frequency tuning to pure tones (40 dB SPL). The BF of 240 Hz is indicated on the plot. Inset plot displays the rate-level function for a pure tone at the BF. ***C***, Response to harmonic complex tones (f_0 = BF, 3 harmonics, cosine phase, individual components at 40 dB SPL) varying in their harmonic composition. All stimuli are MF sounds except the complex tone composed of harmonics 1–3. ***D***, Response to harmonics 1–12 of an f_0 = BF, at 40, 50, and 60 dB SPL. Harmonic 1 is equal to the f_0. ***E***, Fundamental frequency tuning for two harmonic complex (HC) tones [harmonics 1–9 (blue solid) and 2–10 (red dashed), individual components at 30 dB SPL]. ***F***, Spectral tuning to the center frequency of half-octave bandpassed noise stimulus at 50 dB SPL. ***G***, Sensitivity to spectral jitter of the harmonics of a complex tone. Harmonic complex tones have 0% jitter (left inset). Spectral jitter was added by shifting each harmonic by a random amount ranging between 0 and the maximum spectral jitter, proportional to the fundamental frequency (right inset at 50% jitter). The original harmonic complex tone had an f_0 = BF, contained harmonics 4–12, with each component at 30 dB SPL.

We quantified the change in firing rate between COS and ALT stimuli for a population of pitch-selective neurons using the PSI (see Materials and Methods). Because we compared responses to two acoustic stimuli that were spectrally identical, any difference in firing rate was only attributable to differences in the phase relationship between harmonics (which in turn affected the envelope repetition rate). If a pitch-selective neuron was sensitive only to the envelope repetition rate of the acoustic stimulus, the PSI would be 1 for the COS and ALT stimuli at Bf_0 and −1 at Bf_0/2. Conversely, a PSI of 0 indicated that only spectral information was used, because there was no difference in firing rate between the two phase conditions (COS vs ALT).

We observed that the PSI was significantly greater for pitch-selective neurons with a Bf_0 < 450 Hz compared with those with a higher Bf_0 (Fig. 4*E*). This effect was statistically significant when

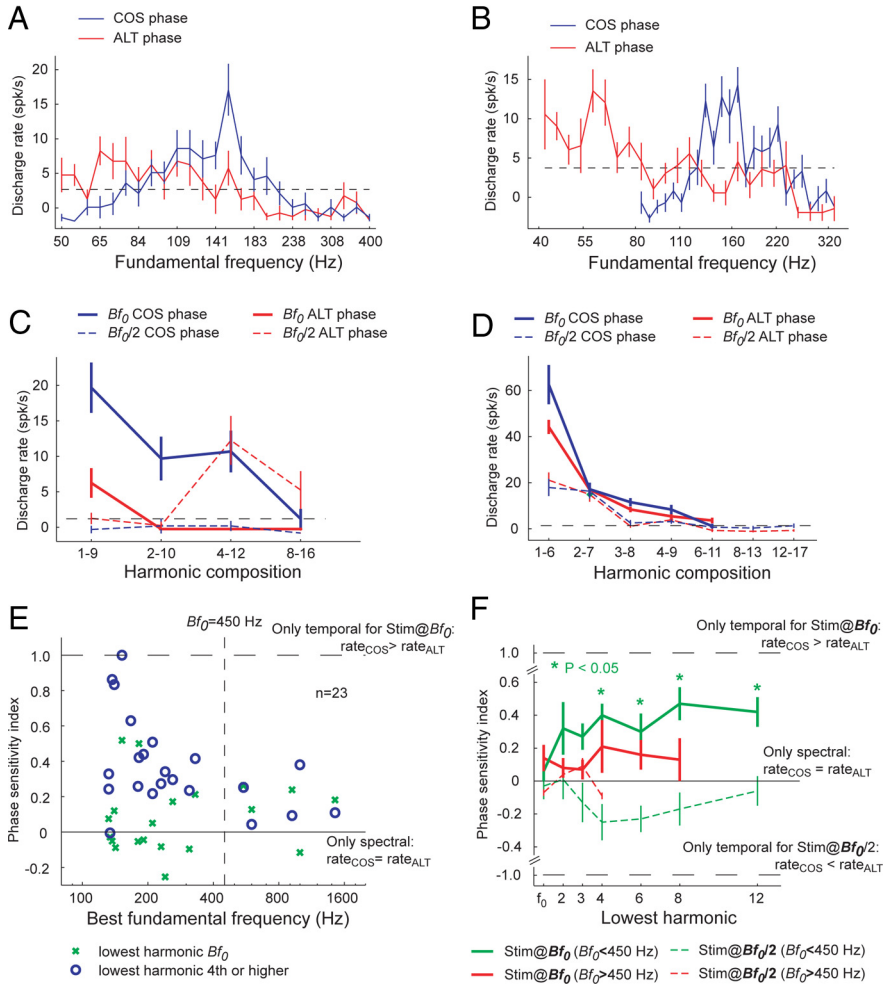

Figure 4. Phase sensitivity experiment in pitch-selective neurons. ***A***, ***B***, Tuning curves of two pitch-selective neurons in response to harmonic complex tones varying in fundamental frequency. COS (blue) and ALT (red) complex tones were used. For these two neurons, the Bf_0 for ALT stimulus is ~1 octave below that of COS stimulus. ***C***, An example of a pitch-selective neuron (unit M36N-534.1) that is sensitive to changes in the phase of harmonics in a complex tone. The Bf_0 of this neuron is 152 Hz. Four types of stimuli were tested in four harmonic compositions (1–9, 2–10, 4–12, and 8–16): COS stimulus at Bf_0 (solid blue curve), COS stimulus at $Bf_0/2$ (dashed blue curve), ALT stimulus at Bf_0 (solid red curve), and ALT stimulus at $Bf_0/2$ (dashed red curve). The dashed line indicates the minimum firing rate criterion for a significant response (2 SDs above the mean spontaneous rate). ***D***, An example of a pitch-selective neuron (unit M32Q-44.1) that is less sensitive to changes in the phase of harmonics in a complex tone. The Bf_0 of this neuron is 1.45 kHz. Four types of stimuli were tested in seven harmonic compositions. Format is the same as in ***A***. ***E***, The PSI is plotted against Bf_0 for a population of pitch-selective neurons. The PSI was calculated for complex tones with low-order harmonics (containing the fundamental) and for complex tones with higher-order harmonics (harmonic 4 or higher). Vertical dashed line indicates fundamental frequency of 450 Hz. The two horizontal lines indicate a PSI of 0 and 1, respectively. ***F***, Mean PSI as a function of the lowest harmonic for pitch-selective neurons. Only for $Bf_0 < 450$ Hz were PSIs significantly different from 0 for harmonic compositions, with the lowest harmonics being the 4, 6, 8, or 12 (signed-rank test, $p < 0.05$, Bonferroni's corrected) when tested at a the Bf_0 of a neuron (solid green curve). The dashed horizontal lines indicate a PSI of 1 and −1, respectively.

