Supplemental Information

"Voice cells in the primate temporal lobe"

Perrodin, C., Kayser, C., Logothetis, N.K. & Petkov, C.I.

Author contribution statement: C.P. and C.I.P. conceived and designed the experiments. C.P. carried out the experiments and analyzed the data. C.P. and C.I.P. wrote the paper. C.K. and N.K.L. provided materials and comments on the manuscript.

Contents:

I. Supplemental Figures

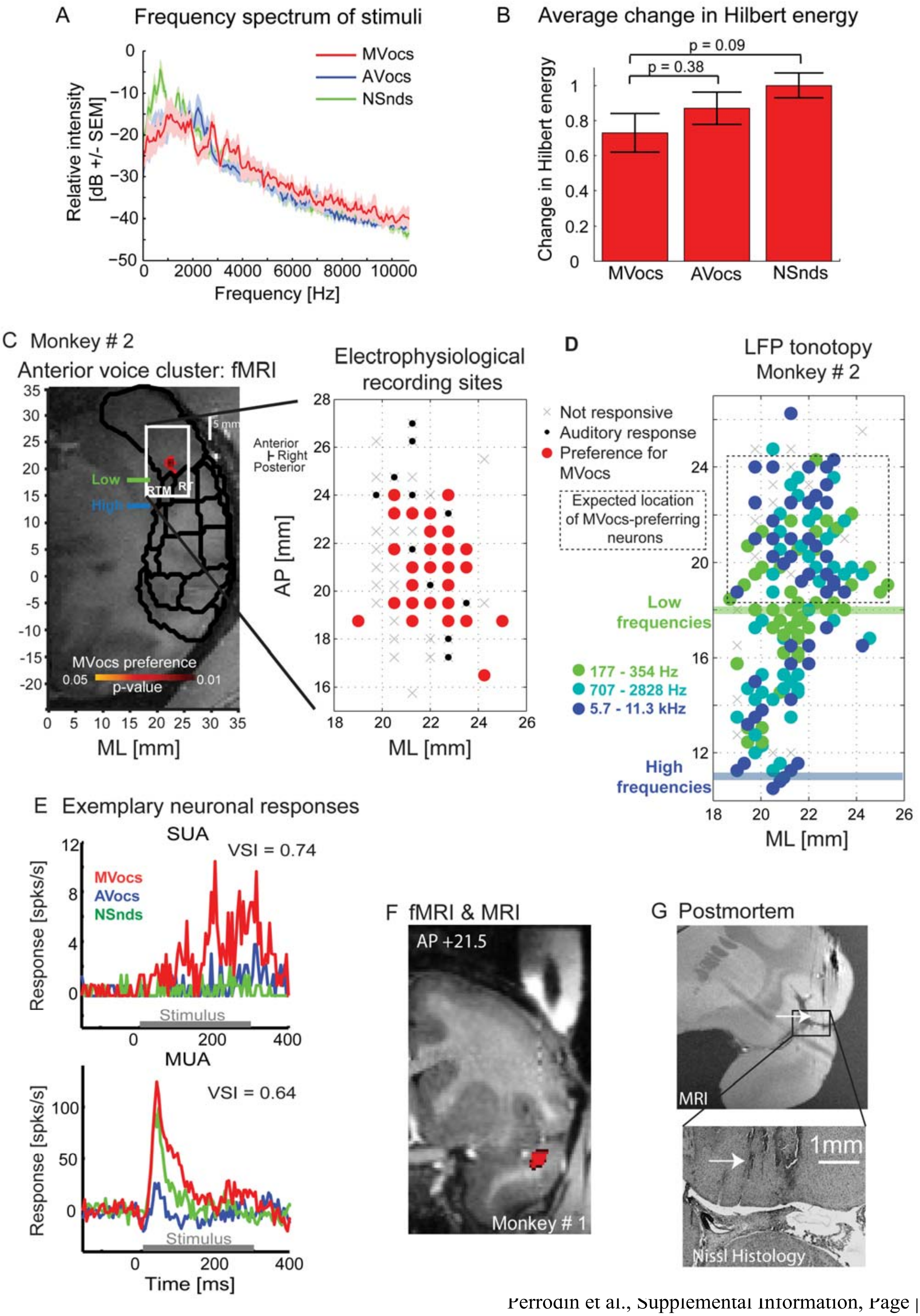
A
Frequency spectrum of stimuli
MVocs
AVocs
NSnds
Relative intensity [dB +/- SEM]
0
−10
−20
−30
−40
−50
0
2000
4000
6000
8000
10000
Frequency [Hz]
B
Average change in Hilbert energy
p = 0.09
p = 0.38
Change in Hilbert energy
0
0.2
0.4
0.6
0.8
1
MVocs
AVocs
NSnds
C
Monkey # 2
Anterior voice cluster: fMRI
35
30
25
20
15
10
5
0
-5
-10
-15
-20
Low
High
RTM
RT
5 mm
Anterior
Right
Posterior
MVocs preference
0.05
0.01
p-value
0
5
10
15
20
25
30
35
ML [mm]
Electrophysiological recording sites
AP [mm]
28
26
24
22
20
18
16
18
20
22
24
26
ML [mm]
D
LFP tonotopy
Monkey # 2
× Not responsive
● Auditory response
● Preference for MVocs
Expected location of MVocs-preferring neurons
Low frequencies
High frequencies
● 177 - 354 Hz
● 707 - 2828 Hz
● 5.7 - 11.3 kHz
24
20
16
12
18
20
22
24
26
ML [mm]
E
Exemplary neuronal responses
SUA
VSI = 0.74
MVocs
AVocs
NSnds
Response [spks/s]
12
8
4
0
Stimulus
0
200
400
MUA
VSI = 0.64
Response [spks/s]
100
50
0
Stimulus
0
200
400
Time [ms]
F
fMRI & MRI
AP +21.5
Monkey # 1
G
Postmortem
MRI
1mm
Nissl Histology

