SCIENTIFIC REPORTS

OPEN

Monkeys share the neurophysiological basis for encoding sound periodicities captured by the frequency-following response with humans

Yaneri A. Ayala[1], Alexandre Lehmann[2,3,4] & Hugo Merchant[1]

The extraction and encoding of acoustical temporal regularities are fundamental for human cognitive auditory abilities such as speech or beat entrainment. Because the comparison of the neural sensitivity to temporal regularities between human and animals is fundamental to relate non-invasive measures of auditory processing to their neuronal basis, here we compared the neural representation of auditory periodicities between human and non-human primates by measuring scalp-recorded frequency-following response (FFR). We found that rhesus monkeys can resolve the spectrotemporal structure of periodic stimuli to a similar extent as humans by exhibiting a homologous FFR potential to the speech syllable /da/. The FFR in both species is robust and phase-locked to the fundamental frequency of the sound, reflecting an effective neural processing of the fast-periodic information of subsyllabic cues. Our results thus reveal a conserved neural ability to track acoustical regularities within the primate order. These findings open the possibility to study the neurophysiology of complex sound temporal processing in the macaque subcortical and cortical areas, as well as the associated experience-dependent plasticity across the auditory pathway in behaving monkeys.

The extraction and encoding of the time-varying features of acoustic signals is fundamental for complex skills. For example, successful vocal communication requires the dynamic sampling and neural representation of the spectral and temporal information of vocal sounds at multiple time scales[1]. The fidelity of neural encoding of acoustic regularities can be captured by the frequency-following response (FFR), an event-related potential reflecting synchronous phase-locked activity of different subcortical and cortical brain areas to the fundamental frequency (F0) and its harmonics of the sound[2–5]. The FFR mimics the envelope and/or fine structure of a sound waveform, capturing most of the temporal features of the evoking sound with extreme granularity, including the duration and periodicity of a stimulus.

The FFR has become increasingly popular as a neural biomarker of the stability of sound processing that correlates with literacy[6–12] and beat entrainment capabilities[13–15], revealing an intricate coupling between the auditory and motor systems[16,17]. More recently, it has been shown that the FFR can also act as a diagnostic tool for mild traumatic brain injuries[18,19]. Importantly, the FFR exhibits flexibility at different time scales and can be modified by changes in the stimulus dynamics[20–22] and by short[23–25] and long-term experiences[26,27]. Short-term interventions, such as auditory or linguistic training, enhance the neural encoding of the F0 which accompanies improvements in perceptual skills[25,28]. Thus, the FFR can act as an index of experience-dependent plasticity of sensory processing.

[1]Instituto de Neurobiología, UNAM, Campus Juriquilla, Boulevard Juriquilla No. 3001, Querétaro, Qro. 76230 Mexico. [2]Department of Otolaryngology Head & Neck Surgery, McGill University, Montreal, QC, Canada. [3]International Laboratory for Brain, Music and Sound Research (BRAMS), Center for Research on Brain, Language and Music (CRBLM), Pavillon 1420, Montreal, QC H3C 3J7, Canada. [4]Department of Psychology, University of Montreal, Montreal, QC, Canada. Correspondence and requests for materials should be addressed to Y.A.A. (email: yaneri.a.ayala@gmail.com) or H.M. (email: hugomerchant@unam.mx)

Because of the clinical and empirical relevance of the FFR, it is of timely interest to disentangle the neurophysiological underpinnings behind the extraction of temporal regularities of complex sounds, its experience-dependent plasticity as well as its association with sensorimotor capabilities. To do so, a mandatory step is to determine the commonalities in the neuronal sensitivity to temporal periodicities across humans and animals. Non-human primate is the closest animal model of human brain structure and function and they are able to discriminate speech sounds[29,30], as well as to detect[31,32] and entrain movements to isochronous auditory metronomes[33–41]. Hence, we compared the scalp-recorded response to the /da/ periodic syllable between human and rhesus monkeys. We hypothesized that the neural evoked responses in rhesus monkeys will show similar properties to the humans′ FFR, since the anatomy and physiology of the early and intermediate stages of the auditory pathway are conserved between the *Homo sapiens* and the macaque[42,43].

Results

Scalp-recorded FFR potentials were obtained from 18 human subjects and two monkeys who passively listened to a /da/ syllable of 40 ms duration[2]. Alternating polarity stimuli were delivered using a previously validated free-field method[44]. We used the same rate (11.1 Hz) and intensity (85 dB SPL) for both species. Temporal and spectral features of the added average of the electroencephalographic (EEG) response to stimulus of positive and negative polarity were estimated for each subject.

Comparable FFR in human and monkey. We found that the monkey FFR to the extensively studied /da/ syllable exhibits a comparable morphology to that of humans. The auditory evoked potential recorded in humans reliably replicated the FFR potential described with close[2,45–47] and free-field protocols[44]. The human FFR exhibited seven characteristic positive peaks comprising the typical transient (peaks 1–3, 7) and sustained features (peaks 4–6) of the response (Fig. 1a,b). Using the same recording set-up and stimulation protocol, we observed that the monkey FFR exhibited six out of the seven peaks of the human response, including the three sustained elements that reflect the encoding of stimulus periodicity (Fig. 1c, Table 1). Peak 2 corresponding to stopping the consonant was absent in both monkeys, as well as the offset peak 7 in monkey A. The other transient peaks corresponding to the sound onset (peak 1) and to the transition between stop burst and onset of voicing (peak 3) were clearly identified. Importantly, analogous periodic peaks comprising the steady-state FFR segment were identified and labeled as 4–6 in both species. We performed a bootstrapping method with a 95% confidence interval (CI) to compare the FFR components in both species. The amplitude and latency of most peaks were larger in monkeys than in humans (Table 2). An accurate encoding of the F0 of the sound was revealed by the inter-peak intervals of the sustained FFR components. The 4–5 and 5–6 interpeak intervals showed a similar duration (~8.6 ms) to the period of the F0 of the stimulus (103–121 Hz) in the two primate species. Specifically, in humans, the 4–5 and 5–6 interpeak intervals were of 8.4 (interquartile range, IQR: 1.62) and 8.42 ms (IQR: 0.68), respectively. Similar values were obtained in monkeys at the 4–5 (monkey Y: 10.16, monkey A: 8.6 ms) and 5–6 interpeak intervals (monkey Y: 8.11, monkey A: 9.42 ms).

Stimulus-response temporal correlation. Cross-correlation analysis demonstrated that the monkey FFR reliably captures the morphology and timing of the stimulus waveform. The estimation of the stimulus-response correlation constitutes a validated measure of similarity between signals[2]. Likewise, the time shift between both signals resulting in maximum correlation is used as an objective estimation of the response onset[2,7,45]. Stimulus and response were maximally positively correlated at time lags of 7.23 (IQR: 0.99) and 8.25 ms in human and monkey subjects, respectively (Fig. 2). These time lags match the neural transmission delay previously reported in human studies[2,45] and the latency of the first peak observed in human and monkey response (Table 1). A significant stimulus-response correlation was obtained in human subjects (r = 0.3; IQR: 0.12, p < 4.17 e-5) and monkeys (monkey Y: r = 0.42, p = 2.37e-32; monkey A: r = 0.57, p = 7e-64).