lower-order harmonics (harmonics 1–3) were not present in the complex tone [median PSI = 0.34 ($Bf_0 < 450$ Hz) and 0.11 ($Bf_0 > 450$ Hz); mean PSI = 0.43 ($Bf_0 < 450$ Hz) and 0.14 ($Bf_0 > 450$ Hz); Wilcoxon's rank-sum test, $p < 0.05$] (Fig. 4*E*, circles). However, when the fundamental (first harmonic) was present in the stimuli, there was no longer a statistically significant difference in PSI between pitch-selective neurons with $Bf_0 < 450$ Hz and $Bf_0 > 450$ Hz [median PSI = −0.03 ($Bf_0 < 450$ Hz) and 0.18 ($Bf_0 > 450$ Hz); mean PSI = 0.06 ($Bf_0 < 450$ Hz) and 0.14 ($Bf_0 > 450$ Hz); Wilcoxon's rank-sum test, $p = 0.29$] (Fig. 4*E*, crosses).

For pitch-selective neurons with $Bf_0 < 450$ Hz, we observed PSIs significantly greater than 0 (signed-rank test, $p < 0.05$, Bonferroni's corrected) when higher-order harmonics (fourth or higher) were present in harmonic complex tones played at the Bf_0 of the pitch-selective neuron (Fig. 4*F*, solid green curve). Although the mean PSI was positive for pitch-selective neurons with $Bf_0 > 450$ Hz, it was not significantly different from 0 (signed-rank test, $p > 0.05$, Bonferroni's corrected) (Fig. 4*F*, solid magenta curve). When harmonic complex tones with f_0 values 1 octave below the Bf_0 of the pitch-selective neuron were used, we observed negative PSIs (Fig. 4*F*, dashed curves). This indicates that the Bf_0 of a harmonic complex tone did not need to match the Bf_0 of a pitch-selective neuron for temporal envelope repetition rate to influence the response of the neuron (as long as the repetition rate matched the Bf_0 of the neuron). This trend was only observed in pitch-selective neurons with $Bf_0 < 450$ Hz, but the mean PSI was not significantly different from 0 (signed-rank test, $p > 0.05$, Bonferroni's corrected) (Fig. 4*F*, dashed green curve). These data show that temporal envelope information is used by pitch-selective neurons with $Bf_0 < 450$ Hz when higher-order harmonics are present in the harmonic complex tone, whereas pitch responses depend more on spectral information for neurons with $Bf_0 > 450$ Hz.

Spectral information processing by pitch-selective neurons

Although changing the phase of harmonics can alter the temporal envelope cues used for pitch extraction, this stimulus manipulation does not allow us to modify the spectral cues used without changing the fundamental frequency. To more closely examine the influence of spectral cues on pitch processing, we next investigated whether pitch-selective responses were sensitive to the absolute frequency of the components of a complex tone using a different acoustic stimulus set. We used frequency-shifted complex tones, for which each component was parametrically shifted in frequency by an amount proportional to f_0 (Fig. 5*A*). For shifts that were multiples of 100%, the resulting complex tones were harmonic such that each component was an integer multiple of f_0. For all other shifts, the resulting complex tones were either inharmonic (components were not integer multiples of the f_0) or harmonic with only odd-number components (effectively decreasing the f_0 1 octave lower). For example, a harmonic complex tone with an f_0 = 100 Hz and three harmonics (1–3) would have components at frequencies 100, 200, and 300 Hz. A frequency shift of 25% relative to the f_0 for each component (125, 225, and 325 Hz) would create an inharmonic complex tone, with none of the components having a frequencies that were integer multiples of the f_0. A frequency shift of 50% relative to the f_0 for each component (150, 250,