Figure S1: Stimuli and targeting the anterior voice-selective cluster, related to manuscript Figure 1.

(A) Frequency spectrum (mean ± SEM) of the 3 categories of sound stimuli used to evaluate the preference for conspecific voices. MVocs: conspecific macaque voices, AVocs: other animal voices, NSnds: natural/environmental sounds.

(B) Average change in Hilbert-extracted energy (power in the temporal envelope) for each sound category (mean ± SEM), normalized to the highest mean change.

(C) Targeting the anterior 'voice' cluster in monkey #2. Antero-posterior (AP) and medio-lateral (ML) coordinates are shown for the anterior fMRI MVocs-preferring cluster (left) and the electrophysiological recording sites (right). Format as in Figure 1B, including black outlines of auditory cortical fields.

(D) Auditory cortical field RT/RTM tonotopic localization in monkey #2. Figure layout is as in C, but note that the grid is expanded in the posterior direction to incorporate the tonotopic fields RT/RTM. For tonotopic localization we mapped LFP responses to one octave wide band-passed noise (with central frequencies ranging from 177 Hz to 11,314 Hz in one-octave steps). The responses were grouped into those that responded to Low (177-354 Hz), Middle (707-2828 Hz) and High (5.7-11.3 kHz) sound frequencies. The color code indicates, for each recording site whether the Low, Mid or High frequency range elicited the maximal ('best') response. Response amplitudes to each sound were calculated as the response averaged over the first 100ms after sound onset. The plot shows the anterior-to-posterior low-to-high frequency tonotopic gradient expected of auditory fields RTM/RT. Anterior to RTM/RT there is no clear tonotopic organization, which is the region labeled as 'expected location of MVocs-preferring neurons'.

(E) Exemplary 'voice' cell (SUA) and multi-unit activity (MUA) responses. Shown are the voice-selectivity index (VSI) values: VSI = (mean response to MVocs – mean response to others categories) / (mean response to MVocs + mean response to other categories). A VSI larger or equal to 1/3 indicates a response to MVocs at least twice larger than the response to the other sound categories.

