Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques

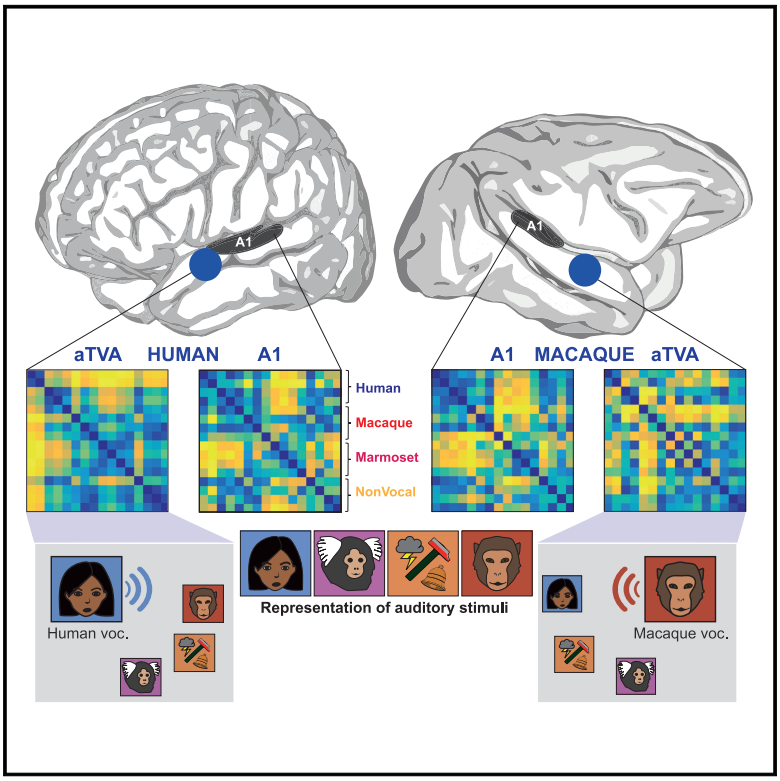
Clémentine Bodin, Régis Trapeau, Bruno Nazarian, Julien Sein, Xavier Degiovanni, Joël Baurberg, Emilie Rapha, Luc Renaud, Bruno L Giordano, Pascal Belin

Report

Current Biology

Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques

Graphical abstract



Highlights

- Both macaques and humans show voice-selective anterior temporal voice areas
- Similar representation of sounds in primary auditory cortex of both species
- The aTVAs categorize conspecific vocalizations apart from other sounds
- Functional homology in high-level auditory cortex of humans and macaques

Authors

Clémentine Bodin, Régis Trapeau, Bruno Nazarian, ..., Luc Renaud, Bruno L. Giordano, Pascal Belin

Correspondence

clementine.bodin@univ-amu.fr (C.B.), regis.trapeau@univ-amu.fr (R.T.), pascal.belin@univ-amu.fr (P.B.)

In brief

Bodin et al. report a functional homology in the cerebral processing of conspecific vocalizations by macaques and humans. Comparative fMRI reveals that both species possess bilateral anterior temporal voice areas that not only prefer conspecific vocalizations but also categorize them apart from all other sounds in a functionally homologous manner.

Bodin et al., 2021, Current Biology *31*, 1–6
November 8, 2021
https://doi.org/10.1016/j.cub.2021.08.043

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques

Clémentine Bodin,[1,4,*] Régis Trapeau,[1,4,*] Bruno Nazarian,[1] Julien Sein,[1] Xavier Degiovanni,[1] Joël Baurberg,[1] Emilie Rapha,[2] Luc Renaud,[2] Bruno L. Giordano,[1] and Pascal Belin[1,3,5,6,*]

[1]La Timone Neuroscience Institute, CNRS and Aix-Marseille University, UMR 7289, 27 bd Jean Moulin, 13005 Marseille, France
[2]Mediterranean Primate Research Center, CNRS, Aix-Marseille University, UAR 3537, 27 bd Jean Moulin, 13005 Marseille, France
[3]Psychology Department, Montreal University, C.P. 6128, succ. Centre-ville, Montreal, QC H3C 3J7, Canada
[4]These authors contributed equally
[5]Twitter: @BancoInt
[6]Lead contact
*Correspondence: clementine.bodin@univ-amu.fr (C.B.), regis.trapeau@univ-amu.fr (R.T.), pascal.belin@univ-amu.fr (P.B.)
https://doi.org/10.1016/j.cub.2021.08.043

SUMMARY

How the evolution of speech has transformed the human auditory cortex compared to other primates remains largely unknown. While primary auditory cortex is organized largely similarly in humans and macaques,[1] the picture is much less clear at higher levels of the anterior auditory pathway,[2] particularly regarding the processing of conspecific vocalizations (CVs). A "voice region" similar to the human voice-selective areas[3,4] has been identified in the macaque right anterior temporal lobe with functional MRI;[5] however, its anatomical localization, seemingly inconsistent with that of the human temporal voice areas (TVAs), has suggested a "repositioning of the voice area" in recent human evolution.[6] Here we report a functional homology in the cerebral processing of vocalizations by macaques and humans, using comparative fMRI and a condition-rich auditory stimulation paradigm. We find that the anterior temporal lobe of both species possesses cortical voice areas that are bilateral and not only prefer conspecific vocalizations but also implement a representational geometry categorizing them apart from all other sounds in a species-specific but homologous manner. These results reveal a more similar functional organization of higher-level auditory cortex in macaques and humans than currently known.