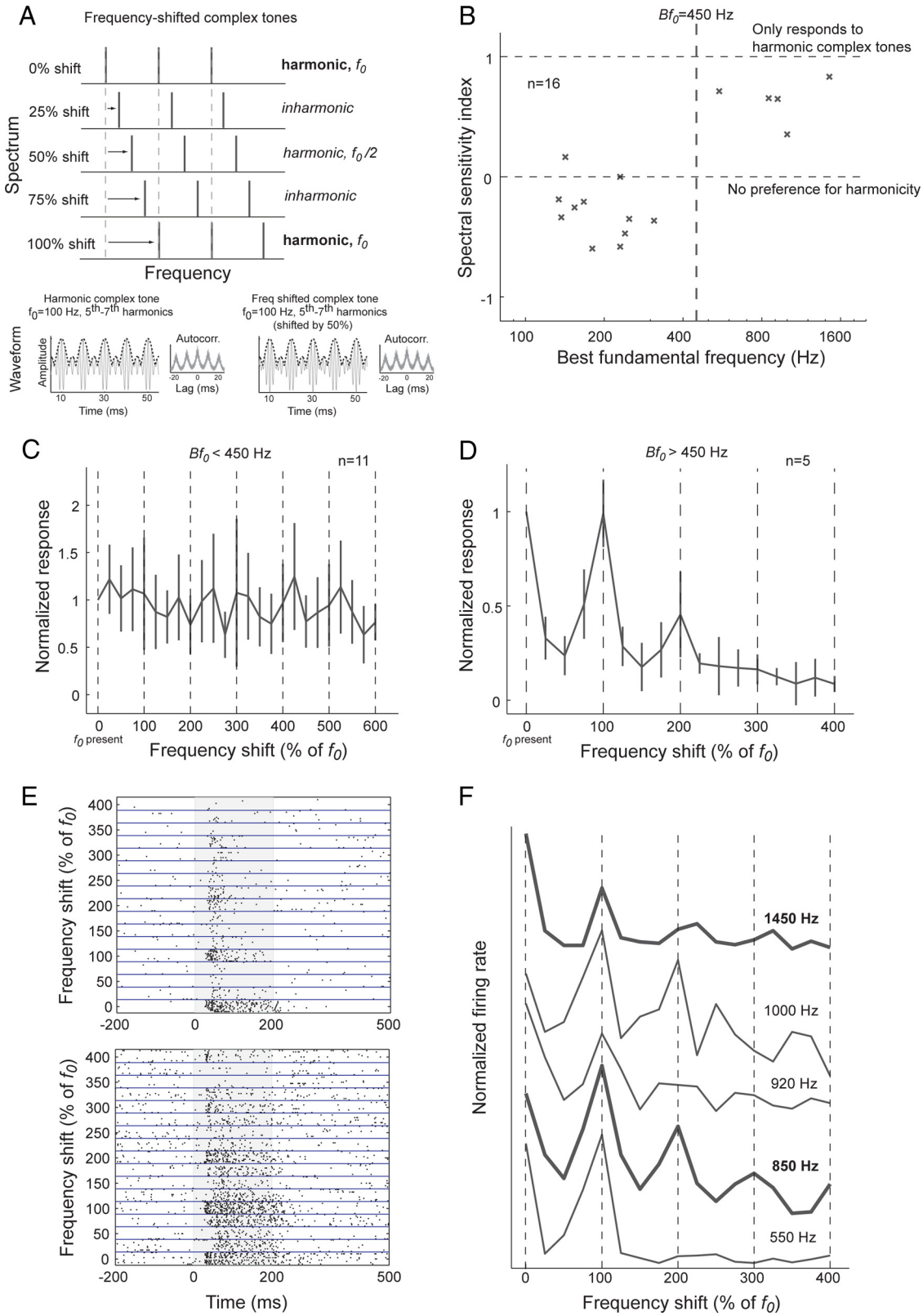

Figure 5. Experiment with frequency-shifted complex tones. ***A***, Each harmonic of a three-component complex tone was shifted in frequency by an amount relative to the fundamental frequency. Thus, a 25% frequency shift in a complex tone (with $f_0 = 100$ Hz) would be a 25 Hz increase in frequency for each harmonic, resulting in an inharmonic complex tone. Frequency shifts of multiples of 100% of f_0 resulted in harmonic complex tones (components were integer multiples of the fundamental frequency). Harmonic complex tones ($f_0 = 100$ Hz, harmonics 5–7, left bottom plot) and odd-harmonic complex tones ($f_0 = 100$ Hz, harmonics 5–7 shifted by 50%, right bottom plot) differed slightly in their waveforms but had similar envelope periodicity (dashed line). The autocorrelations of the half-wave-rectified waveforms of these complex tones were also similar (see insets of each plot). ***B***, Spectral sensitivity of pitch-selective neurons. Vertical dashed line indicates $Bf_0 = 450$ Hz. The horizontal dashed lines indicate a spectral sensitivity index of 1 and 0, respectively. ***C***, The mean normalized response of pitch-selective neurons with $Bf_0 < 450$ Hz to frequency-shifted complex tones. Vertical dashed lines indicate shifts of multiples of 100% of f_0 (i.e., harmonic complex tones). Normalized responses were calculated by dividing the discharge rate by the maximum response of the neuron to the stimulus set. ***D***, Normalized response pitch-selective neurons with $Bf_0 > 450$ Hz to frequency-shifted complex tones. (*Figure legend continues.*)

and 350 Hz) would create an odd-harmonic complex tone because these three components are harmonics 3, 5, and 7 of the harmonic complex tone with an $f_0 = 50$ Hz (1 octave below an f_0 of 100 Hz). In these experiments, the frequency separation between components was equal to f_0 for all acoustic stimuli, and, as a result, the envelope repetition rate was always equal to f_0 despite our spectral manipulations. In humans, pitch shifts or ambiguous pitches are perceived when the components of the inharmonic complex tone are resolved (Patterson and Wightman, 1976; Moore and Moore, 2003). Conversely, when the components are unresolved, inharmonic complex tones have minor or negligible effects on the perceived pitch, and any observed pitch shifts can be attributed to shifts in the spectral "center of gravity" (Moore and Moore, 2003). Although these data can be modeled using a harmonic template in the auditory system for pitch extraction, similar predictions have also been obtained using temporal fine structure cues (Meddis and O'Mard, 1997).

Because the peak firing rate of a pitch-selective neuron occurs at its Bf_0 and decreases when the pitch salience decreases, a pitch shift or ambiguous pitch should be reflected by a decrease in firing rate only when the harmonics are resolved. If the harmonics are not resolved, then only the temporal envelope pitch information remains, and pitch-selective neurons should be insensitive to these frequency shifts. As such, pitch-selective neurons should have lower firing rates for inharmonic complex tones and odd-harmonic complex tones compared with harmonic complex tones when the components are resolved. Pitch is more discriminable for odd-harmonic complex tones compared with inharmonic complex tones (Micheyl et al., 2012), but odd-harmonic complex tones also produce the largest change in f_0 and greatest pitch ambiguity (de Boer, 1956). If pitch-selective neurons are sensitive to fundamental frequency (measured using spectral cues), they should show the largest decrease in firing rate for odd-harmonic complex tones. We calculated the spectral sensitivity index (see Materials and Methods) of pitch-selective neurons, which equaled 1 when neurons only responded to harmonic complex tones (even and odd components) and equaled 0 when neurons responded similarly to odd-harmonic complex tones (shift of 50%, 150%, etc.) and harmonic complex tones with all harmonics (shift of 0%, 100%, etc.). The spectral sensitivity index was significantly greater in pitch-selective neurons with B$f_0 > 450$ Hz compared with those with lower Bf_0 values (Fig. 5*B*) [median = −0.34 (B$f_0 < 450$ Hz) and 0.65 (B$f_0 > 450$ Hz); mean = −0.29 (B$f_0 < 450$ Hz) and 0.64 (B$f_0 > 450$ Hz); Wilcoxon's rank sum test, $p < 5 \times 10^{-4}$]. Pitch-selective neurons with B$f_0 < 450$ Hz showed no decrease in their normalized responses to frequency-shifted complex tones (Fig. 5*C*), whereas every pitch-selective neuron in our dataset with a B$f_0 > 450$ Hz showed a preference for harmonic over both inharmonic complex tones and odd-harmonic complex tones (Fig. 5*D–F*). These data indicate that spectral cues are used by pitch-selective neurons with B$f_0 > 450$ Hz, whereas pitch responses in neurons with B$f_0 < 450$ Hz are less sensitive to spectral cues.

