Complex pitch perception mechanisms are shared by humans and a New World monkey

Xindong Song (宋欣东)[1], Michael S. Osmanski, Yueqi Guo, and Xiaoqin Wang[1]

Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205

Edited by Dale Purves, Duke University, Durham, NC, and approved November 23, 2015 (received for review August 13, 2015)

The perception of the pitch of harmonic complex sounds is a crucial function of human audition, especially in music and speech processing. Whether the underlying mechanisms of pitch perception are unique to humans, however, is unknown. Based on estimates of frequency resolution at the level of the auditory periphery, psychoacoustic studies in humans have revealed several primary features of central pitch mechanisms. It has been shown that (*i*) pitch strength of a harmonic tone is dominated by resolved harmonics; (*ii*) pitch of resolved harmonics is sensitive to the quality of spectral harmonicity; and (*iii*) pitch of unresolved harmonics is sensitive to the salience of temporal envelope cues. Here we show, for a standard musical tuning fundamental frequency of 440 Hz, that the common marmoset (*Callithrix jacchus*), a New World monkey with a hearing range similar to that of humans, exhibits all of the primary features of central pitch mechanisms demonstrated in humans. Thus, marmosets and humans may share similar pitch perception mechanisms, suggesting that these mechanisms may have emerged early in primate evolution.

pitch | marmoset | frequency discrimination | primate | hearing

Extracting pitch from periodic complex sounds is one of the most fundamental functions of the human auditory system, and this periodicity is a critical aspect of music, speech, animal vocalizations, stream segregation, and many other auditory functions. Indeed, almost all naturally occurring, pitch-bearing sounds are periodic, and perceptual sensitivity to acoustic periodicity has been demonstrated in a wide variety of vertebrate species including anurans, songbirds, and mammals. The perceptual mechanisms used by humans to extract a sound's pitch have been extensively studied, but to date there has been little evidence to suggest that any other species use mechanisms similar to those found in humans (1, 2). Further, frequency resolution at the level of the auditory periphery, an important component of pitch perception, is thought to be insufficient in other mammalian species to produce human-like pitch perception (2–5). This lack of adequate frequency resolution limits comparisons of central mechanisms with other species using the same analytical criteria established for humans.

Recent data from two macaque monkey species, representing Old World primates, and the marmoset, a highly vocal New World primate species separated from humans by about 40 million y ago (6) and phylogenetically located roughly between macaques and other nonprimate mammals tested in pitch studies, have begun to cast doubt on whether these pitch perception mechanisms are unique to humans. Both physiology data from the macaque monkey (7) and behavioral data from marmosets (8) suggest that these primate species may exhibit frequency resolution at the auditory periphery similar to that seen in humans. Based on these findings, we hypothesized that there may also be central mechanisms for pitch perception shared between humans and other primates. We show in this report evidence that marmosets exhibit all the primary features of central pitch mechanisms that have been demonstrated in humans. Based on these analyses, we suggest that these central pitch perception mechanisms are not unique to humans but can likely be found in nonhuman primates—including New World primate species—and thus may have originated relatively early in primate evolution.

Most pitch-evoking sounds occurring in natural environments have spectra with harmonic structure, in which the acoustic power is concentrated at frequencies that are integer multiples (harmonics) of a common fundamental frequency (F0). These harmonics are processed at the cochlea by a bank of peripheral auditory filters that separate the incoming signal into individual frequency channels along a tonotopic axis (1). The tuning bandwidth of these filters increases, and thus the frequency resolving power decreases as frequency increases (9, 10). In humans, only the lowest 5~10 harmonics are well-segregated into different auditory filters and can be heard individually from the whole complex sound (11, 12). These harmonics are known as resolved harmonics (RESs). The tuning bandwidth of auditory filters at high frequencies becomes larger than the spacing of adjacent harmonics, which is equal to F0 in a pitch-evoking sound. Thus, each auditory filter receives significant power from more than one harmonic, and these are defined as unresolved harmonics (URSs). In humans, the upper boundary of RESs and the lower boundary of URSs can be assessed behaviorally (11–14), and these measures can be used to determine the relationship between bandwidth and F0 at these two boundaries (10, 13, 14) (Fig. 1*A*, dashed gray lines). We applied these boundary ratios derived in humans to tuning bandwidths measured in marmosets (8) to estimate resolvability boundaries for marmosets (solid black lines). Noticeably, for an F0 of A440 according to the musical tuning standard (440 Hz), marmosets appear to have as many RESs as humans. Based on these data, a model of the excitation pattern at the level of the auditory periphery (15) was applied to marmosets (Fig. 1*B*). The model shows distinct peaks at each harmonic for RESs and a raised smooth plateau across frequency for URSs (also see Fig. S1).

Given the functional properties of the auditory periphery outlined above, there are several possible central mechanisms that can theoretically be used to extract cues to decode the pitch

Significance

Complex pitch perception serves a pivotal role in human audition, especially in speech and music perception. It has been suggested that pitch perception mechanisms demonstrated in humans are not shared by nonhuman species. Here we provide evidence that a New World monkey, the common marmoset, shares all primary features of complex pitch perception mechanisms with humans. Combined with previous findings of a specialized pitch processing region in both marmoset and human auditory cortex, this evidence suggests that pitch perception mechanisms likely originated early in primate evolution.

Author contributions: X.S. and X.W. designed research; X.S., M.S.O., and Y.G. performed research; X.S. analyzed data; X.S., M.S.O., and X.W. wrote the paper; and X.W. supervised the entire project.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1] To whom correspondence may be addressed. Email: songxindong@jhmi.edu or xiaoqin.wang@jhu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516120113/-/DCSupplemental.