RESULTS AND DISCUSSION

We used comparative fMRI and scanned awake rhesus macaques (n = 3) and humans (n = 5) on the same 3T MRI scanner using an identical auditory stimulation paradigm. A first monkey was scanned using a block design, then a sparse-clustered scanning design was used in humans and two monkeys to ensure stimulus delivery in silent periods between volume acquisitions (Figure 1A). Auditory stimuli (n = 96) consisted of brief complex sounds sampled from 16 categories grouped in 4 larger categories: human speech and voice (n = 24), macaque vocalizations (n = 24), marmoset vocalizations (n = 24), and complex non-vocal sounds (n = 24) (Figure 1B). Marmoset vocalizations were included as a category of hetero-specific vocalizations unfamiliar to both humans and macaques, and because we plan to also run this protocol in marmoset monkeys.

General auditory activations in macaques and humans

The comparison of fMRI volumes acquired during sound stimulation versus the silent baseline revealed general auditory activation by the stimulus set. Both humans (Figure 1C) and macaques (Figures 1D, S1, and S2) showed extensive bilateral superior temporal gyrus (STG) activation ($p < 0.05$, corrected for multiple comparisons; *cf.* STAR Methods) centered in both species on core areas of the auditory cortex and extending rostrally and caudally to higher-level auditory cortex.

Primary auditory cortex (A1) of each subject was defined from a probabilistic map of Heschl's gyrus[8] in humans, and from a tonotopy-based parcellation of auditory cortex[7] in monkeys. A1 in both macaques and humans showed robust response profiles across the 16 stimulus categories (Figure 1E) that correlated well across hemispheres particularly in humans and with borderline significance in macaques (human A1, left versus right, median bootstrapped Spearman's rho = 0.708, bootstrapped p = 0.0012, below the Bonferroni-corrected threshold of p = 0.05/8 = 0.0063; macaque A1, left versus right, rho = 0.608, p = 0.007, n.s.). Response profiles also correlated across species in the left, but not right, hemisphere (left A1, humans versus macaques, rho = 0.62, p = 0.006; right A1, humans versus macaques, rho = 0.485, p = 0.029, n.s.). That overall similarity is in agreement with the wealth of anatomical and physiological studies showing that the functional architecture of primary auditory areas is well conserved across primates.[9,10]