←

(*Figure legend continued.*) Vertical dashed lines indicate shifts of multiples of 100% of f_0 (i.e., harmonic complex tones with a fundamental frequency equal to the Bf_0 of the neuron). ***E***, Raster plots of two pitch-selective neurons with $Bf_0 > 450$ Hz responding to frequency-shifted complex tones. Top, Unit M32Q-44.1, $Bf_0 = 1450$ Hz. Bottom, Unit M2P-113.1, $Bf_0 = 850$ Hz. ***F***, Responses of five pitch-selective neurons ($Bf_0 > 450$ Hz) to harmonic and inharmonic complex tones (from top to bottom: unit M32Q-44.1, unit M32Q-35.1, unit M41O-276.1, unit M2P-113.1, and unit M32Q-109.1). Vertical dashed lines indicate harmonic complex tones (100% shifts). The Bf_0 of each neuron is indicated next to each curve. The neurons are ordered from bottom to top by increasing Bf_0. The bold curves indicate the two example neurons shown in ***E***.

Pitch-related responses within the putative pitch center

In addition to pitch-selective neurons that passed our criteria for pitch selectivity (see Materials and Methods), we found evidence of two additional types of responses within the putative pitch center that could contribute to encoding pitch. We identified 10 neurons that were tuned to the repetition rate of the envelope of the acoustic signal but did not respond (or were weakly responsive) to pure tones with frequencies equal to the best envelope repetition rate or harmonic frequencies of this best repetition rate. Examples of two such neurons are shown in Figure 6, *A* and *C*. Responses of these two neurons to sinusoidally modulated (sAM) tones at their best modulation frequency occurred over a range of carrier frequencies (spectral range), an order of magnitude larger than their best envelope repetition rate, demonstrating some degree of spectral invariance (Fig. 6*B*,*D*). Furthermore, no preference for the harmonicity of a complex tone was observed (Fig. 6*B*,*D*). Because of the lack of response in these neurons to pure tones at f_0, they failed the criteria for pitch selectivity (Bendor and Wang, 2005). However, in all other respects, the MF responses of these neurons were similar to the "temporal envelope-based" response of pitch-selective neurons with B$f_0 < 450$ Hz. We tested a group of these neurons with temporally jittered acoustic pulse trains (with a mean envelope repetition rate equal to Bf_0), and all tested neurons showed a preference for temporally regular sounds (Fig. 6*E*). It has been shown that pitch salience covaries with temporal regularity (Pollack, 1968). We have shown previously that sensitivity to temporal regularity is typically observed in pitch-selective neurons; neurons outside the pitch center are generally insensitive to temporal jitter and have similar firing rates for regular and irregular pulse trains with the same mean envelope repetition rate (Bendor and Wang, 2010). Given that these neurons do not respond to pure tones at frequencies equal to their best envelope repetition rate, their responses to MF harmonic complex tones and sensitivity to temporal regularity cannot be a byproduct of distortion products produced in the cochlea (Pressnitzer and Patterson, 2001; McAlpine, 2004).

In addition to envelope repetition rate-tuned neurons, we also found seven neurons exhibiting multipeaked spectral tuning within the putative pitch center. These neurons were tuned to several frequencies that were harmonics of their Bf_0. Response properties of two multipeaked neurons are shown in Figure 7, *A–C* and *D–F*, respectively. When tested with complex tones, a multipeaked neuron typically showed significantly higher firing rates to a harmonic complex tone than to the linear summation of the responses to individual harmonics (Fig. 7*B*). In response to frequency-shifted complex tones (Fig. 7*B*) or sAM tones varying in carrier frequency (Fig. 7*E*), these neurons exhibited the strongest responses to harmonic acoustic stimuli (with a fundamental frequency equal to the Bf_0) compared with frequency shifts creating inharmonic or odd-harmonic acoustic stimuli. Furthermore, responses of these neurons to harmonic complex sounds were not affected by the phase relationship between harmonics when tested by COS, ALT, or Schroeder phase stimuli (Fig. 7*C*,*F*; see Materials and Methods). Both of these multipeaked neurons had B$f_0 > 450$ Hz and had responses that were similar to pitch-selective neurons with comparable Bf_0 (>450 Hz) that were more sensitive to spectral cues than temporal envelope cues. The only observed difference is that these units also responded to each harmonic when played individually as a pure tone (Fig. 7*A*,*D*),

which is why they failed to satisfy the criteria for pitch selectivity. The observed multipeaked tuning for harmonic frequencies creates a harmonic template that could theoretically extract the fundamental frequency using spectral cues.

Discussion

Various neural mechanisms have been proposed over the past century to explain how the auditory system extracts the pitch of an acoustic signal (for recent reviews, see Moore, 2003; Plack et al., 2005; Griffiths and Hall, 2012; Wang and Walker, 2012). Computational models relying entirely on a single mechanism are attractive given their ability to predict many pitch phenomena despite their simplicity (Meddis and Hewitt, 1991a,b; Meddis and O'Mard, 1997; Shamma and Klein, 2000; Plack et al., 2005). However, recent psychophysical data demonstrating pitch perception in human subjects for harmonic complex tones with harmonics above the assumed phase-locking limit cast doubt on a purely temporal model to extract pitch (Oxenham et al., 2011). Furthermore, differences in pitch perception have been observed for resolved and unresolved harmonic complex tones, which has lead some researchers to argue that dual-pitch processing mechanisms provide a more parsimonious explanation of these psychophysical data (Shackleton and Carlyon, 1994; Carlyon, 1998). One difficulty in assessing the success of a model is that better performance could be linked to the use by a model of more free parameters (Plack et al., 2005).