(F) Coronal anatomical MRI slice (AP +21.5) of the right hemisphere in monkey #1, showing the recording chamber (white cylinder above brain), with a tubular marker (darker region in chamber) to mark the center of the chamber. This MRI is co-registered to the fMRI MVocs-preferring ('voice') cluster shown in red on the top of the temporal lobe and the anterior supra-temporal plane (STP).

(G) Monkey #1: Postmortem anatomical MRI (top left) and Nissl stain histological section (bottom right) registered to the coronal slice in (F). The postmortem MRI shows lesions from the electrode tracts leading to the top of the temporal lobe (e.g., white arrow); note that the gray/white matter contrast in this MRI anatomical is reversed relative to the MRI image in the left panel. The histological section is aligned to the approximate location of the imaging slice in black. The histological cutting angle does not exactly match that of the MRI, yet the lesion tracts left by the electrodes are evident with both the postmortem MRI and histology (e.g., white arrows).

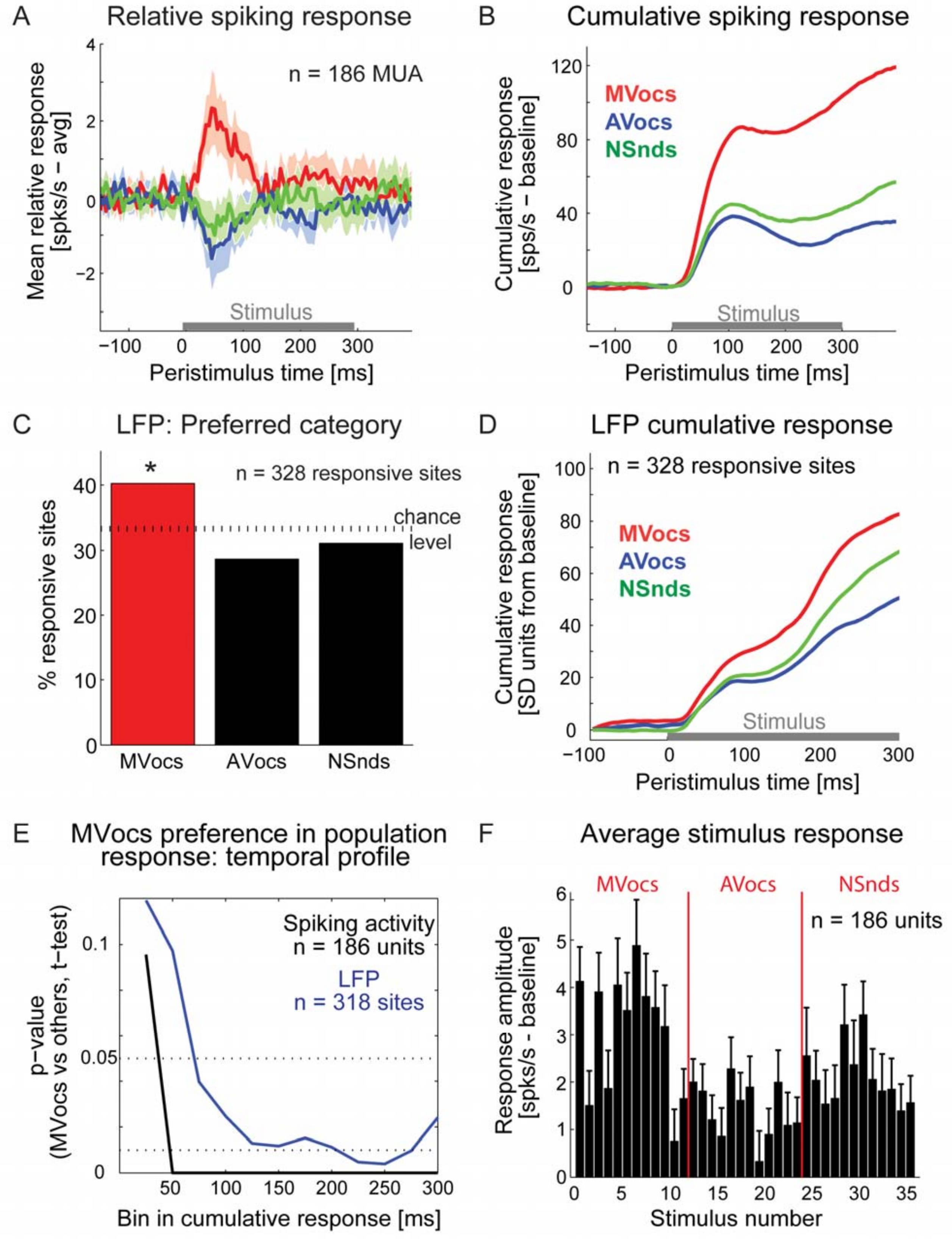

Figure S2: **Supporting results on the sound category preference of the neuronal responses, related to manuscript Figure 2.**

(A) Average relative spiking response across all responsive units. The color shading indicates the 95% confidence intervals for each response. The relative response (see *Electrophysiological Data Analysis* in the Supplemental Experimental Procedures) was calculated for each unit and each sound category by subtracting the mean auditory response to all categories from the average response to each category.