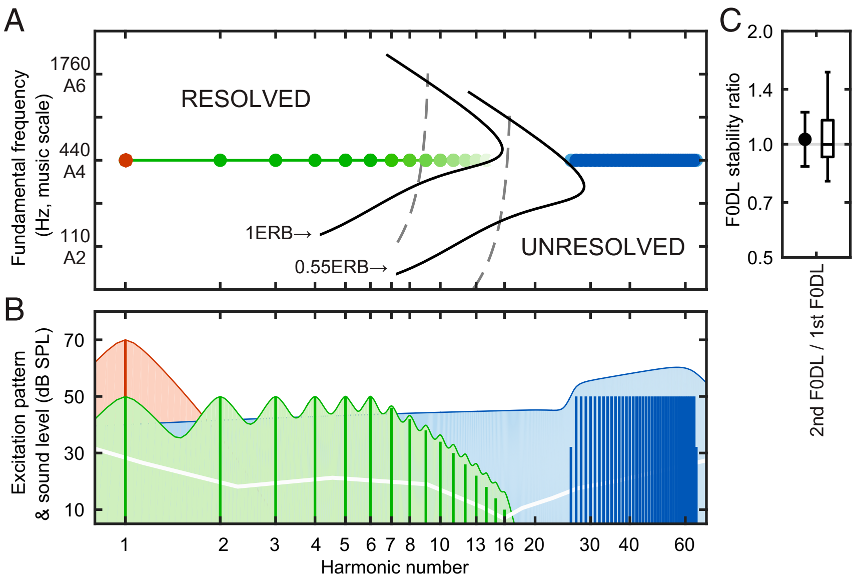

Fig. 1. Harmonic resolvability in pitch perception. (*A*) Estimated harmonic resolvability across F0s. Dashed gray lines indicate the estimated upper boundary of resolved harmonics (1ERB) and lower boundary of unresolved harmonics (0.55ERB) in humans. Black lines indicate the same boundaries estimated in marmosets. At F0 = 440 Hz, PTF0 (red), RES (green), and URS (blue) are indicated by circles. Color definitions are consistent throughout the following figures. (*B*) Vertical lines indicate acoustic spectra and sound levels of the background sounds used in the current marmoset F0DL measurements. Colored areas indicate peripheral excitation patterns (15) in marmosets, in which fluctuations can be seen on RESs (green) but not on URSs (blue). The extended blue tail of URSs on the low-frequency side indicates a noise masker used to mask potential distortion products from URSs back into the resolved side when measuring URSs' F0DL. The white line references the marmoset audiogram (16). (*C*) F0DL stability ratios, defined as the ratio between the second F0DL and the first F0DL measured on the same animal and under the same condition. The error bar indicates the mean value (1.031) and SD (1.181, logarithmic scale), with the box plot to the right (*n* = 30).

of a harmonic sound. The simplest mechanism is to use the lowest-frequency component present in a harmonic sound to extract pitch, which is usually equal to a pure tone at F0 (PTF0) (17). Alternatively, the central auditory system may contain many spectral harmonic templates. A match between one of these templates and the RES components of an incoming harmonic complex sound determines the pitch (18, 19). A third potential mechanism is to extract pitch from the interactions among URS components within an auditory filter, which generate a temporal envelope with a periodicity equal to the pitch (20). Although each of these mechanisms alone is sufficient to evoke pitch, the relative strengths of the perceived pitch based on these mechanisms are different. Pitch strength (a measure of the salience of the perceived pitch) is commonly believed to be correlated with the ability to discriminate changes in F0 (21). The smallest change in F0 that a subject can discriminate, measured as the F0 difference limen (F0DL), is thought to be inversely related to pitch strength. Previous F0DL studies have revealed that harmonic complex tones have a greater pitch strength than a pure tone at the same F0 (22–25), implying that the PTF0 alone cannot dominate the pitch strength of complex tones. Indeed, removing the F0 component from a harmonic complex tone does not affect the robustness and strength of pitch perception (the phenomenon of the "missing fundamental") (26, 27). Further, a significant F0DL increase (and consequently a decrease in pitch strength) was reported as spectral content was shifted out of the lower RES range (28, 29), suggesting that RESs have a greater pitch strength than URSs and thus dominate the pitch strength of harmonic complex tones when both RESs and URSs are present. This is the first primary feature of human pitch perception mechanisms.

The second primary feature of human pitch perception mechanisms is that the pitch of RESs is sensitive to the fidelity of spectral harmonicity. The F0DL of RESs could reflect sensitivity to a globally assembled pitch or, alternatively, to local changes to the individual components within each auditory filter (30). Importantly, sensitivity to individual components would not necessarily require them to be in a harmonic relationship. A series of experiments that measured the change in F0DLs for inharmonic versus harmonic sounds found, where both groups had the same average resolvability and spectral range, that F0DLs were higher under the inharmonic condition (21, 31, 32). These results have been taken as evidence of a role of harmonicity in lowering F0DLs and, because F0DL for RESs is affected by changes to the fidelity of harmonicity, suggest that F0DL reflects sensitivity to a globally assembled pitch.

The third primary feature of human pitch mechanisms is that the pitch strength of URSs is related to the salience of temporal envelope cues. Varying phase relationships among individual components of a harmonic complex sound can induce temporal envelope changes without affecting spectral amplitude. For example, a harmonic complex in Schroeder phase (33) has a flattened temporal envelope, and thus a reduction in the salience of temporal envelope cues, yet retains the same spectral profile as the same harmonic complex in sine phase. Harmonic complex sounds in Schroeder phase show increased F0DL compared with those in sine phase for URSs but not for RESs (28). Thus, changes in temporal envelope cues only affect the perceived pitch of URSs.

The relationships among these three primary features of human pitch perception mechanisms are depicted in Fig. S2. It has been shown that other mammals such as chinchillas (2) and macaque monkeys (34) can perceive sound periodicity through temporal processing. However, to the best of our knowledge, for any nonhuman species, neither the relative importance of RESs in pitch strength nor the sensitivity to spectral harmonicity in RES pitch has been shown behaviorally. The lack of evidence for these two features in animals has led to the proposal that pitch perception arises solely from temporal, rather than spectral, processing in nonhuman mammals (2).

Results

Fundamental Frequency Discrimination Limen Stability. To test whether marmosets also exhibit these primary features of human pitch mechanisms, we measured F0DLs in marmosets under eight different conditions (Fig. S3 and *SI Materials and Methods*). To assure that the measured F0DLs were stable over time, each F0DL value was measured twice and a stability ratio of the two measurements was calculated under each condition for each animal (Fig. 1*C*). There was no significant difference between the two F0DL measurements ($P = 0.517$, Wilcoxon signed-rank test, the same test used in all subsequent analyses), suggesting that our measurements were stable.