Here we report that neurons within the putative pitch center of auditory cortex in marmosets can use either temporal envelope or spectral cues in an acoustic signal for pitch extraction, depending on pitch values and harmonic compositions. Although some neurons appeared to encode pitch using only temporal envelope or spectral cues, other neurons were sensitive to the combination of temporal envelope and spectral cues (Fig. 4*C*). The major determinant of whether pitch was extracted using temporal envelope and/or spectral cues was the fundamental frequency and harmonic order of the complex tone. These observations from pitch-selective neurons were mirrored in the behavioral data from marmosets (Fig. 2), in which both fundamental frequency and harmonic order influence the contribution of spectral and temporal envelope information on the perceived pitch.

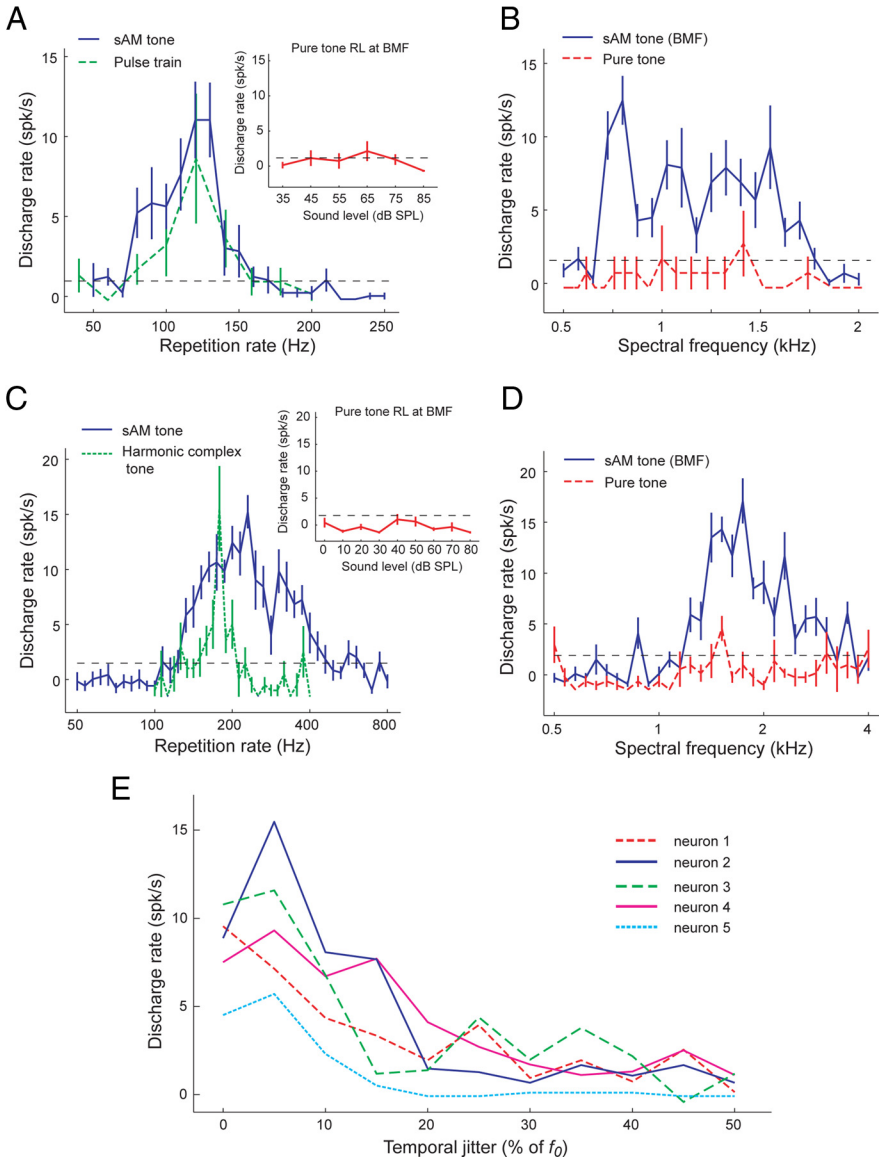

Figure 6. ***A***, ***B***, An example neuron (unit M2P-226.1) in the putative pitch center +tuned to the envelope modulation frequency (but not spectral frequency). ***A***, Similar tuning for sAM tones varying in modulation frequency (blue solid line) and pulse trains varying in envelope repetition rate (green dashed line). Both stimuli were played at 70 dB SPL. The rate-level (RL) function of the unit measured by pure tones at the best modulation frequency (BMF) of the neuron is shown in the inset. No significant response was observed at any sound level tested. ***B***, This neuron responded to sAM tones at its best modulation frequency over a wide range of carrier frequencies (blue solid line) but did not respond to pure tones over the same spectral range (red dashed line). The horizontal black dashed line indicates the minimum firing rate criterion for a significant response (2 SDs above the mean spontaneous rate). ***C***, ***D***, An example neuron (unit M410–309.1) that responded to MF harmonic complex sounds but did not respond to pure tones at its best modulation frequency and had weak or no responses to pure tones at higher frequencies. ***C***, Similar tuning for sAM tones varying in modulation frequency (blue solid line) and harmonic complex tones varying in fundamental frequency (green dashed line). The neuron did not respond to pure tones at its best modulation frequency across sound level (inset). ***D***, This neuron responded to sAM tones at its best modulation frequency over a wide range of carrier frequencies (blue solid line) but did not respond to pure tones over the same spectral range (red dashed line). ***E***, Responses to regular and irregular acoustic pulse trains by five modulation sensitive neurons in the putative pitch center. All five neurons showed decreased responses to temporally irregular pulse trains.