Spectral content of the FFR. The evoked responses of monkeys also reflected the spectral features of speech sound by showing phase-locking properties to the F0 as observed in humans. The spectral information conveyed by the FFR segment (20–60 ms) was determined by the fast Fourier transform, and its association with the spectra of the stimulus was determined with the coherency analysis. Notably, the response amplitude (Fig. 3a,b) and coherency (Fig. 3c) peaked at the F0 of the stimulus, which ramps between 103 and 121 Hz. Also, the power of the response was averaged at the frequency ranges of 90–140 and 200–750 Hz, which include the F0 and harmonics of the stimulus, respectively[44]. The mean amplitude of the grand-added average at the F0 frequency range for human subjects (1.28 μV; IQR: 0.55) and monkey (monkey Y: 3.08 μV, monkey A: 2.48 μV) was larger than 1 μV, whereas the amplitude of the response at harmonic range was considerably small in both species (humans: 0.23 μV; IQR: 0.09, monkey Y: 0.26 μV, monkey A: 0.28 μV). Between groups, the average power across epochs of monkeys was larger than that of humans at the F0 and higher harmonics (HH) range (Bootstrapping, 95% CI). The average amplitude across epochs of humans was of 24.81 (CIs.: 24.71, 24.90) and 9.56 μV (CIs.: 9.54, 9.59) for the F0 and HH range, respectively. For the monkey group, the F0 and HH average amplitude was of 34.35 (CIs.: 33.96, 34.71) and 14.35 μV (CIs.: 14.21, 14.46), respectively.

Intrinsic properties of the FFR. In both species, a robust neural representation of sound periodicities over time was captured by the FFR. The magnitude of neural activation was estimated by the signal-to-noise ratio (SNR), and neural consistency was calculated by the correlation between odd and even subaverages as previously described[44,48]. The amplitude of monkey (monkey Y: 2.05, monkey A: 2.98) and human (1.80; IQR: 0.88) FFR was above noise level (Fig. 4a). The average SNR across epochs indicated a slightly larger neural activation in monkey response (1.66, CIs.: 1.50, 1.87) than in human response (1.19, CIs.: 1.18, 1.19). The FFR was largely consistent across stimulus presentations in both human subjects (r = 0.76; 0.62, 0.89, p < 2.06e-32) and monkeys (monkey Y: r = 0.94, p = 0; monkey A: 0.72, p = 1.82e-15; Fig. 4b).
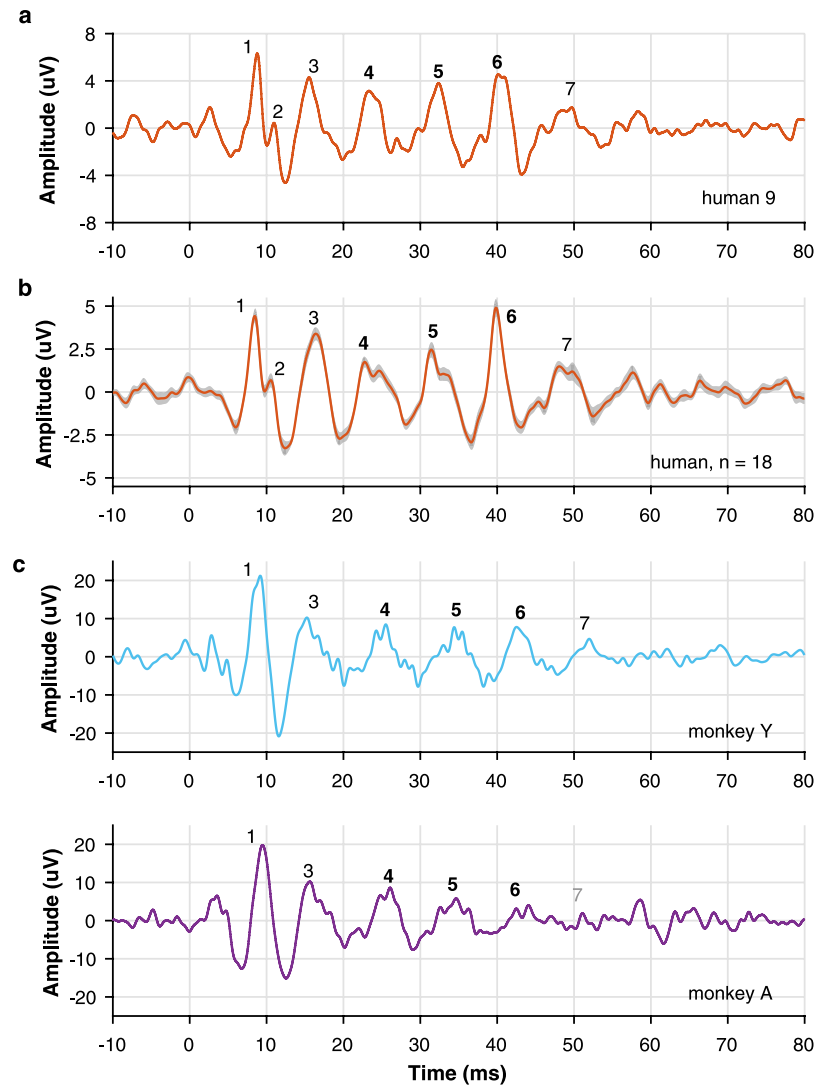
Figure 1. Human and macaque FFR to speech syllable /da/ in the time domain. (**a**) FFR of a representative human case. (**b**) Grand average FFR of the human population is shown as the mean ± SEM (shaded area). (**c**) FFR in two macaques, Y and A. Onset and sustained response peaks are numbered from 1–7. Peaks corresponding to the sustained part of the FFR are in bold. Note the 'not reliable' peak in gray in monkey A.

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7
Amplitude (uV)							
Human	4.80 (1.22)	0.33 (1.85)	3.81 (1.44)	2.63 (1.69)	3.19 (0.70)	4.76 (2.11)	2.91 (1.38)
Monkey Y	19.34	n.r.	9.12	5.66	6.11	7.24	4.02
Monkey A	18.49	n.r.	9.62	7.53	5.13	3.22	n.r.
Latency (ms)							
Human	8.37 (0.55)	10.58 (0.37)	16.15 (1.27)	23.28 (1.27)	31.55 (1.35)	39.91 (0.45)	48.51 (1.8)
Monkey Y	9.19	n.r.	15.25	24.26	34.42	42.53	51.99
Monkey A	9.47	n.r.	15.62	26.06	34.66	44.08	n.r.

Table 1. Amplitude and latency of response peaks estimated from the added-average response. Human data (n = 18) is shown as the median and the 25th and 75th IQR are in parenthesis. n.r., not reliable.