Feature 1: Pitch Strength of a Harmonic Tone Is Dominated by Resolved Harmonics. To test whether RESs dominate pitch strength in marmosets, we measured F0DLs under four stimulus conditions: (*i*) all harmonics covering the marmoset hearing range of a common F0 (ALL); (*ii*) PTF0; (*iii*) RESs only; and (*iv*) URSs only (Fig. 2 *A* and *B*). Example psychometric curves from one marmoset (M13W) are shown in Fig. 2*C* (also see Fig. S3). This animal can easily discriminate frequency changes greater than one semitone under both ALL and RES conditions, showing hit rates above 75%. F0DLs from each condition and across all tested animals are shown in Fig. 2*D*. Average F0DLs in the four stimulus conditions are 0.51 (ALL), 2.68 (PTF0), 0.37 (RESs), and 0.94 (URSs) semitones, respectively. The average pure tone threshold measured in marmosets (2.68 semitones, or 16.7% F0 change at 440 Hz) is comparable to reported pure tone thresholds at a similar frequency in squirrel monkeys, another New World monkey species (13.0% F0 change at 500 Hz) (35). F0DL dominance ratios are (defined and shown in Fig. 2*E*) significantly lower than 1 for both PTF0 and URSs ($P = 0.0039$, $P = 0.0039$). The dominance ratios for RESs, however, are significantly higher than 1 ($P = 0.0039$), suggesting that RESs play a critical role in producing F0DL and likely dominate pitch strength of harmonic complex sounds in marmosets.

Feature 2: Pitch of Resolved Harmonics Is Sensitive to the Quality of Spectral Harmonicity. To test whether F0DL of RESs in marmosets is sensitive to the fidelity of spectral harmonicity, we followed the same rationale as in human studies and sought to demonstrate that F0DLs of RESs in marmosets can be increased using inharmonic shifts. We measured additional F0DLs of modified RESs in which spectral components were inharmonically shifted using two methods based on previous human studies. First, we shifted odd-numbered harmonics upward by 15% of F0, and even-numbered harmonics downward by 15% of F0 (shift condition 1) (31). Second, we shifted all RES components upward by 25% of F0 (shift condition 2) (21, 32). Fig. 3*A* illustrates these spectral shifts. Generally, any inharmonic shift should produce a more ambiguous pitch compared with the harmonic condition (21). Fig. 3*B* shows that shift condition 1 generated significantly larger F0DLs than the harmonic F0DLs ($P = 0.0039$), which is consistent with findings in humans (31). Shift condition 2 showed a trend toward increased F0DLs, although not significantly higher than the harmonic F0DLs ($P = 0.191$, but see Fig. S4 and *SI Materials and Methods*). Together, these results show that inharmonic F0DLs are significantly higher than harmonic F0DLs in marmosets (Fig. 3*B*, shift conditions 1 and 2 combined, $P = 0.0081$), suggesting that spectral harmonicity is required to achieve a lower F0DL. Thus, F0DL of RESs reflects discrimination of a globally assembled pitch in marmosets.

Feature 3: Pitch of Unresolved Harmonics Is Sensitive to the Salience of Temporal Envelope Cues. Finally, we sought to determine the role of temporal envelope cues in URS-induced pitch in marmosets. We used the Schroeder phase to introduce a flattened temporal envelope on both RESs and URSs (Fig. 4 *A* and *B*). The Schroeder phase URS condition was much more difficult for one of our subjects, who failed to produce a comparable F0DL due to a

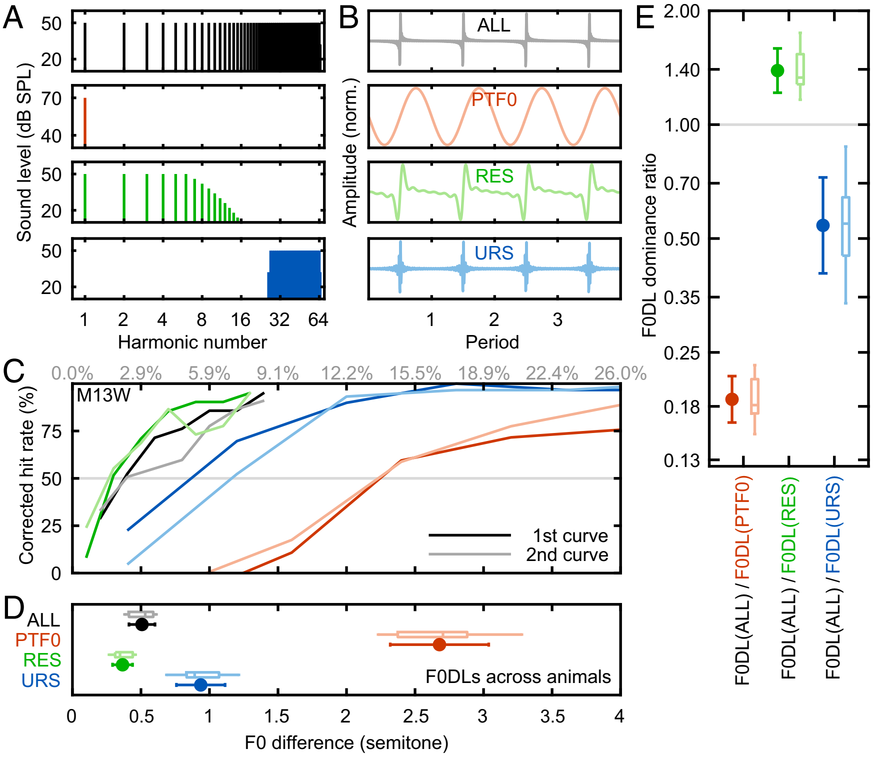

Fig. 2. RESs dominate marmoset pitch strength, similar to humans. (*A* and *B*) Spectra (*A*) and waveforms (*B*) of the background sounds used in marmoset F0DL measurements, for ALL (black), PTF0 (red), RESs (green), and URSs (blue) (noise masker not shown). (*C*) Example psychometric curves from the subject M13W under the four conditions in *A*. Darker and lighter lines indicate the first and the second measured curves, respectively. The gray line indicates 50% corrected hit rate. (*D*) F0DLs under each condition across all tested animals and measurements. Error bars indicate the mean values and SDs, with box plots above ($n = 8$, for each). (*E*) F0DL dominance ratios, defined as the ratio between the F0DL of all harmonics presented together and the F0DL measured under each decomposed condition ($n = 8$, for each). The gray line indicates a reference ratio equal to 1. The error bars indicate the mean values and SDs, with box plots to the right ($n = 8$, for each).