Our data indicate that two different mechanisms are used by the auditory system of the marmoset to extract pitch. The temporally based mechanism is sensitive to both the temporal regularity and repetition rate of the envelope of the acoustic signal. The spectrally based mechanism is sensitive to the harmonicity and f_0 of the harmonics of the complex tone but not to the envelope of the acoustic signal. These data support previous reports of dual-pitch processing mechanisms in macaque monkeys based on electrophysiological recordings (multiunit activity and current source density) in primary auditory cortex in response to same phase and alternating phase click trains (Steinschneider et al., 1998). Temporally based

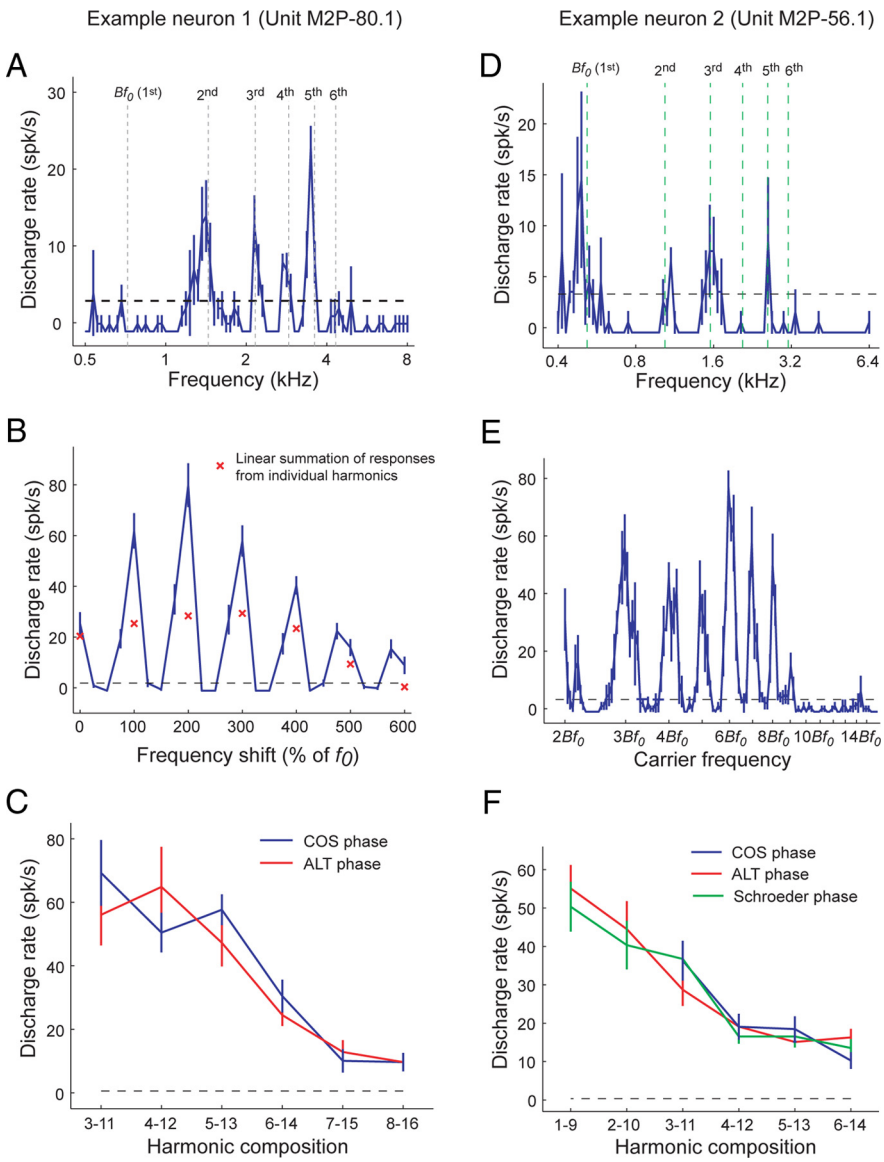

Figure 7. Multipeaked neuron examples. ***A–C***, This neuron (unit M2P-80.1) was not classified as a pitch-selective neuron because it did not have the same tuning for pure tones and MF harmonic complex sounds. However, it did respond to pure tones at harmonics of its Bf_0 at 720 Hz. ***A***, Multipeaked tuning to pure tones. Vertical dashed lines indicate the Bf_0 (720 Hz) and corresponding higher harmonics (2–6). ***B***, Sensitivity to pitch shift. Stimuli were three-component harmonic/inharmonic complex tones (1–3, f_0 = 720 Hz). Responses shown for the complex tone (curve) are greater than the summation of responses to each component played individually (× symbols). Responses to harmonic shifts (100% of f_0) are greater than to inharmonic or odd-harmonic shifts. The dashed lines in ***A*** and ***B*** indicate the minimum firing rate criterion for a significant response (2 SDs above the mean spontaneous rate). ***C***, Similar responses were observed for COS and ALT harmonic complex tones in several harmonic compositions, indicating no PSI for this neuron. ***D–F***, Responses from another multipeaked neuron (unit M2P-56.1). ***D***, Frequency tuning obtained by pure tone stimuli. Vertical dashed lines indicate frequencies that are harmonically related to the Bf_0 (518 Hz) estimated using sAM tones. ***E***, Responses to sAM tones with a modulation frequency equal to the Bf_0 (518 Hz), varying in carrier frequency (labeled relative to the Bf_0). The neuron showed preferences for carrier frequencies that are integer multiples of the Bf_0. The dashed lines in ***D*** and ***E*** indicate the minimum firing rate criterion for a significant response (2 SDs above the mean spontaneous rate). ***F***, Similar responses are observed for harmonic complex tones in COS, ALT, and Schroeder phase. No phase sensitivity is observed in this neuron.

unitary models have provided an alternative explanation of the pitch shift phenomena associated with frequency-shifted complex tones, based on the fine structure processing of the acoustic signal by the auditory system (Meddis and O'Mard, 1997). Although we cannot rule out the use of fine structure information by pitch-selective neurons, the existence of harmonic templates within the putative pitch center (Fig. 7) indicates that pitch-selective neurons have access to explicit spectral information relevant for pitch extraction.

Given that pitch is a percept, any identified neural code for pitch must ultimately be shown to correspond to pitch measured psychophysically (rather than only to an acoustic feature that covaries with pitch). Although there are many potential acoustic manipulations that could be used to investigate pitch processing, we chose to examine phase manipulations in the current experiments because these changes create the most potentially salient differences in both spectral and temporal domains (a doubling of the f_0 and temporal envelope repetition rate, respectively). We observed that the cue (temporal envelope or spectral) used by marmosets to discriminate pitch in our behavioral experiments was similar to the information used by pitch-selective neurons. A fundamental frequency of 450 Hz was the transition point both perceptually and neurally for relying more on using temporal envelope cues (<450 Hz) or spectral cues (> 450 Hz) to extract pitch.