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

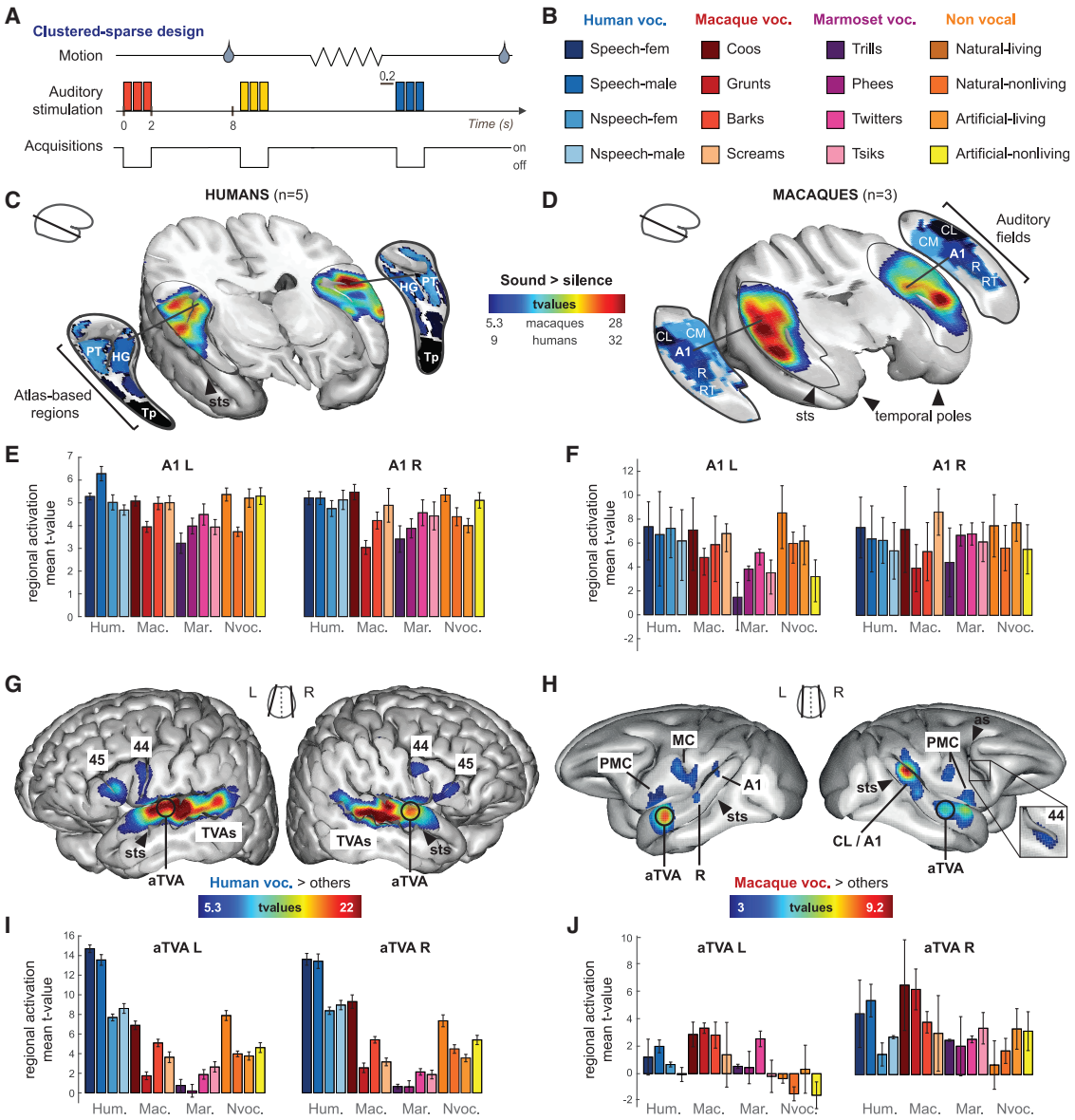


Figure 1. Auditory cerebral activation in humans and macaques

(A) Scanning protocol. Auditory stimuli were repeated three times in rapid succession during silent intervals between scans; macaques were rewarded with juice after 8-s periods of immobility.

(B) Auditory stimuli. Stimuli consisted of 96 complex sounds from 4 large categories divided into 16 subcategories.

(C and D) Areas with significant ($p < 0.05$, corrected) activation to sounds versus the silent baseline. t value threshold as indicated under the color bar. (C) Humans. PT, planum temporale; HG, Heschl's gyrus (from Harvard Oxford atlas); Tp, temporal pole; sts, superior temporal sulcus. (D) Macaques. A1, R, core auditory areas; CL, CM, RT, belt auditory areas from Petkov et al.[7]

(E and F) Group-averaged regional mean activation (t values) for the 16 sound subcategories compared to silence in (E) humans and (F) macaques. Error bars indicate SEM.

(G and H) CV-selective areas showing greater fMRI signal in response to CVs versus all other sounds. White circles indicate the location of bilateral anterior temporal voice areas in both species. (G) Human voc. > others at $p < 0.05$ corrected; TVAS, temporal voice areas; FVAs, frontal voice areas; 44-45, corresponding Brodmann areas from Harvard Oxford Atlas. (H) Macaque voc. > others at $p < 0.001$ uncorrected, $p < 0.05$ cluster-size corrected; as, arcuate sulcus; MC, motor cortex; PMC, premotor cortex; 44, corresponding Brodmann area from D99 atlas.

(I and J) Group-averaged regional mean activation (t values) for the 16 sound subcategories compared to silence in the aTVAs in (I) humans and (J) macaques. See also Figures S1–S3 and Table S1.

Anterior TVAs in humans and macaques

We next searched for conspecific vocalization (CV)-selective activations by contrasting in each species the fMRI signal measured in response to CVs versus all other sounds. In humans this comparison confirmed the classical pattern of three main clusters of voice selectivity along mid-superior temporal sulcus

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

CellPress
OPEN ACCESS

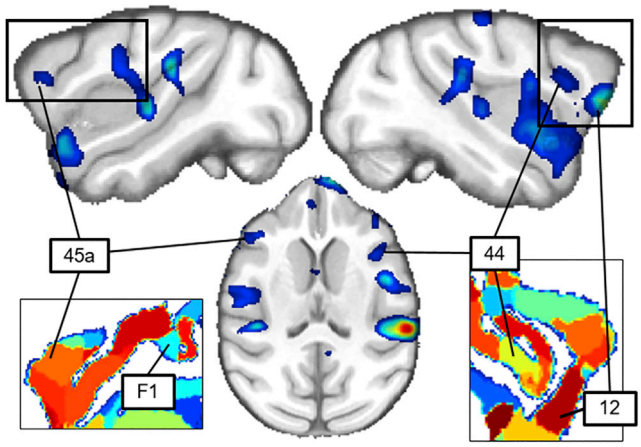


Figure 2. Prefrontal CV-selective activations in macaques

The statistical map of the contrast of CVs versus all other sounds in the three macaque subjects ($p < 0.05$, corrected) in shown in color scale overlaid on a T1-weighted image of the macaque brain in sagittal (top) and axial (bottom) slices. Black rectangles (top) zoom in activations in prefrontal cortex and show them relative to the anatomical parcellation of the D99 template (bottom).[12] Numbers indicate anatomical localization of the maxima of CV selectivity in prefrontal cortex.

(STS) to anterior STG bilaterally—the posterior, middle, and anterior temporal voice areas (TVAs)—with additional voice-selective activations in premotor and inferior frontal cortex including in bilateral BA 44 and 45 (Figures 1G and S2; Table S1).