Discussion

We found that rhesus monkeys exhibit a homologous human FFR, revealing a shared neural sensitivity to acoustic periodicities. Like humans, monkeys are capable of resolving the spectrotemporal structure of the synthesized speech syllable /da/ (Figs 1 and 3). Monkey FFR exhibits the same number of sustained phase-locked peaks

	Peak 1	Peak 3	Peak 4	Peak 5	Peak 6
Amplitude (uV)					
Human	4.63 (4.24–5.02) n = 18 × 6400	4.12 (3.75–4.53) n = 18 × 6400	2.68 (2.30–3.06) n = 18 × 6400	3.33 (2.94–3.72) n = 18 × 6400	4.94 (4.54–5.33) n = 18 × 6400
Monkey	18.91 (17.36–20.41) n = 2 × 6400	9.36 (7.87–10.85) n = 2 × 6400	6.59 (5.13–8.08) n = 2 × 6400	5.62 (4.17–7.09) n = 2 × 6400	5.23 (3.71–6.72) n = 2 × 6400
Latency (ms)					
Human	8.38 (8.38–8.39)	15.99 (15.98–15.99)	23.42 (23.41–23.42)	31.84 (31.83–31.85)	39.86 (39.86–39.87)
Monkey	9.32 (9.31–9.33)	15.44 (15.43–15.45)	25.28 (25.26–25.29)	34.69 (34.68–34.70)	43.31 (43.29–43.32)

Table 2. Average amplitude and latency across epochs of response peaks for the human and monkey groups. Data is shown as the mean and bootstrap 95% CI in parenthesis. The total number of epochs (n=) is indicated as subject *x* epochs.

consistently described in the human FFR[2–4] and occurring at similar latencies (Tables 1 and 2). Furthermore, a high-fidelity and stable representation of the temporal structure of sound occurs in the FFR of both species (Figs 2 and 4).

The periodic activity of the monkey FFR reproduces the F0 of the stimulus as observed in human responses, indicating comparable phase-locking abilities to fast temporal information such as subsyllabic cues in both species. Imaging studies in humans using the diphthong /da/ stimulus have demonstrated that auditory areas, including the cochlear nucleus, inferior colliculus, medial geniculate body and auditory cortex, contribute to the scalp-recorded FFR potential[49,50]. Furthermore, it has been shown that the lemniscal auditory pathway comprising the central nucleus of the inferior colliculus, ventral division of the medial geniculate body together with the core region of the auditory cortex has a strong contribution to the FFR[17]. Lemniscal auditory neurons are highly sensitive to the physical features of the sound exhibiting fast and high-fidelity responses which contrast with the habituating responses to unvarying stimuli of the non-lemniscal neurons[51–53]. Whether analogous auditory neural structures encode temporal periodicities in the order of milliseconds in both species remains to be addressed. Regardless, the principles of organization of auditory ascending pathway in which specialized neurons encoding the acoustical properties are tightly interconnected seems to be a general feature across mammals[51,52,54].

In the rhesus monkey, two ascending auditory pathways are identified: the 'direct' and 'indirect' pathways, both originating in the cochlear nucleus subdivisions and projecting directly or passing though the superior olivary complex to the inferior colliculus respectively[55–57]. Although there are no studies on the response of brainstem and midbrain neurons to periodic sounds in monkey, it has been shown that neurons of the cochlear nucleus[58] and central nucleus of the inferior colliculus[59] of guinea pigs exhibit a similar pattern of periodic activity as synthetic speech sounds. In rhesus monkey, cortical phase-locked activity to periodic speech sounds including vowels[60–62] and the /da/ syllable[63] has been recorded in the primary auditory cortex. Those studies showed that cortical neurons can resolve the spectral fine-structure (harmonics) as well as the envelope of speech sounds using a 'rate-place' code[60–63]. Furthermore, the lag between the periodic peaks of response and /da/ syllable was of 8.5 ms[63], similar to the lag observed in the scalp-recorded FFR (Fig. 2). Overall, animal studies indicate that neurons along the mammalian auditory ascending pathway exhibit a phase-locked patterning of activity that follows the acoustic periodicities in the order of milliseconds.

Further intracerebral recordings are needed to determine the contribution of each of the auditory structures on the scalp-recorded monkey FFR, as well as its differences with the human FFR. Monkey FFR exhibits larger peak amplitude and latency (Fig. 5, Tables 1 and 2). Early[64] and middle latency evoked responses[65] in monkeys also have a larger amplitude than human responses, which could be partially explained by the smaller thickness of the monkey skull and larger ratio between sizes of electrode and neural generators. The longest latencies of monkey FFR (Fig. 5, Tables 1 and 2) could reflect slower dynamics of activation of auditory generators in response to acoustic periodicities. Although, a second possibility is that longer latencies resulted from larger and more frequent movement artifacts occurring in monkey recordings, since animals were not head-restrained and could move spontaneously. While peak 2 in human average response was small (Fig. 1, Table 1) and in some individual responses it was not observed, this peak was absent in both monkey FFRs. Peak 2 is a transient feature of human FFR associated with the transition between the broadband stop burst and onset of voicing associated with the /d/ and /a/ segment, respectively[44,47]. Thus, it is possible that humans have a larger sensitivity to burst-voice transition or that the analogous neural structures in monkeys elicit a weak response not captured by the scalp recording.

Importantly, a shared capacity to track isochronous regularities in the two primates is indicated by the inter-peak intervals and spectra analysis of the sustained features of human and monkey FFR. Nevertheless, recent studies have shown a gradient in the sensitivity to represent the metrical structure of complex sounds across the primate order[35,66–70]. Rhesus monkeys are able to detect regular acoustical metronomes[31,32] but not to perceive the pulse in rhythmic patterns with a complex meter as humans do[31]. Accordingly, the gradual audiomotor evolution hypothesis proposes that the meter hierarchy at which primate species can perceive and entrain movements may have developed as a consequence of a gradient of anatomofunctional changes in the auditory and motor systems[35]. Further comparative studies are needed to determine the stage at which the structure and function of circuits diverges to account for different meter hierarchy sensitivity.