(B) Average cumulative spiking response across all responsive units, computed by summing, for each unit, the average spike count for each of the 3 sound categories over time, starting at the beginning of the baseline period.

(C) Proportion of auditory responsive local-field potentials (LFP) sites responding maximally to each sound category. A significant majority of sites prefer MVocs to other natural sound categories (χ2-test: p = 0.026).

(D) Average LFP cumulative response across all responsive sites, summing amplitude traces over time starting at the beginning of the baseline period.

(E) Temporal dynamics of the population preference for MVocs. The plot shows the p-value of the MVocs vs other comparison (paired *t*-test) as a function of consecutive 25ms bins in the cumulative response, for spiking (black line) and LFP activity (blue line).

(F) Histogram of the average response amplitude (spikes per second in a 200ms window around peak, baseline subtracted) to each of the 36 experimental sounds, across the population of auditory responsive units. Shown is mean ± SEM.

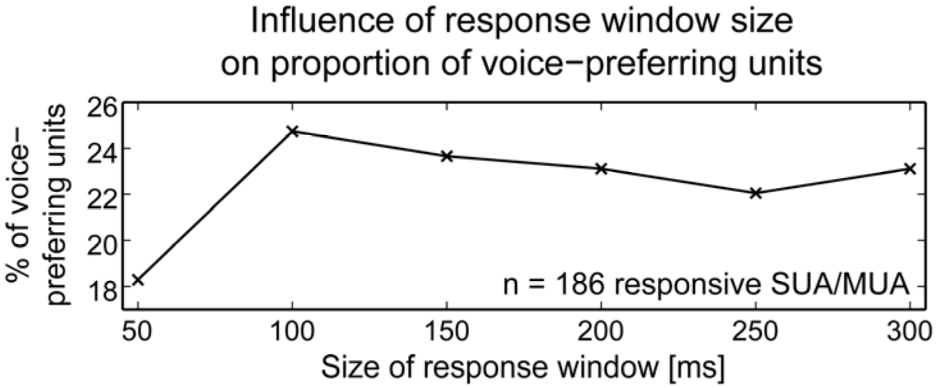

Figure S3: Supporting results on the proportion of voice-preferring units, related to manuscript Figure 3. Proportion of voice-preferring units (mean ± SEM: 22.2 ± 0.9 %) as a function of the size of the response window (centered on the peak category response). All response amplitudes reported in the study are computed using a 200ms response window, which, as shown here, appears to be a representative measure.