In macaques, the contrast of macaque vocalizations versus all other sounds primarily yielded bilateral CV-selective activations ($p < 0.05$, corrected) in anterior STG and extending ventrally to the upper bank of STS in the left hemisphere (Figures 1F, S2, and S3; areas rSTG and RPB; Table S1). These regions correspond to the cytoarchitetonic area ts2, confirming the localization of the macaque voice area previously described in the right hemisphere[5] while emphasizing its bilateral nature: the left voice area was actually more consistently located in our macaques than its right counterpart (*cf.* coincidence maps of Figure S2). We call these bilateral voice areas the macaque aTVAs because of their anterior STG localization analogous to that of the human aTVAs.

The aTVAs in both species responded to most sound categories compared to the silent baseline but (by definition) most strongly to CVs (Figures 1I and 1J). The human aTVAs were most active bilaterally in response to speech sounds (Figure 1I), although their response to non-speech voice stimuli was also greater than to the other sounds. The human aTVAs also responded strongly to the macaque "coo" subcategory—vowel-like affiliative vocalizations that we can easily imitate. The macaque aTVAs responded most strongly to coos and grunts in both hemispheres but also responded robustly to human speech (Figure 1J).

A number of additional CV-selective responses were observed in more posterior regions of the STG, corresponding to core areas (A1 and R) bilaterally and to the caudal lateral field (CL) in the right hemisphere, similar to previously reported cases.[5,11] CV-selective activations were also found in bilateral premotor areas in the inferior prefrontal cortex including BA 44 and 45 in the right hemisphere (Figure 2; Table S1).

Similar representational geometries in human and macaque A1

In order to further probe the potential functional homology between the human and macaque aTVAs, we asked how these areas represent dissimilarities within the stimulus set[13] in comparison with A1. The representational similarity analysis (RSA) framework[14,15] allows quantitative comparisons between various measures of brain activity and theoretical models. In the 2 monkeys who were scanned using the event-related design as well as the 5 human participants, we built 16 × 16 representational dissimilarity matrices (RDMs; STAR Methods; Figure 3A) capturing at different cortical regions (left and right A1 and aTVAs in both species) the pattern of dissimilarities in fMRI responses (group-averaged Euclidean distance measures) to each pair of the 16 stimulus subcategories (Figure 1B).

We compared these RDMs to one another and to two types of comparison RDMs: acoustical RDMs and categorical model RDMs (Figure 3A). Three acoustical RDMs reflect the pattern of difference between the 16 sound subcategories along three measures examining complementary aspects of low-level acoustical structure: loudness, spectral center of gravity (SCG), and pitch (*cf.* STAR Methods). Three binary categorical RDMs capture the theoretical pattern of pairwise dissimilarities in our stimulus set under three separate models of ideal categorical distinction (Figure 3A): (1) a "human" model in which human voices are categorized separately from all other sounds, with no dissimilarity between cerebral responses to pairs of human voices or to pairs of the other sounds, but maximal dissimilarity between responses to a human voice versus another sound; (2) a "macaque" model categorizing macaque vocalizations apart from other sounds; and (3) a "nonvocal" model categorizing vocalizations of all species apart from non-vocal sounds.

In A1, RDMs were strongly correlated across the left and right hemispheres in both humans and macaques (human A1, left versus right, median bootstrapped Spearman's rho = 0.606, bootstrapped $p < 10^{-5}$, below Bonferroni-corrected threshold of $p = 0.05/8 = 0.0063$; macaque A1, left versus right, rho = 0.536, $p < 10^{-5}$). Remarkably, A1 RDMs were also strongly correlated across species in both hemispheres (left A1, human versus macaque, rho = 0.594, $p < 10^{-5}$; right A1, human versus macaque, rho = 0.664, $p < 10^{-5}$).

Comparisons of A1 RDMs with the acoustical RDMs yielded strong associations in both species for all three acoustical measures (Figure 3B; loudness: left human A1, Spearman's rho = 0.672, $p = 0.0015$, below Bonferroni-corrected threshold of $p < 0.05/24 = 0.0021$; right human A1, rho = 0.542, $p < 10^{-5}$; left macaque A1, rho = 0.222, $p = 0.0082$, n.s.; right macaque A1, rho = 0.315, $p = 2.75\ 10^{-4}$; SCG: left human A1, rho = 0.358, $p = 2.64 \times 10{-}5$; right human A1, rho = 0.612, $p < 10^{-5}$; left macaque A1, rho = 0.217, $p = 0.0086$, n.s.; right macaque A1, rho = 0.435, $p < 10^{-5}$; pitch: all rhos > 0.401, $p < 10^{-5}$).

In contrast, none of the four A1 RDMs showed significant associations with any of the three categorical model RDMs, as assessed by comparing via two-sample t tests the distributions of distance percentile values in the within versus the between portions of the A1 RDMs predicted by each model (Figure 3C; human model, all t values < 1.69, $p > 0.047$, above Bonferroni-corrected threshold of $p < 0.05/24 = 0.0021$, n.s.; macaque model, all $t < 0.989$, $p > 0.162$, n.s.; nonvocal model, all t negative, n.s.).