The data presented here indicate that how pitch is perceived in marmosets depends on fundamental frequency (in addition to harmonic order). Although marmosets and humans likely hear the same pitch for the majority of acoustic stimuli, our data suggest that, for those acoustic stimuli in which temporal envelope and spectral cues indicate different fundamental frequencies, marmosets and humans may perceive different pitches. For example, for an alternating phase harmonic complex tone with an f_0 = 150 Hz and harmonics 5–8, our data suggest that marmosets hear a pitch 1 octave above the f_0 (extracted from temporal envelope cues), whereas humans hear a pitch equal to the f_0 (extracted from spectral cues) (Shackleton and Carlyon, 1994). This is most likely a consequence of a smaller cochlea size in marmosets (compared with humans) that limits the species' spectral resolvability (Shera et al., 2002, 2010). Pitch discrimination thresholds have been shown to be worse in other nonhuman species compared with humans, also potentially a consequence of a smaller cochlear size (Shofner, 2002; Kalluri et al., 2008; Walker et al., 2009). A notable exception to this is in birds; behavioral experiments in birds have demonstrated that they have an exquisite sensitivity to small changes in the temporal fine structure, including phase, of periodic sounds (Cynx et al., 1990; Lohr and Dooling, 1998; Dooling et al., 2002).

There are several possible mechanisms that can provide the spectral and temporal envelope information necessary for pitch

extraction. It is unlikely that temporal envelope cues are extracted above the level of primary auditory cortex, because the frequency limit of envelope locking for the majority of cortical neurons in the core areas of the marmoset (Lu et al., 2001b; Liang et al., 2002; Lu and Wang, 2004; Bendor and Wang, 2007) is typically near to or less than the lower limit of pitch (Krumbholz et al., 2000). Examples of stimulus locking at higher frequencies have been observed from multiunit responses (Steinschneider et al., 1998) and single-unit responses in auditory cortex (De Ribaupierre et al., 1972; Wallace et al., 2002). Stimulus-locked responses at frequencies >200 Hz may be from thalamic fibers (Steinschneider et al., 1998) and only reflect the temporal fidelity of the medial geniculate body (as opposed to the temporal fidelity of auditory cortex). Given that the upper limit of pitch perception in humans extends several octaves above the cortical stimulus synchronization limit, extracting the fundamental frequency of a sound using temporal cues is more likely to happen subcortically, at least for higher fundamental frequencies. As such, possible candidates for the conversion of temporal envelope information (both fine structure and envelope) in the form of temporal firing patterns into a rate code are the pathways leading to auditory cortex from the ventral and dorsal divisions of the medial geniculate body or the inferior colliculus (Kaas and Hackett, 2000; Bartlett and Wang, 2007). This rate code must be sensitive to both the envelope repetition rate and temporal envelope regularity of the acoustic stimulus, both of which are represented in the temporal firing patterns of neurons at lower levels of the auditory system. Although previous studies have shown tuning in the inferior colliculus for modulation frequency (envelope repetition rate) (Langner and Schreiner, 1988), it is unknown whether these neurons show any preference for temporally regular sounds. If they do not pass this stricter criterion, then the cortical representation of pitch must rely on another subcortical pathway for its sensitivity to temporal envelope regularity.

Only within the putative pitch center have we found neurons consistently tuned to the temporal envelope regularity of the acoustic stimulus (Bendor and Wang, 2010), an acoustic feature required for pitch perception. This raises another important distinction between neurons found throughout auditory cortex that covary their firing rates with pitch-related information (Bizley et al., 2009, 2010; Bendor and Wang, 2010; Wang and Walker, 2012) and neurons within the putative pitch center that are defined as pitch selective using stricter criteria (Bendor and Wang, 2005). Further supporting their role in extracting pitch, here we have shown that the responses of pitch-selective neurons in the marmoset's pitch center also closely match the pitch discrimination of the marmoset (Figs. 2*B*, 4*C*).

In our experiments, we observed multipeaked frequency tuning from neurons within the putative pitch center that could provide a sufficient spectral input for pitch neurons to perform pitch extraction (Fig. 7). Although these harmonic template neurons exist in auditory cortex, it is entirely possible that they also exist at lower levels of the auditory system (Shamma and Klein, 2000). Unlike the temporal envelope processing of pitch, which must occur at a level of the auditory system in which neuronal inputs are synchronized to the fundamental frequency, spectral pattern matching can occur at higher levels of the auditory system. Given the higher degree of plasticity (Recanzone et al., 1993) and sharper spectral tuning of neurons in auditory cortex relative to lower levels of the auditory system (Bitterman et al., 2008; Bartlett et al., 2011), the auditory cortex may be a more optimal location for the formation of harmonic templates. If this is true, then the putative pitch center identified in auditory cortex may be the first stage of the auditory system in which temporally and spectrally based pitch extraction pathways are unified into a central neural representation of pitch.

References

Bartlett EL, Wang X (2007) Neural representations of temporally modulated signals in the auditory thalamus of awake primates. J Neurophysiol 97:1005–1017.

Bartlett EL, Sadagopan S, Wang X (2011) Fine frequency tuning in monkey auditory cortex and thalamus. J Neurophysiol 106:849–859.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165.

Bendor D, Wang X (2007) Differential neural coding of acoustic flutter within primate auditory cortex. Nat Neurosci 10:763–771.

Bendor D, Wang X (2010) Neural coding of periodicity in marmoset auditory cortex. J Neurophysiol 103:1809–1822.

Bernstein JG, Oxenham AJ (2003) Pitch discrimination of diotic and chotic complexes: harmonic resolvability or harmonic number? J Acoust Soc Am 113:3323–3334.

Bidelman GM, Heinz MG (2011) Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. J Acoust Soc Am 130:1488–1502.

Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature 451:197–201.

Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW (2009) Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J Neurosci 29:2064–2075.

Bizley JK, Walker KM, King AJ, Schnupp JW (2010) Neural ensemble codes for stimulus periodicity in auditory cortex. J Neurosci 30:5078–5091.

Cariani PA, Delgutte B (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol 76:1698–1716.

Cariani PA, Delgutte B (1996b) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J Neurophysiol 76:1717–1734.

Carlyon RP (1998) Comments on "A unitary model of pitch perception." J Acoust Soc Am 104:1118–1121.

Cedolin L, Delgutte B (2005) Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. J Neurophysiol 94:347–362.