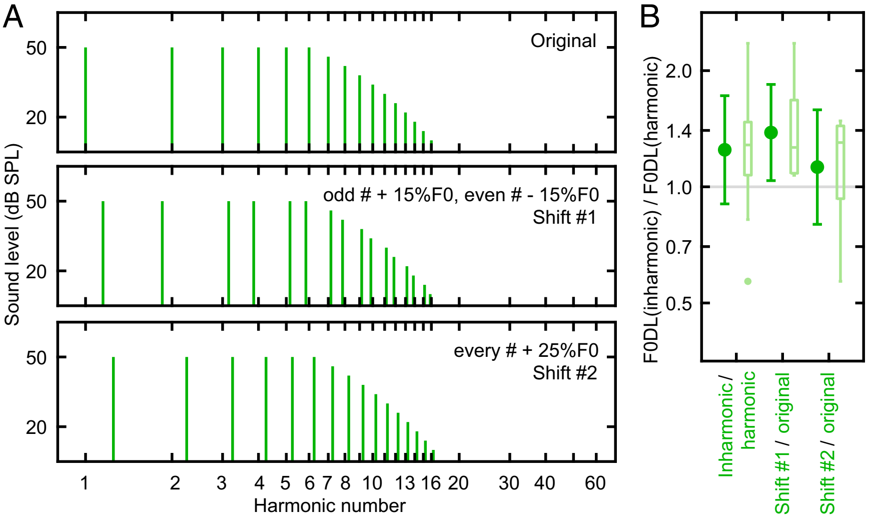

Fig. 3. F0DLs of RESs are sensitive to the quality of spectral harmonicity in marmosets. (*A*) Spectra of the background sounds used for testing harmonicity sensitivity on RESs. Vertical ticks serve as references for integer harmonic numbers. (*B*) F0DL harmonicity ratios, defined as ratios between inharmonic F0DL and harmonic F0DL. The gray line indicates a reference ratio equal to 1. The error bars indicate the mean values with SDs, with box plots to the right ($n = 16$, for inharmonic/harmonic; $n = 8$, for shift#1/original and shift#2/original).

persistent high false alarm rate. For the remaining three animals, the Schroeder phase did not introduce a significant F0DL increase for RESs ($P = 0.422$) but did so for URSs ($P = 0.016$) (Fig. 4*C*), suggesting that marmosets, like humans, are sensitive to the salience of temporal envelope cues of URSs. That is, marmosets appear to discriminate periodicity changes on URSs in a manner similar to that of human subjects (28).

Discussion

The findings described above show that marmosets share three primary features of pitch perception mechanisms demonstrated in humans. First, both species have a higher pitch strength for RESs compared with either URSs or PTF0. Second, both species are sensitive to changes in the fidelity of spectral harmonicity for RES components. Third, both species are sensitive to the salience of temporal envelope cues for URS components. The first two of these features have been previously demonstrated only in humans, and were not believed to be a component of auditory perception of nonhuman mammals (2). The present data are thus the first, to our knowledge, to show that a nonhuman species shares all three primary features of central pitch processing mechanisms with humans. The majority of previous work with nonhuman species has been done with rodents [e.g., chinchillas (2), gerbils (3)] and has generally shown an impoverished peripheral frequency resolution, which called into question the existence of human-like pitch perception mechanisms in these species. It was suggested, for example, that chinchillas, unlike humans, may rely solely on temporal envelope cues for perceiving periodicity (2).

Rodents share a common ancestor with primates ∼90 million y ago, whereas the separation of New World and Old World monkeys occurred only about 40 million y ago (6). Importantly, evidence from behavioral studies in a New World primate, the common marmoset (8), and physiological studies in Old World monkeys (7) suggests that both of these primate groups at least share similar peripheral frequency resolution with humans.

In addition to these perceptual data, a putative cortical pitch center has been described in marmoset auditory cortex (36), at the anterolateral low-frequency border of primary auditory cortex, which contains neurons responsive to pitch-evoking sounds in humans. Depending on harmonic resolvability, these neurons extract pitch using either spectral harmonicity or temporal envelope cues (37), and thus mirror features (*ii*) and (*iii*) of the pitch perception mechanisms mentioned above. In humans, correspondingly, sensitivity to pitch strength and harmonic composition has also been reported near the same functional cortical location as identified in marmoset auditory cortex (38, 39), suggesting a homologous cortical pitch processing center shared by humans and marmosets. All of these data together suggest that the marmoset is a valuable model system to study the neuronal circuitry underlying human-like pitch perception mechanisms and related auditory attributes.

Marmosets have a rich vocal repertoire that contains a variety of harmonic structures. Some of their vocalizations (e.g., “phee” and “twitter” calls) contain high-frequency F0s (>2–3 kHz), whereas others (e.g., “egg,” “moan,” and “squeal” calls) contain

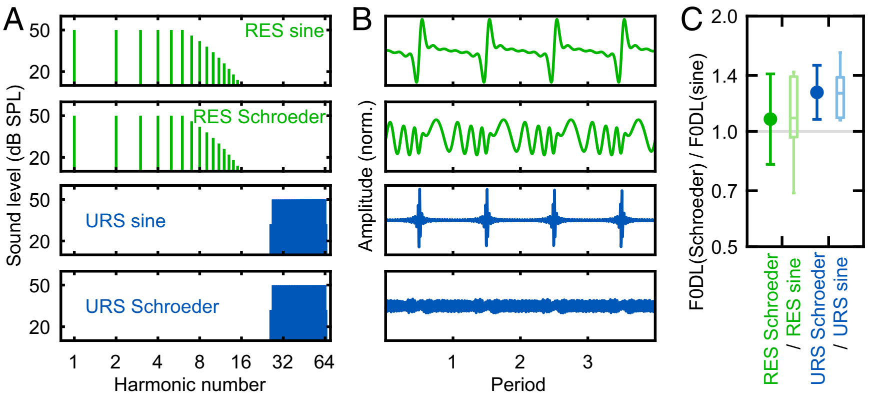

Fig. 4. F0DLs of URSs are sensitive to the salience of temporal envelope cues in marmosets. (*A* and *B*) Spectra (*A*) and waveforms (*B*) of the background sounds used for testing URS sensitivity to the salience of temporal envelope cues. (*C*) F0DL Schroeder/sine ratios, defined as ratios between the F0DL measured using Schroeder phase sounds and the F0DL measured using sine phase sounds. The gray line indicates a reference ratio equal to 1. The error bars indicate the mean values with SDs, with box plots to the right ($n = 6$, for each).

low-frequency F0s (<2 kHz, in the range of pitch) (40–42). In addition, harmonic structures are also commonly found throughout the marmoset's natural acoustic environment in the South American rainforest, including the vocalizations of various heterospecific species such as insects, birds, amphibians, mammals, and so forth. Marmosets thus can, and likely do, make use of pitch perception mechanisms in the roles of both predator and prey in their natural habitat as they interact with these other species.