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

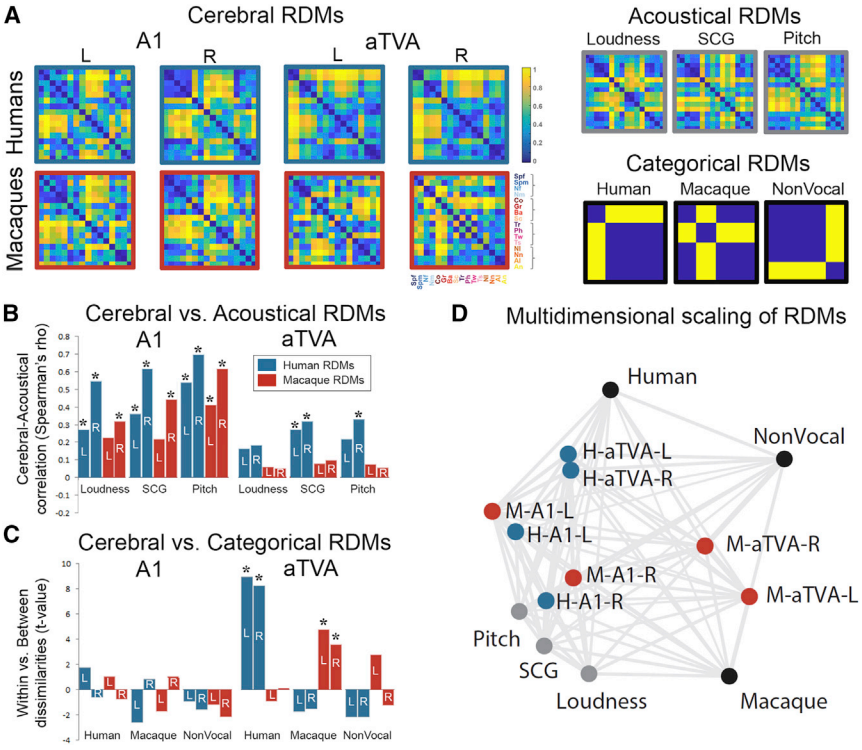


Figure 3. Representational similarity analysis in A1 and the aTVAs

(A) Representational dissimilarity matrices (RDMs) showing percentile dissimilarities in pairwise fMRI response to the 16 sound subcategories for left and right A1 and aTVAs in both species, along with 3 comparison acoustical RMS (right column, top row) and 3 categorical RDMs (bottom row).

(B) Comparison between brain RDMs and acoustical RDMs (Spearman correlation). *p < 0.05, Bonferroni-corrected.

(C) Comparison between brain RDMs and categorical model RDMs.

(D) 2D representation of dissimilarities within brain and comparison RDMs via multidimensional scaling. Large distances indicate large dissimilarities (low correlations). Blue disks, human RDMs; red disks, macaque RDMs; black disks, model RDMs; gray disks, acoustical RDMs; L, left hemisphere; R, right hemisphere.

Species-specific, but functionally homologous representational geometries in the aTVAs

At the level of the aTVAs, RDMs were also strongly correlated across hemispheres in both species (human aTVAs, left versus right, rho = 0.937, $p < 10^{-5}$; macaque aTVAs, left versus right, rho = 0.367, $p = 1.49 \times 10^{-5}$) and they correlated across species in the right, and with near-significance in the left, hemisphere (left aTVA, human versus macaque, rho = −0.0761, p = 0.789, n.s.; right aTVA, human versus macaque, rho = 0.258, p = 0.0024). Comparisons with the acoustical RDMs yielded smaller rho values than for A1 that only reached significance for the human aTVAs (Figure 3B; loudness: all rhos < 0.181, n.s.; SCG: left human aTVA, rho = 0.268, p = 0.0016; right human aTVA, rho = 0.315, $p = 2.12 \times 10^{-4}$; macaque aTVAs, all rhos < 0.092, n.s.; pitch: left human aTVA, rho = 0.215, p = 0.0093, n.s.; right human aTVA, rho = 0.325, $p = 1.3 \times 10^{-4}$; macaque aTVAs, all rhos < 0.073,n.s.).

Unlike in A1, however, there were significant associations between aTVA RDMs and the models in both species, but only with the species-specific model (Figure 2C; human model, left human aTVA, t = 8.914, $p < 10^{-5}$; right human aTVA, t = 8.179, $p < 10^{-5}$; left macaque aTVA, t = −0.887, p = 0.811, above Bonferroni-corrected threshold of p < 0.05/24 = 0.0021, n.s.; right macaque aTVA, $t = 6.98 \times 10^{-4}$, p = 0.5, n.s.; macaque model, left human aTVA, t = −1.713, p = 0.955, n.s.; right human aTVA, t = −1.489, p = 0.93, n.s.; left macaque aTVA, t = 4.761, $p < 10^{-5}$; right macaque aTVA, t = 3.538, $p = 5.779 \times 10^{-4}$; nonvocal model, all t values < 2.704, p > 0.0039, n.s.).