II. Supplemental Experimental Procedures

Animal procedures

Two adult male rhesus macaques (*Macaca mulatta*) participated in these experiments (M1 and M2, respectively). The macaques were from a group-housed colony. All procedures were approved by the local authorities (Regierungspräsidium Tübingen, Germany) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

Acoustical stimuli

All sounds were sampled at 22050 Hz, stored as WAV files, amplified using a Yamaha amplifier (AX-496), and delivered from 2 free-field speakers (JBL Professional, Northridge, CA), which were positioned at ear level 70 cm from the head and 50 degrees to the left and right. Sound presentation was calibrated using a condenser microphone (Brüel & Kjær 4188, Bremen, Germany) and sound level meter (Brüel & Kjær, 2238 Mediator) to ensure a linear (±4 dB) transfer function of sound delivery (between 88 Hz and 20 kHz).

To balance the acoustical features of the experimental sound categories while maintaining their ethological relevance, three categories of 12 complex natural sounds were sub-sampled from a larger set of vocalizations and natural/environmental sounds that we have previously used, see Experiment 1 in [7]. The categories of sounds consisted of the following: (1) macaque vocalizations from 12 different callers (MVocs); (2) other animal vocalizations from 12 different callers (AVocs); and, (3) 12 natural/environmental sounds (NSnds). The category of MVocs was recorded from many different macaques that were housed separately from the experimental animals and thus consisted of vocalizations unfamiliar to them prior to the start of the experiments. Similarly, the AVocs consisted of calls from many animals, and the macaques had not heard the AVocs and NSnds sounds prior to the start of the experiments.

The original larger sets of sounds consisted of 99 sounds (33 from each category). Sub-sampling was conducted to select 12 sounds from each category that best matched the average overall frequency spectrum of all the sounds combined. The sub-selection of the MVocs category was further constrained to result in a fairly balanced sampling of the following more commonly produced call types: 3 coos, 3 grunts, 2 barks, 2 harmonic arches and 2 screams (one tonal and one noisy). The sub-sampling of AVocs resulted in a mixture of hetero-specific primate (e.g., chimp call and monkey calls from other species) and bird, amphibian, wild and domesticated animal calls. The sub-sampling of NSnds resulted in a mixture of samples of running water, insect sounds, the flapping of bird wings, thunder and hands clapping. Importantly, each vocalization in the MVocs and AVocs categories was produced by a different individual, thereby consisting of a category of many voices (analogous to the categories of faces used to study face processing [1-5, 8, 39-42]). Moreover, these categories were composed of a mixture of commonly produced call types, to balance the impact of any particular form of referential information in the vocalizations [7].

The resulting 3 categories of 12 sounds did not significantly differ in their overall frequency spectrum (*t*-test, MVocs vs. AVocs, $p = 0.09$; MVocs vs. NSnds, $p = 0.56$; Fig. S1A) and had comparable fluctuation in their Hilbert extracted temporal envelope (*t*-test, MVocs vs. AVocs, $p = 0.38$; MVocs vs. NSnds, p = 0.09; Fig. S1B). The sound categories also did not statistically differ in their duration (mean duration: MVocs, 0.291 ms; AVocs, 0.300 ms; NSnds, 0.300 ms; *t*-test, MVocs vs. AVocs, $p = 0.44$; MVocs vs. NSnds, $p = 0.42$). The intensity of all of the sounds was normalized in RMS level and was calibrated at the position of the head to be presented at an average intensity of 65 dB SPL.

It was not possible to further control the stimulus acoustical features without affecting their ethological relevance. Moreover, whereas previously we have used acoustical control categories (such as phase-scrambled MVocs or noise shaped by the Hilbert envelope of the MVocs [7]), we did not use these acoustical control categories here since such acoustical controls appeared weakly effective in eliciting activity in hierarchically higher-level auditory cortical regions, see [7] and [26]. Thereby, we opted for spectro-temporally complex natural sounds to compare with the MVocs, such as the AVocs and NSnds sound categories.