In sum, we suggest that human-like pitch perception mechanisms may have originated relatively early in primate evolution, perhaps as early as or even earlier than ~40 million y ago, when New World and Old World primates separated on the evolutionary tree. To more precisely localize the evolutionary origin of pitch perception, more species, including both primate and nonprimate species, need to be tested in a manner similar to what has been conducted in humans and marmosets. The resultant dataset would allow us to more fully describe the evolutionary development of peripheral frequency resolution and central processing mechanisms that have resulted in human-like pitch perception.

Materials and Methods

All experimental procedures were approved by the Johns Hopkins University Animal Care and Use Committee and were in compliance with the guidelines of the National Institutes of Health. See *SI Materials and Methods* for more details.

Estimation of Harmonic Resolvability. The absolute frequency resolution at a particular frequency at the auditory periphery can be quantified by the tuning bandwidth of an auditory filter centered at that frequency, measured as the equivalent rectangular bandwidth (ERB) (9). The relative resolution for a specified F0 can be described as the ratio α between this F0 and the measured ERB, as $\alpha = \mathrm{F0/ERB}$ or $\mathrm{ERB} = 1/\alpha \bullet \mathrm{F0}$. The smaller α goes, the more harmonics can pass through this auditory filter's bandwidth, thus increasing the likelihood that adjacent harmonics become unresolved. Alternatively, the higher α goes, the fewer harmonics (or none) pass within the bandwidth of the auditory filter, thus increasing the likelihood that adjacent harmonics become resolved.

By definition, RESs can be heard out individually from the complex. In humans, the highest RES of a wide range of F0s was assessed behaviorally (11, 12). These data show that, for a particular F0, it is proportional to the ERB of its highest RES, with a fixed ratio α around 1~1.25 (10). This ratio provides a link between peripheral frequency resolution and the upper boundary of RESs. Another path for deriving the link between peripheral frequency resolution and harmonic resolvability boundaries is to examine the saliency of temporal envelope cues. When adjacent harmonics have an alternating phase relationship (e.g., sine and cosine starting phase), the amplitude profile of the sound spectrum remains unchanged but the overall periodicity is doubled compared with the case when all harmonics begin in the same phase. If a sound's temporal envelope is the dominate cue for perceiving pitch, then an alternating phase complex should be reported as bearing a pitch an octave higher than a purely sine–cosine phase complex. Indeed, this was observed on URSs but not on RESs (13). The upper boundary of RESs and the lower boundary of URSs were then estimated, having α values 0.9 and 0.55, respectively. In addition, a recent study estimated the upper boundary of RESs to be where the peak-to-valley ratio of the excitation pattern is 1.98 dB (14), from where $\alpha = 1.18$ can be derived mathematically.

Based on the findings described above, we used α values of 1 and 0.55 to derive resolvability boundaries in marmosets based on previously estimated ERB data for this species at 500, 1,000, 7,000, and 16,000 Hz (8). The α value of 1 for the upper boundary of RESs was validated by tuning bandwidths of single units recorded at each unit's best sound level in awake marmoset auditory cortex (8, 43). Those boundaries are shown as black lines in Fig. 1*A*.

A one-parameter rounded exponential filter was borrowed from the human excitation pattern model (15) to build an analogous model of the marmoset excitation pattern. ERBs were interpolated ("pchirp" in MATLAB R2013b; MathWorks) from behavioral measurements of peripheral tuning bandwidths in marmosets (8) to cover the entire marmoset hearing range. Parameter p in the rounded exponential filter model is given as $p = 4 \cdot f/\mathrm{ERB}$ (15).

Operant Conditioning Task. The basic operant task and apparatus have been described in detail previously (8, 16, 44). In the current study, animals were presented with repeating "background" sounds that had an F0 equal to 440 Hz, also known as A440 or A4 in musical tuning standard (45). Each testing trial had a variable duration waiting time that lasted from 3 to 15 s, during which this background sound was repeatedly presented to the animal. After this waiting period, a "target" sound, which was always higher in "equivalent" fundamental frequency than the background, began to alternate with the background sound. Both the background and target sounds had a duration of 200 ms with a 10-ms linear ramp (rise/fall time). The interstimulus interval was fixed at 300 ms. Animals could respond anytime during the alternation period (i.e., the response window), which lasted for 4.8 s in total. The subject had to detect the F0 change and respond by licking at a feeding tube placed in front of its mouth during the response window ("hit") to receive a food reward. However, if the subject licked before the response window onset, the chamber light was extinguished for 2–5 s as a warning signal. If the subject did not respond during the trial at all, a "miss" was recorded and the system automatically started the next trial. Each experimental session contained at least 100 but not more than 200 trials, in which 70% of trials were measuring hit rates on real targets randomly chosen from seven possible target choices corresponding to seven different F0 distances from the background sound. These possible F0 changes were equally spaced in the semitone scale (Fig. S3*A*). The remaining 30% of trials were sham trials in which no target sound was presented. Sham trials were used to measure false alarm rate as an indicator of how much the subject relied on guessing during the task. Sample raw hit rates are shown in Fig. S3*A* for subject M13W under URS condition. As the F0 difference decreases, the raw hit rate drops from nearly perfect 100% to around false alarm rate. The corrected hit rates were calculated by corrected hit rate = (raw hit rate − false alarm rate)/(1 − false alarm rate). Discrimination thresholds were defined as that F0 difference that the animals correctly identified 50% of the time (46). Experimental sessions with a false alarm rate >25% or with a corrected hit rate curve passing below 50% multiple times were excluded from analysis. Testing continued until at least three consecutive experimental sessions produced discrimination thresholds within one F0 spacing between adjacent targets. Qualified experimental sessions from the same subject under the same condition were combined together in temporal order and then equally divided into two analysis parts. The first and second F0DLs were calculated as described previously for calculating discrimination thresholds for each experimental session, but based on the data from those two analysis parts. Each analysis part contained at least 24 repetitions for each target. For comparisons of F0DLs, ratios were calculated between F0DLs; Wilcoxon signed-rank test was used to test statistical significance.