Thus, while associations with A1 RDMs were observed with acoustical RDMs, but not with categorical RDMs, a nearly opposite pattern was found with the aTVAs, with weak correlations with the acoustical RDMs and strongest associations with their own, species-specific categorical RDM. This pattern of result is well illustrated by the two-dimensional representation of the relative position of the RDMs via multidimensional scaling in Figure 3D: human and macaque A1 RDMs cluster together close to the acoustical RDMs and far from the categorical models, indicating similar representational geometries across both species and hemispheres that largely reflect low-level acoustical differences in the stimulus set. The aTVA RDMs, in contrast, are separated by species and displaced away from the acoustical RDMs toward their respective categorical RDM, indicating representational geometries more abstracted from acoustics that tend to categorize conspecific vocalizations apart from other sounds in the two species.

Note that only a small number of models were compared here, largely to mitigate the multiple comparisons problem, such that it is entirely possible that other models based on other acoustical features or combinations of features may better account for the patterns of activity observed in the aTVAs. Note also that the brain-acoustics correlations observed particularly in the human aTVAs could also be due to intrinsic correlations between the acoustical and categorical models—an issue to be pursued in future studies.

We did not observe particular hemispheric differences in either the localization of the macaque aTVAs or their response profile and representational geometries. This is consistent with lateralization analyses in several hundreds of human subjects that indicate a slight, non-significant right-hemispheric bias[4] in an otherwise largely symmetrical, bilateral pattern of voice sensitivity. The well-known hemispheric asymmetries in human auditory processing[16] likely arise at higher processing stages, more specialized for a specific type of information in voice (e.g., speaker versus phoneme identity). In any case, given the large inter-individual variability in hemispheric lateralization known in humans, strong claims on hemispheric lateralization can hardly be made based on samples of 2–3 individuals and will require larger samples in future studies.

The handful of neuroimaging studies of vocalization processing in macaques[5,11,17] have produced mixed results so far, in part because of large residual movements.[18] Only few studies directly compared humans and macaques in comparable conditions of auditory stimulation.[19] The first study to have done so highlighted a distributed pattern of CV sensitivity in the macaque akin to that observed in humans, including premotor and prefrontal regions, although results in macaques failed to reach standard significance thresholds.[20] A more recent comparative fMRI study observed important differences in the cerebral processing of harmonic sounds in the anterior temporal suggesting a "fundamental divergence" in the organization of higher-level auditory cortex.[18] Our findings suggest that such difference in cerebral processing of synthetic harmonic sounds does not extend to natural vocalizations, perhaps in part because harmonicity is not the only defining feature of macaque vocalizations.

Overall, these results reveal a much more similar functional organization of higher-level auditory cortex in macaques and humans than currently known. Rather than a repositioning,[6] these results instead suggest a complexification of the voice processing network in the human lineage—but one that preserved a key voice processing stage in anterior temporal cortex. These findings further validate the macaque as a valuable model of higher-level vocalization processing, opening the door to more detailed investigations with techniques such as fMRI-guided electrophysiology.[21] In the visual domain, the macaque model has generated considerable advances in our understanding of human face processing:[22,23] our findings set the stage for comparable efforts into cerebral voice processing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human participants
  - Macaque subjects and surgical procedures
- METHOD DETAILS
  - Auditory stimuli
  - Experimental Protocol
  - fMRI acquisition
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - fMRI data preprocessing
  - fMRI data analysis
  - Representational similarity analysis (RSA)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.08.043.

ACKNOWLEDGMENTS

We thank C. Amiez, S. Ben Hamed, T. Brochier, B. Cottereau, F. Chavanne, O. Joly, S. Love, G. Masson, C. Petkov, W. Vanduffel, and B. Wilson for useful discussions. This work was funded by Fondation pour la Recherche Medicale (AJE201214 to P.B. and FDT201805005141 to C.B.); Agence Nationale de la Recherche grants ANR-16-CE37-0011-01 (PRIMAVOICE), ANR-16-CONV-0002 (Institute for Language, Communication and the Brain), and ANR-11-LABX-0036 (Brain and Language Research Institute); the Excellence Initiative of Aix-Marseille University (A*MIDEX); and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 788240).

AUTHOR CONTRIBUTIONS

Conceptualization, C.B., R.T., and P.B.; Methodology, C.B., R.T., B.N., J.S., X.D., J.B., and P.B.; Data acquisition, C.B., R.T., J.S., E.R., and L.R.; Analyses, C.B., R.T., B.L.G., and P.B.; Funding, C.B. and P.B.; Writing, C.B., R.T., and P.B.

DECLARATION OF INTERESTS

The authors declare no competing interest.

Received: April 15, 2021
Revised: July 8, 2021
Accepted: August 13, 2021
Published: September 9, 2021

REFERENCES

1. Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. USA *97*, 11793–11799.
2. Rauschecker, J.P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. Science *268*, 111–114.
3. Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature *403*, 309–312.
4. Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., and Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage *119*, 164–174.
5. Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N.K. (2008). A voice region in the monkey brain. Nat. Neurosci. *11*, 367–374.
6. Ghazanfar, A.A. (2008). Language evolution: neural differences that make a difference. Nat. Neurosci. *11*, 382–384.
7. Petkov, C.I., Kayser, C., Augath, M., and Logothetis, N.K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS Biol. *4*, e215.
8. Penhune, V.B., Zatorre, R.J., MacDonald, J.D., and Evans, A.C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb. Cortex *6*, 661–672.
9. Kaas, J.H., Hackett, T.A., and Tramo, M.J. (1999). Auditory processing in primate cerebral cortex. Curr. Opin. Neurobiol. *9*, 164–170.
10. Baumann, S., Petkov, C.I., and Griffiths, T.D. (2013). A unified framework for the organization of the primate auditory cortex. Front. Syst. Neurosci. *7*, 11.
11. Gil-da-Costa, R., Martin, A., Lopes, M.A., Muñoz, M., Fritz, J.B., and Braun, A.R. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. Nat. Neurosci. *9*, 1064–1070.