Similarly, it is of timely interest to study how early encoding of acoustic regularities is exploited and modulated by downstream sensory and motor neural circuits to support perception and production of more complex
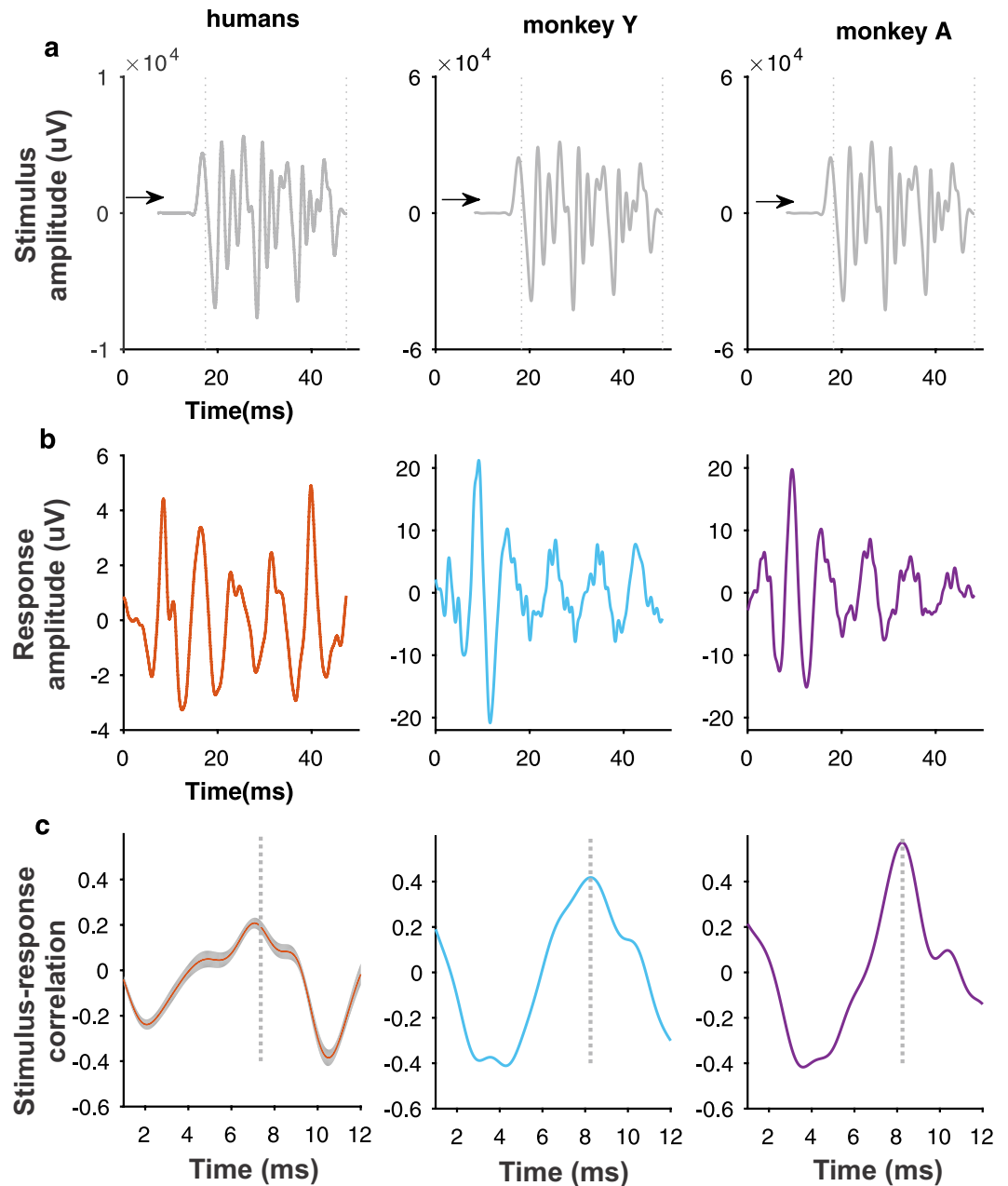
Figure 2. Stimulus-response correlation. (**a**) Low-pass filtered version of the stimulus /da/ syllable. Stimulus onset is shifted in time (arrow) to account for the time lag resulting in maximal correlation between signals with the aim of maximizing the visual coherence between stimulus and response waveforms. The dotted lines delimit the sustained part of the stimulus (10–40 ms). (**b**) FFR waveform of human subjects (grand average) and macaques (individual averages). (**c**) Correlation coefficients between stimulus and response as a function of the time shift between them. The maximal correlation is reached at time displacement of 7–8 ms indicated by the dotted line.

temporal patterns[69–71]. Human studies suggest an overlap between auditory circuits underlying the FFR and higher level neural circuits involved in beat entrainment. It has been observed that subjects that can entrain, with high precision, to an external beat have a larger phase locking and inter-trial consistency of the FFR potential[12,72]. These observations strongly suggest that the precise representation of the temporal structure of sound occurring at low hierarchical levels are used to plan and align motor outputs in phase to the stimulus[17,70,73,74]. Furthermore, top-down signals also sharpen the early processing of temporal regularities, as revealed by the finding that FFR strength is modified by short-term training protocols[23,75] and on-line cognitive factors[46,76].

Non-human primates are able to perceive and respond predictively to isochronous auditory sequences[33,36,37] and notably, the stimulus-response correlation and the neural consistency of monkey FFR was comparable (Fig. 4b) to human indices, reflecting a stable and faithful representation of the regular temporal structure of sound over time in both species. Therefore, the neurophysiological basis of audio-motor entrainment[34,36,37,77]
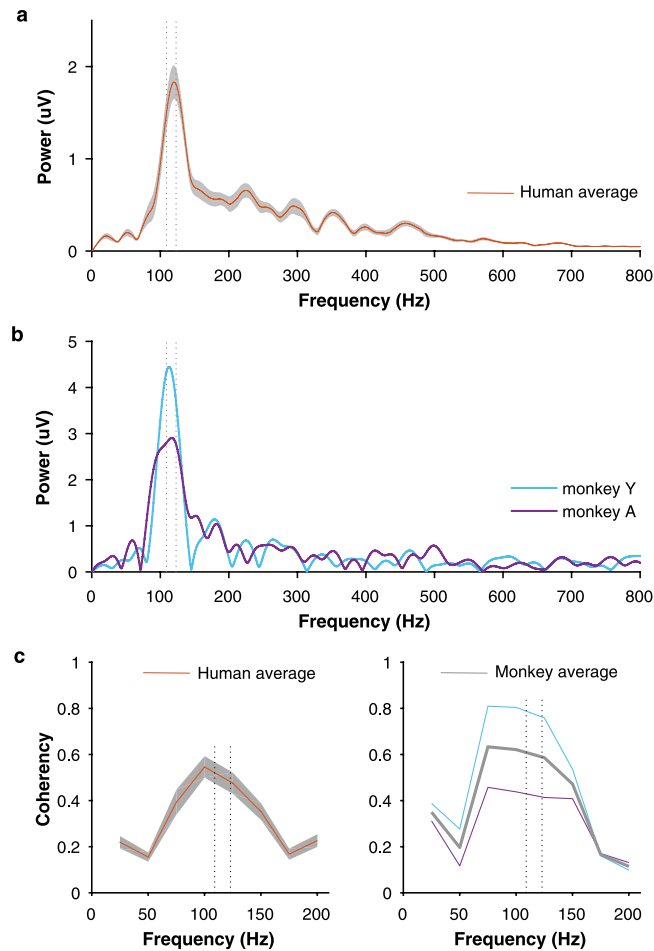
Figure 3. Spectral amplitude of the FFR. (**a**) Average frequency spectra of the FFR from the human population. (**b**) Frequency spectra of the individual FFR of each macaque. (**c**) Coherency index between the FFR segment and /da/ stimulus for human subjects (left) and monkeys (right). The mean ± SEM of frequency spectrum and coherency are shown for human data. The gray line shows the coherency average of the two macaque values. Dotted lines illustrate the F0 of the stimulus (103–121 Hz).

and the associated experience-dependent plasticity of the FFR potential can be studied in the behaving monkey to understand the mechanisms behind the correlation between the extraction of periodic auditory features and precision of synchronization performance observed in humans.

In summary, our findings reveal a conserved neural tracking accuracy for stimulus regularities between human and non-human primates. Thus, the rhesus monkey can be a suitable animal model not only to investigate the neurophysiological basis of the FFR, but also to study the neural mechanisms behind the association between beat entrainment abilities and experience-dependent auditory plasticity.