For PTF0 discrimination, the background level was calibrated to be around 40 dB sensation level (SL) [~70 dB sound pressure level (SPL)]. Targets were adjusted in level to match the sensation level of the background, based on the marmoset audiogram (16), to eliminate level differences as a potential cue. For the other stimuli, the maximum level of harmonics was calibrated to be around 50 dB SPL. Previous results in chinchillas showed that the F0DL of complex tones composed of the first 10 harmonics was lower than the F0DL of a pure tone of the same F0 (47). However, these results cannot exclude the possibility that discrimination in that task was based on the relative location of the highest harmonic on the spectral edge alone rather than discriminating pitch per se (48). To minimize the possibility that our subjects could use spectral edges as a cue for discrimination, we implemented roll-offs on the spectral edges. Upper spectral edges of RES sounds were rolled off as level = (50 − (f − 2,640 Hz)/110 Hz) dB SPL, ending at 7,040 Hz. Upper spectral edges of ALL sounds and both edges of URS sounds were rolled off as level = (50 + 20 · log10((10^(1 − |f − *Fedge*|/660) − 1)/9)) dB SPL (49), where upper *Fedge* = 28.16 kHz and lower *Fedge* = 11.88 kHz. The inharmonic shifts on RESs are described in the main text; both inharmonic shifted RESs and harmonic RESs follow the same spectral envelope, as described above. Schroeder phases with a negative sign were generated for both RESs and URSs. It is believed that Schroeder phases with a negative sign have a flatter temporal envelope compared with Schroeder phases with a positive sign, after phase dispersion in the inner ear (50). Schroeder phase sounds have the same spectral envelope as their sine-phased counterparts.

Background sound level was randomly roved within ±3 dB. Target sound level was always fixed. For URS measurements, a fixed-level band-pass white noise was generated online to prevent subjects from using potential nonlinear distortion products on the lower-frequency side to do the discrimination. The cutoff frequencies of the noise masker were 100 and 12,000 Hz. Noise was estimated as 40~46 dB SPL per ERB. All sound levels were calibrated using a 1/2-inch free-field microphone (Brüel and Kjaer; type 4191) positioned at the same location as the animal's head, with a customized program.

During testing, marmosets were seated in a plexiglass restraint chair mounted in the center of a single-walled sound isolation chamber (IAC Acoustics; model 400A) lined with 2-inch acoustic absorption foam (Pinta Acoustic; model PROSPEC). Sounds were played through a speaker (Tannoy;

model Arena) powered by an amplifier (Crown Audio; D-75A) mounted 40 cm in front of the animal's head. All sound signals were generated using a customized MATLAB program (MathWorks) and delivered at a nominal sampling rate of 100 kHz through a multiprocesser digital signal processing unit (Tucker-Davis Technologies; RX6), followed by a programmable attenuator (Tucker-Davis Technologies; PA5) and an audio amplifier (Crown Audio; model D-75A).

Subjects. Four adult common marmosets (all male) were used in the current study, ranging from 2 to 6 y old during testing. Each had at least 8 mo of experience in discrimination tasks, either with previous auditory peripheral tuning bandwidth measurements (8) or with pure tone discrimination training. Two were head-fixed during all testing sessions (44). All subjects finished all eight testing conditions except subject M62U, who failed to obtain a comparable F0DL for Schroeder phase URSs due to a persistent high false alarm rate and was consequently excluded from Schroeder phase RES testing. Testing order for each subject is listed in Table S1. Subjects were housed in individual cages in a large colony at the Johns Hopkins University School of Medicine. All subjects were maintained at ∼90% of their free-feeding weight on a diet consisting of monkey chow, fruit, and yogurt and had ad libitum access to water. Subjects were tested 5 or 6 d/wk between 0900 and 1800 hours.

ACKNOWLEDGMENTS. We thank Nate Sotuyo and Shanequa Smith for their assistance with animal care. This work was supported by NIH Grants DC003180 (to X.W.) and DC013150 (to M.S.O.).