12. Reveley, C., Gruslys, A., Ye, F.Q., Glen, D., Samaha, J., Russ, B.E., Saad, Z., Seth, A.K., Leopold, D.A., and Saleem, K.S. (2017). Three-dimensional digital template atlas of the macaque brain. Cereb. Cortex *27*, 4463–4477.
13. Edelman, S. (1998). Representation is representation of similarities. Behav. Brain Sci. *21*, 449–467, discussion 467–498.
14. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.
15. Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. PLoS Comput. Biol. *10*, e1003553.
16. Zatorre, R.J., Belin, P., and Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. *6*, 37–46.
17. Ortiz-Rios, M., Kuśmierek, P., DeWitt, I., Archakov, D., Azevedo, F.A., Sams, M., Jääskeläinen, I.P., Keliris, G.A., and Rauschecker, J.P. (2015). Functional MRI of the vocalization-processing network in the macaque brain. Front. Neurosci. *9*, 113.
18. Norman-Haignere, S.V., Kanwisher, N., McDermott, J.H., and Conway, B.R. (2019). Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. Nat. Neurosci. *22*, 1057–1060.
19. Erb, J., Armendariz, M., De Martino, F., Goebel, R., Vanduffel, W., and Formisano, E. (2019). Homology and specificity of natural sound-encoding in human and monkey auditory cortex. Cereb. Cortex *29*, 3636–3650.
20. Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., and Orban, G.A. (2012). Processing of vocalizations in humans and monkeys: a comparative fMRI study. Neuroimage *62*, 1376–1389.
21. Perrodin, C., Kayser, C., Logothetis, N.K., and Petkov, C.I. (2011). Voice cells in the primate temporal lobe. Curr. Biol. *21*, 1408–1415.
22. Freiwald, W.A., and Tsao, D.Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science *330*, 845–851.
23. Hesse, J.K., and Tsao, D.Y. (2020). The macaque face patch system: a turtle's underbelly for the brain. Nat. Rev. Neurosci. *21*, 695–716.
24. Ashburner, J. (2012). SPM: a history. Neuroimage *62*, 791–800.
25. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage *23* (*Suppl 1*), S208–S219.
26. Avants, B.B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTS). Insight J. *2*, 1–35.
27. Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., and Evans, A.C. (2002). A general statistical analysis for fMRI data. Neuroimage *15*, 1–15.
28. Moerel, M., De Martino, F., and Formisano, E. (2012). Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J. Neurosci. *32*, 14205–14216.
29. Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. Behav. Res. Methods *40*, 531–539.
30. Hauser, M.D. (1991). Sources of acoustic variation in rhesus macaque (Macaca mulatta) vocalizations. Ethology *89*, 29–46.
31. Ghazanfar, A.A., and Liao, D.A. (2018). Constraints and flexibility during vocal development: Insights from marmoset monkeys. Curr. Opin. Behav. Sci. *21*, 27–32.
32. Capilla, A., Belin, P., and Gross, J. (2013). The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cereb. Cortex *23*, 1388–1395.
33. Leite, F.P., Tsao, D., Vanduffel, W., Fize, D., Sasaki, Y., Wald, L.L., Dale, A.M., Kwong, K.K., Orban, G.A., Rosen, B.R., et al. (2002). Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. Neuroimage *16*, 283–294.
34. Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., and Collins, D.L.; Brain Development Cooperative Group (2011). Unbiased average age-appropriate atlases for pediatric studies. Neuroimage *54*, 313–327.
35. Baumann, S., Griffiths, T.D., Rees, A., Hunter, D., Sun, L., and Thiele, A. (2010). Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates. Neuroimage *50*, 1099–1108.
36. Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., and Evans, A.C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. *4*, 58–73.
37. Aglieri, V., Chaminade, T., Takerkart, S., and Belin, P. (2018). Functional connectivity within the voice perception network and its behavioural relevance. Neuroimage *183*, 356–365.
38. Glasberg, B.R., and Moore, B.C.J. (2002). A model of loudness applicable to time-varying sounds. J. Audio Eng. Soc. *50*, 331–342.
39. de Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am. *111*, 1917–1930.