fMRI localization of voice-preferring regions

The two macaques had previously participated in fMRI experiments to localize their voice-preferring regions, including the anterior voice clusters, which were identified as being voice-identity sensitive, see: [7]. All procedures have been detailed previously [7]. Briefly, M1 was trained on a visual fixation task during sound stimulation and scanning. This individual was scanned awake in a 7-Tesla MRI scanner (Bruker Medical). The fMRI scanning of the second macaque, M2, was conducted under anesthesia in a 4.7T MRI scanner; for details on the anesthesia and scanning parameters see [7]. Also see Fig. 1B and Fig. S1C which show the coordinates of the fMRI activity clusters that were used to guide the electrophysiological recordings. Here the fMRI activity cluster that prefers MVocs is analyzed using the MVocs > max [activity response of other sound categories] criterion, see [7], which is comparable to the max comparison of the electrophysiological data (Fig. 1B, 2A; S1C, S2C). Analyzing the fMRI results using the MVocs > average [activity response of other categories] comparison results in broader activation clusters. The 'max' comparison was preferred since it is a more conservative measure of MVocs preference and results in more focal clusters, the stereotactic coordinates of the centers of which were used to guide the electrophysiological recording electrodes.

Electrophysiological site localization with fMRI, MRI and histology

Prior to the experiments, a form-fitting, custom-designed head post was implanted during an aseptic surgical procedure[44]. Then in a later procedure a standard recording chamber was positioned based on the preoperatively obtained stereotaxic coordinates of the individual fMRI maps of the animals allowing access to the auditory regions on the supra-temporal plane (see Fig. 1A, B; S1C, F). The stereotactic coordinates used the Frankfurt-zero standard, where the origin is defined as the midpoint of the interaural line and the infraorbital plane. Different electrode penetration angles of 5, 10 and 15 degrees anterior were used as required to access and sufficiently sample the center and surrounding area of the fMRI identified cluster. The precise angle of the recording electrodes and depth to reach the center of the fMRI cluster were obtained by the stereotactic coordinates of the fMRI maps in each animal, and in monkey 1 (M1) by using the BrainSight neurosurgical targeting system which combines MRI- and fMRI-based markers (Rogue Research, Inc). BrainSight was also used after surgical recovery of the animal to confirm the target coordinates to the anterior fMRI-based voice cluster. In monkey 2 (M2), the anterior voice-preferring cluster directly neighbored the tonotopically organized fields RT and RTM that were posterior to it (Fig. S1C). Because of this close proximity to the tonotopically organized auditory cortex, in this animal we opted for electrophysiological mapping using band-passed noise stimuli, combined with data from a previous auditory study of his tonotopic organization of fields RT and RTM [45, 46]. These data further confirmed the electrophysiological location of the anterior voice cluster relative to the tonotopically organized fields RT/RTM (see Fig. S1C,D). Fig. S1D illustrates that no clear tonotopic organization was observed in the region of the expected location of the anterior MVocs-preferring cortex.

Coordinates were also confirmed at the completion of the experiments with postmortem MRI and histology (Fig. S1G). At the termination of all experiments, the animals were sedated with ketamine and deeply anesthetized with an overdose of pentobarbital (60-80 mg/kg i.v.). Immediately after euthanasia, they were perfused transcardially with 4 L of heparinized 0.9% saline followed by 4 L of 4% paraformaldehyde in phosphate buffer (PB, 0.1 M, pH 7.4). The brains were removed and stored at 4°C in phosphate buffer saline (PBS, 0.01 M, pH 7.4) containing 30% sucrose and 0.1% sodium azide.

Then a post-mortem MRI of the fixated brain stabilized in Agar gel was obtained using the following MRI parameters for M1: MSME sequence with 9.6 x 9.6 x 6.4 cm^3 field of view; matrix of 384 x 384 x 256 voxels; Echo time: 16 ms; Repetition time: 250 ms. The MRI parameters for M2 were: RARE sequence, FOV: 6.4 x 6.4 x 6.4 cm^3, matrix of 256 x 256 x 256 voxels; Echo time: 20 ms; Repetition time: 3000ms. Voxel size in both scans: 0.25mm.