1. Plack CJ, Oxenham AJ, Fay RR, Popper AN (2005) *Pitch: Neural Coding and Perception* (Springer, New York).
2. Shofner WP, Chaney M (2013) Processing pitch in a nonhuman mammal (*Chinchilla laniger*). *J Comp Psychol* 127(2):142–153.
3. Klinge A, Itatani N, Klump GM (2010) A comparative view on the perception of mistuning: Constraints of the auditory periphery. *The Neurophysiological Basis of Auditory Perception*, eds Lopez-Poveda EA, Palmer AR, Meddis R (Springer, New York), pp 465–475.
4. Shera CA, Guinan JJ, Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99(5): 3318–3323.
5. Shera CA, Guinan JJ, Jr, Oxenham AJ (2010) Otoacoustic estimation of cochlear tuning: Validation in the chinchilla. *J Assoc Res Otolaryngol* 11(3):343–365.
6. Worley KC, et al.; Marmoset Genome Sequencing and Analysis Consortium (2014) The common marmoset genome provides insight into primate biology and evolution. *Nat Genet* 46(8):850–857.
7. Joris PX, et al. (2011) Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *Proc Natl Acad Sci USA* 108(42):17516–17520.
8. Osmanski MS, Song X, Wang X (2013) The role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (*Callithrix jacchus*). *J Neurosci* 33(21):9161–9168.
9. Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47(1-2):103–138.
10. Moore BCJ (2012) *An Introduction to the Psychology of Hearing* (Emerald, Bingley, UK), 6th Ed.
11. Plomp R (1964) The ear as a frequency analyzer. *J Acoust Soc Am* 36(9):1628–1636.
12. Plomp R, Mimpen AM (1968) The ear as a frequency analyzer. II. *J Acoust Soc Am* 43(4):764–767.
13. Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95(6): 3529–3540.
14. Bernstein JGW, Oxenham AJ (2006) The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *J Acoust Soc Am* 120(6):3916–3928.
15. Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J Acoust Soc Am* 74(3):750–753.
16. Osmanski MS, Wang X (2011) Measurement of absolute auditory thresholds in the common marmoset (*Callithrix jacchus*). *Hear Res* 277(1-2):127–133.
17. Ohm GS (1843) Uber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann Phys* 135(8):513–565. German.
18. Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54(6):1496–1516.
19. 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(5 Pt 1): 2631–2644.
20. Schouten JF (1940) The residue and the mechanism of hearing. *Proc K Ned Akad Wet* 43:991–999.
21. Micheyl C, Divis K, Wrobleski DM, Oxenham AJ (2010) Does fundamental-frequency discrimination measure virtual pitch discrimination? *J Acoust Soc Am* 128(4): 1930–1942.
22. Zeitlin LR (1964) Frequency discrimination of pure and complex tones. *J Acoust Soc Am* 36(5):1027.
23. Henning GB, Grosberg SL (1968) Effect of harmonic components on frequency discrimination. *J Acoust Soc Am* 44(5):1386–1389.
24. Fastl H, Weinberger M (1981) Frequency discrimination for pure and complex tones. *Acustica* 49(1):77–78.
25. Spiegel MF, Watson CS (1984) Performance on frequency-discrimination tasks by musicians and nonmusicians. *J Acoust Soc Am* 76(6):1690–1695.
26. Schouten JF (1938) The perception of subjective tones. *Proc K Ned Akad Wet* 41: 1086–1093.
27. Licklider JCR (1956) Auditory frequency analysis. *Information Theory*, ed Cherry C (Academic, New York), pp 253–268.
28. Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87(1):304–310.
29. Kaernbach C, Bering C (2001) Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *J Acoust Soc Am* 110(2):1039–1048.
30. Faulkner A (1985) Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations? *J Acoust Soc Am* 78(6):1993–2004.
31. Moore BCJ, Glasberg BR (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J Acoust Soc Am* 87(5):2163–2177.
32. Micheyl C, Ryan CM, Oxenham AJ (2012) Further evidence that fundamental-frequency difference limens measure pitch discrimination. *J Acoust Soc Am* 131(5):3989–4001.
33. Schroeder M (1970) Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans Inf Theory* 16(1):85–89.
34. Joly O, et al. (2014) A perceptual pitch boundary in a non-human primate. *Front Psychol* 5:998.
35. Wienicke A, Häusler U, Jürgens U (2001) Auditory frequency discrimination in the squirrel monkey. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 187(3): 189–195.
36. Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436(7054):1161–1165.
37. Bendor D, Osmanski MS, Wang X (2012) Dual-pitch processing mechanisms in primate auditory cortex. *J Neurosci* 32(46):16149–16161.
38. Norman-Haignere S, Kanwisher N, McDermott JH (2013) Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J Neurosci* 33(50):19451–19469.
39. 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(30):6810–6815.
40. Epple G (1968) Comparative studies on vocalization in marmoset monkeys (Hapalidae). *Folia Primatol (Basel)* 8(1):1–40.
41. Bezerra BM, Souto A (2008) Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int J Primatol* 29(3):671–701.
42. Agamaite JA, Chang C-J, Osmanski MS, Wang X (2015) A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *J Acoust Soc Am* 138(5):2906–2928.
43. Bartlett EL, Sadagopan S, Wang X (2011) Fine frequency tuning in monkey auditory cortex and thalamus. *J Neurophysiol* 106(2):849–859.
44. Remington ED, Osmanski MS, Wang X (2012) An operant conditioning method for studying auditory behaviors in marmoset monkeys. *PLoS One* 7(10):e47895.
45. ISO (1975) *ISO16:1975 - Acoustics - Standard Tuning Frequency (Standard Musical Pitch)* (International Organization for Standardization, Geneva).
46. Gescheider GA (1985) *Psychophysics: Method, Theory, and Application* (Lawrence Erlbaum, NJ), 2nd Ed.
47. Shofner WP (2000) Comparison of frequency discrimination thresholds for complex and single tones in chinchillas. *Hear Res* 149(1-2):106–114.
48. Nelson DA, Kiester TE (1978) Frequency discrimination in the chinchilla. *J Acoust Soc Am* 64(1):114–126.
49. Moore BCJ, Moore GA (2003) Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hear Res* 182(1-2):153–163.
50. Kohlrausch A, Sander A (1995) Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *J Acoust Soc Am* 97(3): 1817–1829.
51. Slaney M (1998) Auditory Toolbox (Interval Research, Palo Alto, CA), Version 2, Tech Rep 1998-010. Available at https://engineering.purdue.edu/~malcolm/interval/1998-010.

Supporting Information

Song et al. 10.1073/pnas.1516120113

SI Materials and Methods

The spatiotemporal activity pattern was also modeled to show both resolvability change across different frequencies and temporal envelope cue along the temporal axis, as shown in Fig. S1*B*. Auditory filters were modeled based on the Gammatone filter bank algorithm (51). Bandwidth parameters were also taken from marmoset ERB data (8). The incoming sound signal was passed through this filter bank and then rectified.

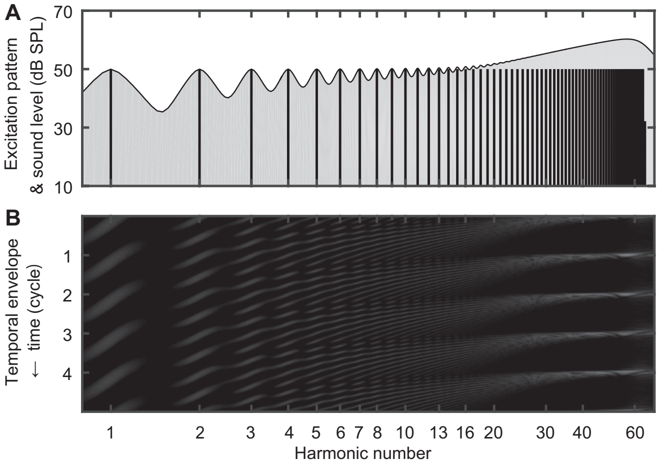

Fig. S1. Illustration of harmonic resolvability in marmosets. (*A*) Vertical lines indicate the acoustic spectrum and sound levels of the background sound used in “ALL” condition F0DL measurements. The shadowed area indicates its excitation pattern in marmoset auditory peripheries. (*B*) The spatiotemporal activity pattern of marmoset auditory peripheries. Five F0 cycles are shown along the vertical axis, against harmonic numbers along the horizontal axis. Vertically oscillating stripes are discretely distributed on each RES, according to its frequency. However, on higher-frequency URSs, harmonics can no longer be spectrally separable from adjacent harmonics. In return, their temporal interactions generate a temporal envelope repetition rate equal to F0, and can serve as a pitch cue.