Methods

Subjects.   Non-invasive electroencephalographic activity was recorded on two male rhesus monkeys (*Macaca mulatta*, 10 and 9 years old, 6–7 kg BW) and 18 human subjects (mean age ± SD: 26 ± 4 years, 10 females) while passively listening to free-field auditory stimuli as previously demonstrated to elicit reliable FFR responses[44]. Subjects had no history of hearing disorders and all gave written informed consent. Experimental procedures on human subjects were approved by the UNAM Research Ethics Committee and were compliant with the Declaration of Helsinki. All the animal care, housing, and experimental procedures were approved by the National University of Mexico Institutional Animal Care and Use Committee and strictly conformed to the principles outlined in the Guide for Care and Use of Laboratory Animals (NIH, publication number 85–23, revised 1985). The researchers and animal care staff monitored the two monkeys daily to ensure their health and well-being. To ameliorate their condition of life, we routinely introduced toys (often containing food items that they liked) to their home cage (1.3 m³) environment to promote their exploratory behavior.

Acoustic stimulus.   The acoustic stimulus was the 40 ms consonant–vowel /da/ syllable, one of the most extensively used speech sounds in FFR human studies[2]. The stimulus has an F0 that rises linearly from 103 to 121 Hz; voicing begins at 5 ms and onset noise burst occurs during the first 10 ms. The acoustic stimulus
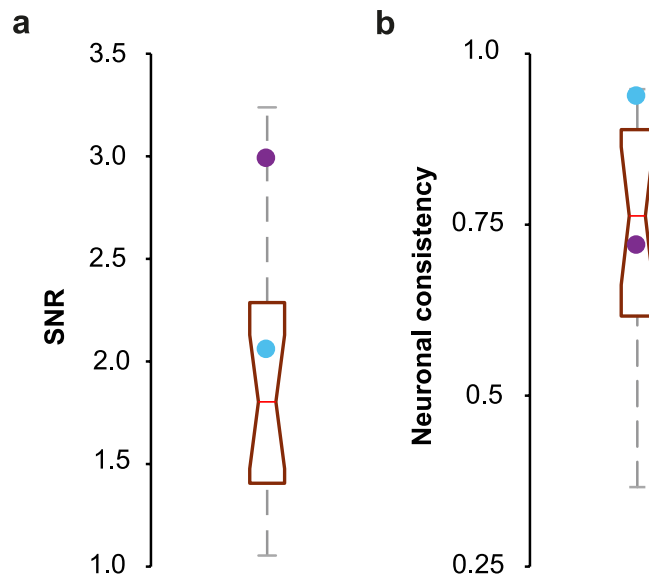
Figure 4. Intrinsic measures of the FFR. (**a**) Signal-to-noise ratio (SNR) between the FFR segment (20–60 ms) and baseline period (−10 ms) estimated by root-mean-square values. Data is shown as a box plot for the human data and as discrete values for monkeys (blue: monkey Y; purple, monkey A). The red line within each box represents the median values, the edges of the box delimit the 25th and 75th percentiles and the whiskers indicate the 10th and 90th percentiles. (**b**) Neural consistency index estimated by correlating the response waveforms resulting from averaging the odd and even epochs.
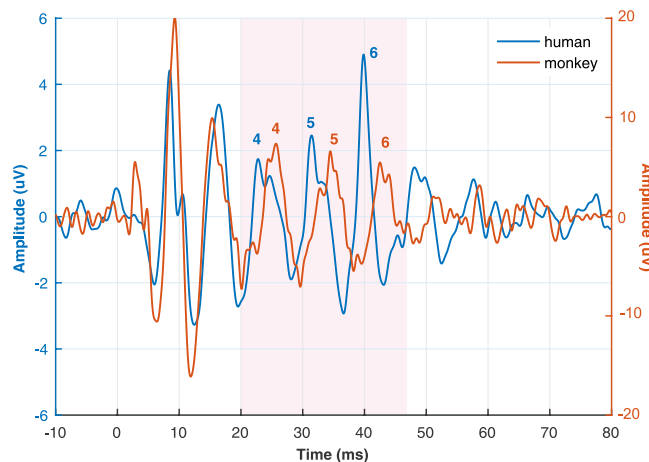

Figure 5. Human and monkey FFR potential. Average FFR across human subjects (n = 18) and monkeys (n = 2) recorded in the head vertex. The overall morphology of both primate responses exhibited three sustained peaks labeled as 4–6 reflecting the phase-locked activity of large neural ensembles to the fundamental frequency of the /da/ stimulus. Note the absence of peak 2 and larger amplitude and latencies in monkey FFR.

was resampled at 24414 Hz and a cosine ramp of 2 ms was applied at the beginning and end of the stimulus (*GenerateEnvelope*, Psychoacoustics ToolBox, MathWorks, Natick, MA, USA)[47].

Stimulation procedure. Human subjects and monkeys were awake; head unrestrained and comfortably seated inside an electrically isolated and sound-attenuated room without performing any behavioral task. To prevent drowsiness and minimize motion, human subjects watched a silent subtitled movie from a laptop running on battery power. The /da/ syllable was binaurally delivered through magnetically shielded studio monitors (KRK 5-G3, USA) with a 55-cm distance between the speakers and the subject's ears. Blocks of 2,000 stimuli in alternating positive and negative polarity at a rate of 11.1 Hz and an intensity of 85 dB SPL were delivered. Sound intensity was measured with an omnidirectional microphone (M 101, Beyer Dynamic, Germany) placed at the site of subject's head. The /da/ stimulus blocks were randomly interspersed among blocks of other simple (click) and complex sounds (human word and macaque vocalizations) in two recording sessions per subject. Sound delivery was controlled through an RZ2 TDT BioAmp Processor (TDT, System 3) using customized RPvdsEx

circuits. The audio file (.wav) containing the /da/ stimulus was loaded into RPvdsEx buffers and triggered at the desired rate and voltage.

Recording set-up.   Scalp-recorded activity was collected with Grass gold-cup electrodes (Natus Neurology, #FS-E5GH-60) at Fz, Cz, Pz, F3 and F4 sites according to the 10/20 EEG system. The reference and ground electrodes were placed on the inion and forehead, respectively. Electrodes were attached to the subject's scalp using conductive EEG paste (Ten20, D.O. Waver and Company, USA). The scalp of the macaques was shaved and the scalp of human subjects was cleaned with mild abrasive gel (Nuprep, D.O. Weaver and Company, USA) before each recording session to reduce scalp impedance. Here, the signal recorded at the Cz site was analyzed since it exhibited the largest SNR and it has been reported in previous EEG monkey studies[31]. Furthermore, artifacts due to the postauricular muscle reflex are diminished by using the vertex-inon derivation[78].

Signal acquisition.   The EEG signal, as well as the sound trigger and waveform, were simultaneously acquired at a sampling rate of 24414.0625 Hz using the same RZ2 TDT BioAmp Processor (TDT System 3) that controlled the presentation of stimuli. The signal was re-referenced to the signal recorded at inion and amplified using the Medusa preamplifier (RA16PA, TDT systems). Next, the EEG signal was band-pass filtered between 1–3000 Hz, and a notch filter at 60 Hz was applied to the signal to remove the line frequency using RPvdsEx TDT filter components. The signal timeline was corrected by 1.75 ms to compensate for the delays introduced by the sound delivery set-up.