Following the post-mortem MRI, the brain was cut into serial 100 micron-thick sections in the coronal plane on a horizontal freezing microtome and collected in PBS. The sections were then mounted onto microscope glass slides, dried overnight, colored with thionin blue using a standard Nissl stain protocol, after which the slides were coverslipped and photographed during microscopy. The histology sections were aligned to the MRI coordinates system using anatomical landmarks and a standard macaque brain atlas [47]. Nissl histology revealed the electrode tract lesions leading to the anterior supra-temporal plane, on the top of the temporal lobe (Fig. S1G).

Electrophysiological recording procedures

A custom-made multi-electrode system was used to independently advance up to 5 epoxy-coated tungsten microelectrodes (FHC Inc., Bowdoinham, ME; 0.8-2 MOhm impedance). Electrophysiological signals were amplified using an Alpha Omega amplifier system (Alpha Omega GmbH, Ubstadt Weiher, Germany), filtered between 4 Hz and 10 kHz (4-point Butterworth filter) and digitized at a 20.83 kHz sampling rate. Recordings were performed in a darkened and sound-insulated booth (Illtec, Illbruck Acoustic GmbH, Germany). The animals were awake during recordings and passively listened to the sounds. To help them to stay awake and listening to the sounds, the monkeys were intermittently rewarded with juice in between the electrophysiological recording trials and/or recording runs.

The electrodes were advanced down to the MRI-calculated depth of the anterior auditory cortex on the supra-temporal plane through an angled grid placed on the recording chamber. The coordinates of each electrode along the AP and ML axes were noted, as were the angle of the grid and the depth of the recording sites. During each recording session, each electrode was advanced to the depth of interest. Sites in the auditory cortex were distinguished from deeper recording sites in the superior-temporal sulcus (STS) using the depth of the electrodes, the crossing of the lateral sulcus that is devoid of neuronal activity (i.e., the occurrence of over 2 millimeters of white matter between auditory cortex and STS) and the prominence of visual responses in the STS. To further avoid recording from the STS or insula we excluded recording sites with depths that were too shallow, too deep or too medial based on the MRI coordinates.

Auditory LFP and/or spiking activity for recording was identified as follows. Experimental recordings were initiated if at least one electrode had LFP or neurons that could be driven by any of a large set of search sounds, including tones, frequency modulated sweeps, band-passed noise, clicks, musical samples and other natural sounds from a large library. No attempt was made to select neurons with a particular response preference. Rather any neuron or LFP site that appeared responsive to sound was recorded. Once a responsive site was isolated, the experiment began. The experimental sounds were then presented individually in randomized order, using a rapid stimulus presentation procedure (similar to [4]) with a randomly varying inter-stimulus interval ranging from 100 to 175ms. Each stimulus was repeated 20 times. After data collection was completed each electrode was advanced at least 250 micrometers to a new recording site and until the neuronal activity pattern considerably changed.

Electrophysiological data analysis

The data were analyzed in Matlab (Mathworks). The recorded broadband signal was separated into spiking activity by high-pass filtering the raw signal at 500 Hz. The low-frequency signal from 4-150Hz yielded the local-field potential.

Offline spike-sorting was performed using commercial spike sorting software (Plexon Offline Sorter; Plexon). The procedure involved the clustering of spike waveforms based on principal component analysis, valley height/peak amplitude and time information. We retained single units if the waveform signal-to-noise ratio (SNR) was larger than 4 (SNR = average waveform peak amplitude / average waveform standard deviation), combined with a clear refractory period (less than 1.5% of the total number of spikes occurring in the first 1.5ms following a spike). For other sites where the spike-sorting did not yield well separated clusters, the activity was combined into multi-unit activity (MUA). The results for single unit activity (SUA) and of combined single/multi-unit activity (SUA/MUA, which we refer to throughout as MUA) were separately analyzed and are described in the manuscript where they are reported. Spike times were saved at a resolution of 1ms. Peri-stimulus time histograms (PSTHs) were obtained using bins of 5 ms and 10 ms Gaussian smoothing (full-width at half-maximum, FWHM). A significant response to sensory stimulation (auditory-responsive activity) was determined by comparing the response amplitude of the average response to the response variability during the baseline period. Arithmetically this involved normalizing the average response to standard deviation units (SD) with respect to baseline (i.e., z-scores), and a response was regarded as significant if the z-score exceeded 2.5 SDs during a continuous period of at least 25ms during stimulus presentation. A unit was considered auditory responsive if it breached this threshold for any of the 36 experimental auditory stimuli.