Cynx J, Williams H, Nottebohm F (1990) Timbre discriminations in zebra finch (*Taeniopygia guttata*) song syllables. J Comp Psychol 104:303–308.

de Boer E (1956) Pitch of inharmonic signals. Nature 178:535–536.

De Ribaupierre F, Goldstein MH Jr, Yeni-Komshian G (1972) Cortical coding of repetitive acoustic pulses. Brain Res 48:205–225.

Dooling RJ, Leek MR, Gleich O, Dent ML (2002) Auditory temporal resolution in birds: discrimination of harmonic complexes. J Acoust Soc Am 112:748–759.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (1998) Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. Brain Res 786:18–30.

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

Flanagan JL, Guttman N (1960) On the pitch of periodic pulses. J Acoust Soc Am 32:1308–1319.

Gescheider GA (1985) Psychophysics: method, theory, and application. New York: Erlbaum.

Glasberg BR, Moore BC (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138.

Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. J Acoust Soc Am 54:1496–1516.

Griffiths TD, Hall DA (2012) Mapping pitch representation in neural ensembles with fMRI. J Neurosci 32:13342–13347.

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

Kalluri S, Depireux DA, Shamma SA (2008) Perception and cortical neural coding of harmonic fusion in ferrets. J Acoust Soc Am 123:2701–2716.

Krumbholz K, Patterson RD, Pressnitzer D (2000) The lower limit of pitch as determined by rate discrimination. J Acoust Soc Am 108:1170–1180.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799–1822.
Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J Neurophysiol 87:2237–2261.
Lohr B, Dooling RJ (1998) Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). J Comp Psychol 112:36–47.
Lu T, Wang X (2004) Information content of auditory cortical responses to time-varying acoustic stimuli. J Neurophysiol 91:301–313.
Lu T, Liang L, Wang X (2001a) Neural representation of temporally asymmetric stimuli in the auditory cortex of awake primates. J Neurophysiol 85:2364–2380.
Lu T, Liang L, Wang X (2001b) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131–1138.
McAlpine D (2004) Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. J Neurophysiol 92:1295–1311.
Meddis R, Hewitt MJ (1991a) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. J Acoust Soc Am 89:2866–2882.
Meddis R, Hewitt MJ (1991b) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. J Acoust Soc Am 89:2883–2894.
Meddis R, O'Mard L (1997) A unitary model of pitch perception. J Acoust Soc Am 102:1811–1820.
Micheyl C, Ryan CM, Oxenham AJ (2012) Further evidence that fundamental-frequency difference limens measure pitch discrimination. J Acoust Soc Am 131:3989–4001.
Moore BCJ (2003) An introduction to the psychology of hearing. London: Academic.
Moore GA, Moore BC (2003) Perception of the low pitch of frequency-shifted complexes. J Acoust Soc Am 113:977–985.
Osmanski MS, Wang X (2011) Measurement of absolute auditory thresholds in the common marmoset (*Callithrix jacchus*). Hear Res 277:127–133.
Oxenham AJ (2012) Pitch perception. J Neurosci 32:13335–13338.
Oxenham AJ, Shera CA (2003) Estimates of human cochlear tuning at low levels using forward and simultaneous masking. J Assoc Res Otolaryngol 4:541–554.
Oxenham AJ, Micheyl C, Keebler MV (2009) Can temporal fine structure represent the fundamental frequency of unresolved harmonics? J Acoust Soc Am 125:2189–2199.
Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S (2011) Pitch perception beyond the traditional existence region of pitch. Proc Natl Acad Sci U S A 108:7629–7634.
Patterson RD, Wightman FL (1976) Residue pitch as a function of component spacing. J Acoust Soc Am 59:1450–1459.
Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.
Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci 24:6810–6815.
Plack CJ, Oxenham AJ, Fay RR, Popper AN (2005) Pitch: neural coding and perception. Springer handbook of auditory research. New York: Springer Science.
Plomp R (1964) The ear as a frequency analyzer. J Acoust Soc Am 36:1355–1364.
Plomp R, Mimpen AM (1968) The ear as a frequency analyzer II. J Acoust Soc Am 43:764–767.
Pollack I (1968) Discrimination of mean temporal interval within jittered auditory pulse trains. J Acoust Soc Am 43:1107–1112.
Pressnitzer D, Patterson RD (2001) Distortion products and the perceived pitch of harmonic complex tones. In: Physiological and psychophysical bases of auditory function (Breebart DJ, Houtsma AJ, Kohlrausch A, Prijs VF, Schoonoven R, eds), pp 97–104. Maastricht, The Netherlands: Shaker Publishing.
Puschmann S, Uppenkamp S, Kollmeier B, Thiel CM (2010) Dichotic pitch activates pitch processing centre in Heschl's gyrus. Neuroimage 49:1641–1649.
Recanzone GH, Schreiner CE, Merzenich MM (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci 13:87–103.
Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A (2005) Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat Neurosci 8:1241–1247.
Schönwiesner M, Zatorre RJ (2008) Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. Exp Brain Res 187:97–105.
Schroeder MR (1970) Synthesis of low peak-factor signals and binary sequences with low autocorrelation. IEEE Trans Inform Theory 16:85–89.
Schwarz DW, Tomlinson RW (1990) Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). J Neurophysiol 64:282–298.
Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J Acoust Soc Am 95:3529–3540.
Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. J Acoust Soc Am 107:2631–2644.
Shera CA, Guinan JJ Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci U S A 99:3318–3323.
Shera CA, Guinan JJ Jr, Oxenham AJ (2010) Otoacoustic estimation of cochlear tuning: validation in the chinchilla. J Assoc Res Otolaryngol 11:343–365.
Shofner WP (2002) Perception of the periodicity strength of complex sounds by the chinchilla. Hear Res 173:69–81.
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.
Walker KM, Schnupp JW, Hart-Schnupp SM, King AJ, Bizley JK (2009) Pitch discrimination by ferrets for simple and complex sounds. J Acoust Soc Am 126:1321–1335.
Wallace MN, Shackleton TM, Palmer AR (2002) Phase-locked responses to pure tones in the primary auditory cortex. Hear Res 172:160–171.
Wang X, Walker KM (2012) Neural mechanisms for the abstraction and use of pitch information in auditory cortex. J Neurosci 32:13339–13342.
Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435:341–346.