Signal processing.   A total of 8,000 stimulus presentations were recorded for macaques and human subjects. Half the set of epochs corresponded to stimuli of positive polarity and the other half to stimuli of negative polarity. Twenty percent of epochs having the largest amplitude were discarded from each individual set of epochs regardless of the stimulus polarity to remove muscle and movement artifacts that ensure a stable SNR[50]. Next, the signal was further band-pass filtered (100–1000 Hz, Butterworth 4th-order filter) and epoched with respect to stimulus onset (−10–80 ms). EEG epochs were sorted into positive and negative polarity stimulus presentations and sub-averages were computed for each polarity condition. The added polarity grand average was obtained and used for further analysis to avoid any transduction stimulus artifact and to minimize the cochlear and microphonic potentials[2]. The median followed by the 25–75th IQR are indicated for human data while individual values are reported for monkeys.

Signal analysis.   FFR analysis was conducted using custom routines in Matlab 2016b (The MathWorks Inc, Natick, Massachusetts, US) and built functions from the Brainstem Toolbox (BT_2013) and Chronux Software (Chronux 2.12). The EEG signal was analyzed following measures of acoustic and intrinsic properties, as well as of response reliability typically employed in FFR studies in open-[6] and close-field stimulation experiments[45].

Acoustic properties
Peak amplitude and latency.   Prominent, on-going peaks were identified following the nomenclature used by Lehmann and colleagues in a previous FFR study in humans using open-field stimulation[44]. For human responses, we used the latency of the peaks at the grand average to guide the peak identification in individual responses. A peak was coded 'reliable' and included in the analysis if visual observation by two raters revealed that the peak's amplitude was above the pre-stimulus amplitude. The amplitude of each peak was estimated as the mean value from a window of ±0.5 ms around the maximum value. The latency was estimated as the time relative to the onset of the stimulus corresponding to the maximum amplitude. The response window containing the FFR segment was set at 20 to 60 ms in both the human and monkey response.

Temporal correlation between stimulus and response waveform.   The FFR segment was correlated with the 10–40 ms portion of a low-pass filtered version of the stimulus (500 Hz, Butterworth 3rd-order filter) that includes its harmonic segment[7]. The low-pass filtered version of the stimulus was employed to accentuate the envelope of the stimulus[7]. The function *bt_xcorrelation* of the BT_2012 Toolbox was used to calculate Pearson's correlation coefficients.

Spectral amplitude.   Response amplitudes in the spectral domain were computed for the corresponding FFR portion using the *bt_fftsc* function from the BT_2013 toolbox[2]. A 2 ms on– 2 ms off Hanning ramp was applied to the waveform. Zero padding was employed to increase the number of frequency points where spectral estimates were obtained. Average spectral amplitude was calculated for two different frequency ranges around the fundamental frequency (F0: 90–140 Hz) and its harmonics (H: 200–750 Hz)[44]. To estimate the strength of association between the spectral content of the stimulus (0–40 ms) and the response at F0, a coherency index was estimated using the *coherencyc* function of Chronux Toolbox[79]. Coherency is the cross-spectrum magnitude of stimulus and response signals at frequency index *j*, divided by the power spectrum magnitude of each signal at frequency index *j*. For this aim, response subaverages of 800 epochs randomly selected with replacement were calculated for each subject.

Intrinsic properties
Root mean square and signal-to-noise ratio.   To estimate the response strength, we estimated the magnitude of neural activation in the FFR and 10 ms pre-stimulus period using the root mean squared values (RMS). Subsequently, the SNR was estimated as the ratio between the RMS of the FFR segment and the pre-stimulus segment[2]. The RMS and SNR were estimated using the *bt_rms* function of BT_2013 toolbox.

Neural consistency.   Neural consistency was estimated for individual FFRs to assess the extent to which the brainstem's representation of sound varies from trial to trial. The FFR portions of the neural signal from each subject were split into even-odd halves taking into consideration both stimulus polarities[44]. In detail, the epochs were first sorted into positive- or negative-polarity trials, then, the epochs of each set of polarity were further split into odd or even trials according to their serial order within each set. Finally, even-ordered positive polarities were averaged, even-ordered negative polarities were averaged and those two were further averaged into an even-ordered FFR sub-average. The same procedure was done for odd-ordered epochs. Finally, the consistency index was computed as the correlation between the even-ordered FFR sub-average and odd-ordered FFR sub-average, with coefficient values close to one representing a more consistent response[44,48].

Comparison between human and monkey data.   Human and monkey groups were compared using the bootstrapping method on trial-by-trial estimations of (1) response peak amplitude and latency, (2) power spectrum and (3) SNR of the FFR segment. This method relies on random sampling with replacement to determine the confidence intervals of sample estimates[80]. A total of 10,000 bootstrap samples were obtained. Non-overlapping 95% CIs between groups indicated significant difference. For the power spectrum and SNR analysis, the human group comprised 18 subjects ×6,400 epochs, while for the response peak analysis, the number of subjects varied according to the identification (reliable or not reliable) of each peak (indicated in Table 2). The monkey group comprised two subjects ×6,400 epochs.