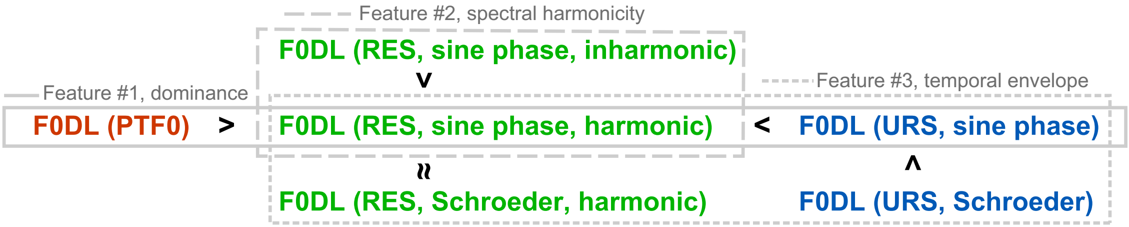
Feature #2, spectral harmonicity
F0DL (RES, sine phase, inharmonic)
v
Feature #1, dominance
Feature #3, temporal envelope
F0DL (PTF0)
>
F0DL (RES, sine phase, harmonic)
<
F0DL (URS, sine phase)
≈
^
F0DL (RES, Schroeder, harmonic)
F0DL (URS, Schroeder)

Fig. S2. Summary of human pitch perception mechanisms. Feature 1: Lower resolved harmonics (RESs) have a stronger pitch strength, and thus a smaller F0 discrimination threshold, compared with both pure tones at the fundamental frequency (F0) and higher unresolved harmonics (URSs). Feature 2: Pitch perception based on resolved harmonics is sensitive to the fidelity of spectral harmonicity. Feature 3: Pitch perception based on unresolved harmonics is sensitive to the saliency of temporal envelope cues.

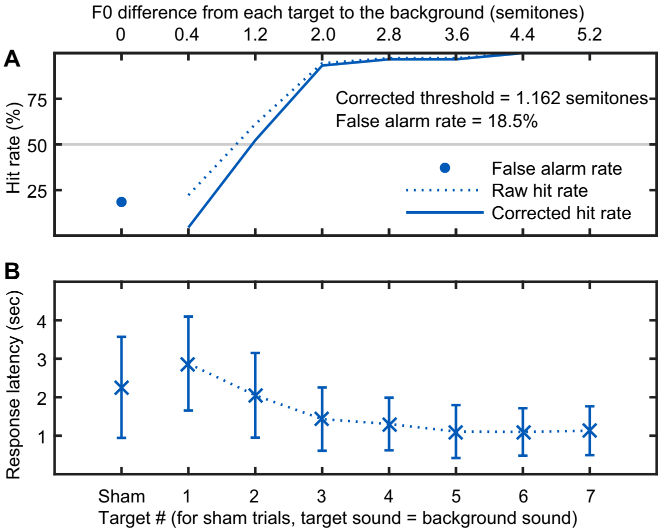

Fig. S3. Summary of behavioral methods. (*A*) The corrected psychometric curve along seven different targets in the second F0DL measure, on subject M13W, under the URS condition. The gray line indicates a threshold reference line for 50% corrected hit rate. Raw hit rates and the false alarm rate are also shown. (*B*) Response latencies across different targets. The error bars indicate the mean values with SDs. False alarm response latency is also shown, labeled as "Sham".

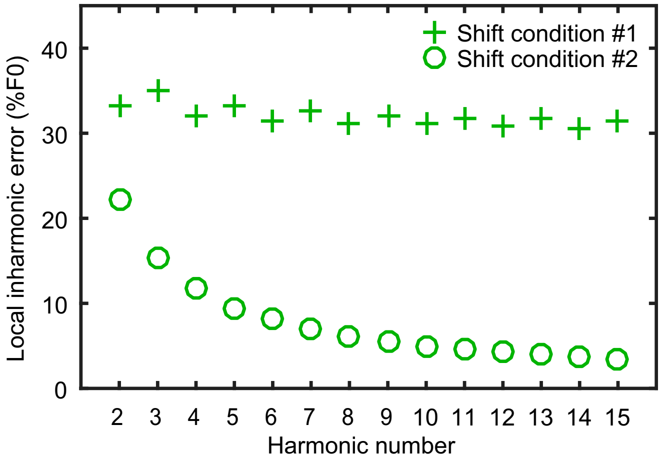

Fig. S4. Inharmonicity produced by shift conditions. For each of the inharmonic shift conditions in the present experiment, we measured the degree of deviation from harmonicity around each component. To do this, for each component together with its neighboring two components, we found the nearest three harmonic templates that best approximated the inharmonic complex. The frequency difference between each of the three inharmonic components and its corresponding harmonic component was calculated as individual inharmonic error. A local inharmonic error measure was calculated by summing the individual inharmonic errors across all three adjacent components and dividing by the F0 of the harmonic template. As shown in the figure, shift condition 1 introduces consistently significant amounts of inharmonicity across the entire RES frequency range. However, shift condition 2 can only introduce significant amounts of inharmonicity for the first several harmonics, and the local inharmonic error drops below 10% of F0 after the fifth harmonic. In humans, the spectral region that contributes most to the globally assembled pitch perception is reported to be roughly between the first and the fourth harmonic (1). However, in marmosets, the spectral region that contributes most to the globally assembled pitch might be composed of the higher-numbered harmonics, although still in the RES range. If that is true, then, for marmosets, shift condition 2 approximates a harmonic series inside this region, whereas shift condition 1 remains inharmonic across the entire RES range. Our data show that shift condition 1 introduces a significant increase in F0DL, whereas shift condition 2 only shows an increased trend, and may suggest that the spectral region that contributes most to the globally assembled pitch in marmosets might be higher than those in humans, at least for an F0 of 440 Hz.

Table S1. Testing order of different conditions on each animal

Subject/order	M62U	M13W	M11X	M4Y
1	PTF0	PTF0	PTF0	PTF0
2	ALL	ALL	ALL	ALL
3	URS, sine phase	RES, shift 2	URS, sine phase	RES, sine phase
4	RES, shift 1	RES, sine phase	URS, Schroeder phase	RES, shift 2
5	RES, shift 2	RES, Schroeder phase	RES, sine phase	RES, Schroeder phase
6	RES, sine phase	RES, shift 2	RES, Schroeder phase	RES, shift 1
7	URS, Schroeder phase	URS, Schroeder phase	RES, shift 2	URS, sine phase
8		URS, sine phase	RES, shift 1	URS, Schroeder phase