The local-field potential (LFP, low-frequency range of the mean extracellular field potential) was obtained by low-pass filtering the raw data at 150 Hz (3rd order Butterworth filter). The signals were full-wave rectified. The auditory LFP response was obtained by normalizing the responses to units of standard deviations from baseline (see Fig. 2C). A recording site was deemed auditory responsive if the response breached 2.5 SD for a continuous period of at least 50ms for any of the sound stimuli.

For each response type, the mean of the baseline response was subtracted to compensate for fluctuations in spontaneous activity. Response amplitudes were defined by first computing the mean response for each category (MVocs, AVocs, NSnds) across trials and different sounds. For the category response, the peak of the category average response was calculated and the response amplitude was defined as the average response in a 200 ms window centered on the peak of the category average response. The preferred category for each unit was defined as the one eliciting the largest (maximal) response amplitude. 'Voice' cells were classified according to the face-preference criterion used in visual studies. In our case, the response to MVocs is defined as being at least twice larger than the response to the other categories. Formally this was based on the approach used in [4], as follows. We defined a 'voice selectivity index' as $VSI = \frac{mean(MVocs) - mean(others)}{mean(MVocs) + mean(others)}$ using the average response amplitudes to the different sound categories. A single unit was identified as a 'voice' cell if its VSI was larger or equal to 1/3, also see [4]. We also used two normalization procedures to account for the difference in firing rate between individual units. First we defined a relative firing rate as follows: for each unit the mean response across all three conditions was subtracted from the response to each individual condition (see Fig. S2A).

Finally, we computed a sparseness index [17] of the form $s = \frac{1- a}{1- \frac{1}{n}}$ where $a = \left(\frac{\left(\sum_{i=1}^{n} \frac{r_i}{n}\right)^2}{\sum_{i=1}^{n} \frac{r_i^2}{n}}\right)$, r_i is the trial-averaged, baseline-corrected response amplitude to the ith stimulus of the MVocs sound category and n is the total number of stimuli in that category (n = 12 calls in the MVocs category). For direct comparison of our results with the 'a' sparseness values reported for face cells [10], we normalized a to s for the results shown in Fig. 4D. Specifically, we used the values of a and n = 23 face stimuli as reported in [10] to obtain the corresponding values of s for face cells. Rolls and Tovee report standard deviation values for their population of 14 face cells, which we converted into standard error in Fig. 4D.

All auditory responsive units, distributed across an approx. 66mm^2 area on the anterior STP of both monkeys, were used in the analyses. Since this was a broad recording region and to allow better comparison to a recent visual study on face cells where the authors recorded from the center of a face cluster/patch [4], we defined in each monkey a focal cluster of the same dimensions as in the visual study, i.e., three adjacent grid holes (spacing of 0.75 mm) with the highest density of MVocs-preferring units. See the manuscript text and Fig. 3B for further details.

III. Supplemental References

44. Logothetis, N.K., Guggenberger, H., Peled, S., and Pauls, J. (1999). Functional imaging of the monkey brain. Nat Neurosci *2*, 555-562.

45. Kayser, C., Petkov, C.I., and Logothetis, N.K. (2008). Visual modulation of neurons in auditory cortex. Cereb Cortex *18*, 1560-1574.

46. Kayser, C., Petkov, C.I., and Logothetis, N.K. (2007). Tuning to sound frequency in auditory field potentials. J Neurophysiol *98*, 1806-1809.

47. Paxinos, G., Huang, X., and Toga, A.W. (2000). The Rhesus Monkey Brain in Stereotaxic Coordinates, (San Diego: Academic Press).