References

1. Moore, B. C. & Gockel, H. E. Properties of auditory stream formation. *Philos Trans R Soc Lond B Biol Sci* **367**, 919–931, https://doi.org/10.1098/rstb.2011.0355 (2012).
2. Skoe, E. & Kraus, N. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* **31**, 302–324, https://doi.org/10.1097/AUD.0b013e3181cdb272 (2010).
3. Kraus, N., Anderson, S. & White-Schwoch T. *The Frequency-Following Response: A Window into Human Communication* (eds Kraus, N. *et al.*) Ch. 1, 1–15 (Springer-Nature, 2017).
4. Chandrasekaran, B. & Kraus, N. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* **47**, 236–246, https://doi.org/10.1111/j.1469-8986.2009.00928.x (2010).
5. Aiken, S. J & Picton, T. W. Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* **245**, 35–47, https://doi.org/10.1016/j.heares.2008.08.004 (2008).
6. Wible, B., Nicol, T. & Kraus, N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* **67**, 299–317, https://doi.org/10.1016/j.biopsycho.2004.02.002 (2004).
7. Akhoun, I. *et al.* The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* **119**, 922–933, https://doi.org/10.1016/j.clinph.2007.12.010 (2008).
8. White-Schwoch, T. *et al.* Auditory Processing in Noise: A Preschool Biomarker for Literacy. *PLoS Biol* **13**, e1002196, https://doi.org/10.1371/journal.pbio.1002196 (2015).
9. Kraus, N. & Anderson, S. *Translational Research in Audiology, Neurotology, and the Hearing Sciences* (eds Le Prell, C. G. *et al.*) Ch. 3, 51–80 (Springer International Publishing, 2016).
10. Banai, K. *et al.* Reading and subcortical auditory function. *Cereb Cortex* **19**, 2699–2707, https://doi.org/10.1093/cercor/bhp024 (2009).
11. Abrams, D. A., Nicol, T., Zecker, S. & Kraus, N. Abnormal cortical processing of the syllable rate of speech in poor readers. *J Neurosci* **29**, 7686–7693, https://doi.org/10.1523/JNEUROSCI.5242-08.2009 (2009).
12. Tierney, A., White-Schwoch, T., MacLean, J. & Kraus, N. Individual Differences in Rhythm Skills: Links with Neural Consistency and Linguistic Ability. *J Cogn Neurosci* **29**, 855–868, https://doi.org/10.1162/jocn_a_01092 (2017).
13. Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L. & Kraus, N. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc Natl Acad Sci USA* **111**, 14559–14564, https://doi.org/10.1073/pnas.1406219111 (2014).
14. Woodruff Carr, K., Fitzroy, A. B., Tierney, A., White-Schwoch, T. & Kraus, N. Incorporation of feedback during beat synchronization is an index of neural maturation and reading skills. *Brain Lang* **164**, 43–52, https://doi.org/10.1016/j.bandl.2016.09.005 (2017).
15. Tierney, A. & Kraus, N. Getting back on the beat: links between auditory-motor integration and precise auditory processing at fast time scales. *Eur J Neurosci* **43**, 782–791, https://doi.org/10.1111/ejn.13171 (2016).
16. Tierney, A. & Kraus, N. Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Front Hum Neurosci* **8**, 949, https://doi.org/10.3389/fnhum.2014.00949 (2014).
17. Kraus, N. & White-Schwoch, T. Unraveling the Biology of Auditory Learning: A Cognitive-Sensorimotor-Reward Framework. *Trends Cogn Sci* **19**, 642–654, https://doi.org/10.1016/j.tics.2015.08.017 (2015).
18. Kraus, N. *et al.* Auditory biological marker of concussion in children. *Sci Rep* **6**, 39009, https://doi.org/10.1038/srep39009 (2016).
19. Kraus, N. *et al.* The neural legacy of a single concussion. *Neurosci Lett* **646**, 21–23, https://doi.org/10.1016/j.neulet.2017.03.008 (2017).
20. Slabu, L., Grimm, S. & Escera, C. Novelty detection in the human auditory brainstem. *J Neurosci* **32**, 1447–1452, https://doi.org/10.1523/JNEUROSCI.2557-11.2012 (2012).
21. Shiga, T. *et al.* Deviance-Related Responses along the Auditory Hierarchy: Combined FFR, MLR and MMN Evidence. *PLoS One* **10**, e0136794, https://doi.org/10.1371/journal.pone.0136794 (2015).
22. Gorina-Careta, N., Zarnowiec, K., Costa-Faidella, J. & Escera, C. Timing predictability enhances regularity encoding in the human subcortical auditory pathway. *Sci Rep* **6**, 37405, https://doi.org/10.1038/srep37405 (2016).
23. Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A. & Kraus, N. Auditory training improves neural timing in the human brainstem. *Behav Brain Res* **156**, 95–103, https://doi.org/10.1016/j.bbr.2004.05.012 (2005).
24. Sweetow, R. & Palmer, C. V. Efficacy of individual auditory training in adults: a systematic review of the evidence. *J Am Acad Audiol* **16**, 494–504 (2005).
25. Song, J. H., Skoe, E., Wong, P. C. & Kraus, N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* **20**, 1892–1902, https://doi.org/10.1162/jocn.2008.20131 (2008).
26. Krishnan, A., Xu, Y., Gandour, J. T. & Cariani, P. A. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* **189**, 1–12, https://doi.org/10.1016/S0378-5955(03)00402-7 (2004).
27. White-Schwoch, T., Woodruff Carr, K., Anderson, S., Strait, D. L. & Kraus, N. Older adults benefit from music training early in life: biological evidence for long-term training-driven plasticity. *J Neurosci* **33**, 17667–17674, https://doi.org/10.1523/JNEUROSCI.2560-13.2013 (2013).
28. Kraus, N., Skoe, E., Parbery-Clark, A. & Ashley, R. Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann N Y Acad Sci* **1169**, 543–557, https://doi.org/10.1111/j.1749-6632.2009.04549.x (2009).
29. Sinnott, J. M., Beecher, M. D., Moody, D. B. & Stebbins, W. C. Speech sound discrimination by monkeys and humans. *J Acoust Soc Am* **60**, 687–695 (1976).

30. Sinnott, J. M. & Kreiter, N. A. Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans. *J Acoust Soc Am* **89**, 2421–2429 (1991).
31. Honing, H., Merchant, H., Haden, G. P., Prado, L. & Bartolo, R. Rhesus monkeys (Macaca mulatta) detect rhythmic groups in music, but not the beat. *PLoS One* **7**, e51369, https://doi.org/10.1371/journal.pone.0051369 (2013).
32. Selezneva, E. *et al.* Rhythm sensitivity in macaque monkeys. *Front Syst Neurosci* **7**, 49, https://doi.org/10.3389/fnsys.2013.00049 (2013).
33. Zarco, W., Merchant, H., Prado, L. & Mendez, J. C. Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J Neurophysiol* **102**, 3191–3202, https://doi.org/10.1152/jn.00066.2009 (2009).
34. Merchant, H., Zarco, W., Perez, O., Prado, L. & Bartolo, R. Measuring time with different neural chronometers during a synchronization-continuation task. *Proc Natl Acad Sci USA* **108**, 19784–19789, https://doi.org/10.1073/pnas.1112933108 (2011).
35. Merchant, H. & Honing, H. Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front Neurosci* **7**, 274, https://doi.org/10.3389/fnins.2013.00274 (2013).
36. Crowe, D. A., Zarco, W., Bartolo, R. & Merchant, H. Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *J Neurosci* **34**, 11972–11983, https://doi.org/10.1523/JNEUROSCI.2177-14.2014 (2014).
37. Merchant, H. *et al.* Sensorimotor neural dynamics during isochronous tapping in the medial premotor cortex of the macaque. *Eur J Neurosci* **41**, 586–602, https://doi.org/10.1111/ejn.12811 (2015).
38. Yu, L. & Tomonaga, M. Interactional synchrony in chimpanzees: Examination through a finger-tapping experiment. *Sci Rep* **5**, 10218, https://doi.org/10.1038/srep10218 (2015).
39. Large, E. & Palmer, C. Perceiving temporal regularity in music. *Cogn Sci* **26**, 1–37, https://doi.org/10.1207/s15516709cog2601_1 (2002).
40. Nagasaka, Y., Chao, Z. C., Hasegawa, N., Notoya, T. & Fujii, N. Spontaneous synchronization of arm motion between Japanese macaques. *Sci Rep* **3**, 1151, https://doi.org/10.1038/srep01151 (2013).
41. Takeya, R., Kameda, M., Patel, A. D. & Tanaka, M. Predictive and tempo-flexible synchronization to a visual metronome in monkeys. *Sci Rep* **7**, 6127, https://doi.org/10.1038/s41598-017-06417-3 (2017).
42. Strominger, N. *Sensory systems of primates* (ed. Noback, C. R.) Ch. 3, 53–91 (Springer, 1978).
43. Baumann, S., Petkov, C. I. & Griffiths, T. D. A unified framework for the organization of the primate auditory cortex. *Front Syst Neurosci* **7**, 11, https://doi.org/10.3389/fnsys.2013.00011 (2013).
44. Gama, N., Peretz, I. & Lehmann, A. Recording the human brainstem frequency-following-response in the free-field. *J Neurosci Methods* **280**, 47–53, https://doi.org/10.1016/j.jneumeth.2017.01.016 (2017).
45. Russo, N., Nicol, T., Musacchia, G. & Kraus, N. Brainstem responses to speech syllables. *Clin Neurophysiol* **115**, 2021–2030, https://doi.org/10.1016/j.clinph.2004.04.003 (2004).
46. Lehmann, A. & Schonwiesner, M. Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS One* **9**, e85442, https://doi.org/10.1371/journal.pone.0085442 (2014).
47. Lehmann, A., Skoe, E., Moreau, P., Peretz, I. & Kraus, N. Impairments in musical abilities reflected in the auditory brainstem: evidence from congenital amusia. *Eur J Neurosci* **42**, 1644–1650, https://doi.org/10.1111/ejn.12931 (2015).
48. Hornickel, J., Zecker, S. G., Bradlow, A. R. & Kraus, N. Assistive listening devices drive neuroplasticity in children with dyslexia. *Proc Natl Acad Sci USA* **109**, 16731–16736, https://doi.org/10.1073/pnas.1206628109 (2012).
49. Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S. & Zatorre, R. J. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun* **7**, 11070, https://doi.org/10.1038/ncomms11070 (2016).
50. Coffey, E. B., Musacchia, G. & Zatorre, R. J. Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. *J Neurosci* **37**, 830–838, https://doi.org/10.1523/JNEUROSCI.1265-16.2016 (2017).
51. Lee, C. C. & Sherman, S. M. Topography and physiology of ascending streams in the auditory tectothalamic pathway. *Proc Natl Acad Sci USA* **107**, 372–377, https://doi.org/10.1073/pnas.0907873107 (2010).
52. Lee, C. C. & Sherman, S. M. On the classification of pathways in the auditory midbrain, thalamus, and cortex. *Hear Res* **276**, 79–87, https://doi.org/10.1016/j.heares.2010.12.012 (2011).
53. Gruters, K. G. & Groh, J. M. Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Frontiers in neural circuits* **6**, 96, https://doi.org/10.3389/fncir.2012.00096 (2012).
54. Hackett, T. A. Information flow in the auditory cortical network. *Hear Res* **271**, 133–146, https://doi.org/10.1016/j.heares.2010.01.011 (2011).
55. Jones, E. G. Chemically defined parallel pathways in the monkey auditory system. *Ann N Y Acad Sci* **999**, 218–233 (2003).
56. Gray, D. T., Engle, J. R. & Recanzone, G. H. Age-related neurochemical changes in the rhesus macaque cochlear nucleus. *The Journal of comparative neurology* **522**, 1527–1541, https://doi.org/10.1002/cne.23479 (2014).
57. Engle, J. R., Gray, D. T., Turner, H., Udell, J. B. & Recanzone, G. H. Age-related neurochemical changes in the rhesus macaque inferior colliculus. *Frontiers in aging neuroscience* **6**, 73, https://doi.org/10.3389/fnagi.2014.00073 (2014).
58. Moore, T. J. & Cashin, J. L. Jr. Response of cochlear-nucleus neurons to synthetic speech. *J Acoust Soc Am* **59**, 1443–1449 (1976).
59. White-Schwoch, T., Nicol, T., Warrier, C. M., Abrams, D. A. & Kraus, N. Individual Differences in Human Auditory Processing: Insights From Single-Trial Auditory Midbrain Activity in an Animal Model. *Cereb Cortex*, 1–21, https://doi.org/10.1093/cercor/bhw293 (2016).
60. Fishman, Y. I., Steinschneider, M. & Micheyl, C. Neural representation of concurrent harmonic sounds in monkey primary auditory cortex: implications for models of auditory scene analysis. *J Neurosci* **34**, 12425–12443, https://doi.org/10.1523/JNEUROSCI.0025-14.2014 (2014).
61. Fishman, Y. I., Micheyl, C. & Steinschneider, M. Neural Representation of Concurrent Vowels in Macaque Primary AuditoryCortex. *eNeuro* **3**, https://doi.org/10.1523/ENEURO.0071-16.2016 (2016).
62. Fishman, Y. I., Micheyl, C. & Steinschneider, M. Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. *J Neurosci* **33**, 10312–10323, https://doi.org/10.1523/JNEUROSCI.0020-13.2013 (2013).
63. Steinschneider, M., Arezzo, J. & Vaughan, H. G. Jr. Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. *Brain Res* **198**, 75–84 (1980).
64. Lasky, R. E., Maier, M. M., Snodgrass, E. B., Laughlin, N. K. & Hecox, K. E. Auditory evoked brainstem and middle latency responses in Macaca mulatta and humans. *Hear Res* **89**, 212–225 (1995).
65. Uno, A., Kaga, K., Tsuzuku, T. & Kuroki, M. Middle-latency response of awake and anesthetized japanese macaques. *Audiology* **32**, 302–307 (1993).
66. Fitch, W. T. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front Syst Neurosci* **7**, 68, https://doi.org/10.3389/fnsys.2013.00068 (2013).
67. Brosch, M. & Scheich, H. *Primate Audition. Ethology and neurobiology* (ed. Ghazanfar A. A.) Ch. 9, 151–176 (CRC Press, 2003).
68. Patel, A. D. The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol* **12**, e1001821, https://doi.org/10.1371/journal.pbio.1001821 (2014).
69. Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y. & ten Cate, C. Searching for the origins of musicality across species. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140094, https://doi.org/10.1098/rstb.2014.0094 (2015).
70. Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M. & Fitch, W. T. Finding the beat: a neural perspective across humans and non-human primates. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140093, https://doi.org/10.1098/rstb.2014.0093 (2015).

71. Merchant, H. & Yarrow, K. How the motor system both encodes and influences our sense of time. *Current Opinion in Behavioral Sciences* **8**, 22–27 (2016).
72. Tierney, A. & Kraus, N. The ability to move to a beat is linked to the consistency of neural responses to sound. *J Neurosci* **33**, 14981–14988, https://doi.org/10.1523/JNEUROSCI.0612-13.2013 (2013).
73. Repp, B. H. Sensorimotor synchronization: a review of the tapping literature. *Psychon Bull Rev* **12**, 969–992 (2005).
74. Repp, B. H. & Su, Y. H. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon Bull Rev* **20**, 403–452 (2013).
75. Henshaw, H. & Ferguson, M. A. Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. *PLoS One* **8**, e62836, https://doi.org/10.1371/journal.pone.0062836 (2013).
76. Skoe, E. & Kraus, N. Hearing it again and again: on-line subcortical plasticity in humans. *PLoS One* **5**, e13645, https://doi.org/10.1371/journal.pone.0013645 (2010).
77. Merchant, H., Perez, O., Zarco, W. & Gamez, J. Interval tuning in the primate medial premotor cortex as a general timing mechanism. *J Neurosci* **33**, 9082–9096, https://doi.org/10.1523/JNEUROSCI.5513-12.2013 (2013).
78. Legatt, A. D. Electrophysiologic auditory tests. *Handbook of clinical neurology* **129**, 289–311, https://doi.org/10.1016/B978-0-444-62630-1.00017-2 (2015).
79. Mitra, P. & Bokil, P. *Observed Brain Dynamics*. (Oxford University Press, 2008).
80. Eforn, B. & Tibshirani, R. *An Introduction to the Bootstrap* (Chapman & Hall, 1993).

Acknowledgements

Author Contributions

Y.A.A., A.L., H.M. designed the study; Y.A.A., A.L., H.M. designed the data analysis tools; Y.A.A. performed the experiments and analyzed data; Y.A.A., A.L., H.M. wrote the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.