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Rhesus Macaque (Macacca Mulatta)	Station de Primatologie, UAR 846, Centre National de la Recherche Scientifique, D56 Rousset sur arc 13790, France	N/A
Deposited data		
Human fMRI data	This paper	Zenodo: https://doi.org/10.5281/zenodo.5071389
Macaque fMRI data	This paper	Zenodo: https://doi.org/10.5281/zenodo.5074859
Software and algorithms		
MATLAB R2015b	MathWorks	http://www.mathworks.com/products/matlab/; RRID: SCR_001622
SPM12	[24]	http://www.fil.ion.ucl.ac.uk/spm/; RRID: SCR_007037
FSL v5.0.10	[25]	http://www.fmrib.ox.ac.uk/fsl/; RRID: SCR_002823
ANTS - Advanced Normalization ToolS	[26]	http://stnava.github.io/ANTs/; RRID: SCR_004757
FMRISTAT - A general statistical analysis for fMRI data	[27]	http://www.math.mcgill.ca/keith/fmristat/; RRID: SCR_001830
Code used in the present paper	This paper	Zenodo: https://doi.org/10.5281/zenodo.5075675

RESOURCE AVAILABILITY

Lead contact

Further information and requests for stimuli and data should be directed to and will be fulfilled by the lead contact, Pascal Belin (pascal.belin@univ-amu.fr).

Materials availability

This study did not generate any new materials or reagents.

Data and code availability

- The raw data have been deposited at Zenodo and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human participants

Five native French human speakers were scanned (one male (author R.T.) and four females; 23-38 years old). Participants gave written informed consent and were paid for their participation.

Macaque subjects and surgical procedures

Three adult rhesus monkeys (Macaca mulatta) were scanned, one 7-year-old male (M1) weighing 10 kg and two females (M2, M3) of 4 and 5 years of age and weighing between 4 and 5 kg. Each animal was implanted with a custom-made MRI-compatible head-post under sterile surgical conditions. The animals recovered for several weeks before being acclimated to head restraint via positive reinforcement (juice rewards). All experimental procedures were in compliance with the National Institutes of Health's Guide for the Care and Use of Laboratory Animals and approved by the Ethical board of Institut de Neurosciences de la Timone (ref 2016060618508941).

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

METHOD DETAILS

Auditory stimuli

Four main categories of sounds were used in the experiment: human voices, macaque vocalizations, marmoset vocalizations and non-vocal sounds, each containing 24 stimuli, for a total of 96 sound stimuli. Each main category was divided into 4 subcategories of 6 stimuli, forming 16 subcategories in total (*cf.* Table S1). The set of stimuli used during training was different from the one used during scanning in order to minimize familiarization effects. Human voices contained both speech (sentence segments from the set of stimuli used in a previous study,[28] n = 12), and non-speech (vocal affect bursts selected from the Montreal Affective Voices dataset;[29] n = 12), equally distributed into positive (pleasure, laugh; n = 4), neutral (n = 4) and negative (angry, fear; n = 4) vocalizations. Macaque vocalizations, kindly provided by Marc Hauser,[30] included both positive (coos 25%, n = 6, grunts 25%, n = 6) and negative (aggressive calls 25%, n = 6, screams 25%, n = 6) calls. Marmoset vocalizations, kindly provided by Asif Ghazanfar,[31] were divided into supposed positive (trill 25%, n = 6), neutral (phee 25%, n = 6, twitter 25%, n = 6) and negative (tsik 25%, n = 6) calls. These three primate call categories contained an equal number of female and male callers. Non-vocal sounds included both natural (living 25%, n = 6, non-living 25%, n = 6) and artificial sounds (human actions 25%, n = 6, or not 25%, n = 6) from previous studies from our group[3,32] or kindly provided by Christopher Petkov[5] and Elia Formisano.[28] Stimuli were adjusted in duration, resampled at 48828 Hz and normalized by root mean square amplitude. Finally, a 10-ms cosine ramp was applied to the onset and offset of all stimuli. During experiments, stimuli were delivered via MRI-compatible earphones (S14, SensiMetrics, USA) at a sound pressure level of approximately 85 dB (A).

Experimental Protocol

Two different protocols were used to train and scan the monkeys. Monkey M1 was involved in protocol 1, monkeys M2 and M3 were trained and scanned a year later, using protocol 2. Human data were acquired using the fMRI design of protocol 2 Table S2.

Functional scanning was done using a block-design paradigm with continuous acquisitions in protocol 1 and using an event-related paradigm with clustered-sparse acquisitions in protocol 2. The marmoset sound category was not included in protocol 1, whereas all 96 stimuli described above were presented in pseudo-random order in protocol 2. M1 underwent scanning sessions both without and with ferrous oxide contrast agent (monocrystalline iron oxide nanoparticle, MION). MION was used for all sessions of M2 and M3. No contrast agent was used for human participants.

Both protocols used an auditory listening task for which subjects were instructed (humans) or trained (monkeys) to stay still in the scanner for sessions of about one h and a half. Monkeys received juice rewards after remaining motionless for a fixed period of time (4 s for protocol 1, 8 s for protocol 2). Head and body movements of the monkeys were monitored online by analyzing the frame-by-frame differences in the images provided by a camera placed in front of the animal. Movements were not monitored online for humans. To minimize body motion, monkeys tested with protocol 2 were also required to hold a bar with both hands. Hand detection was achieved using two optical sensors. To increase engagement in the task, protocol 2 also included a visual feedback that indicated the presence of each hand on the bar, reward delivery, as well as a gauge of the time remaining until reward delivery.

In protocol 1, blocks of the same category of sound stimuli were presented for duration of 6 s during non-MION sessions and 30 s during MION sessions. Juice reward was delivered at the end of each motionless period of 4 s, independently of sound stimulation. Protocol 2 was dependent on monkey behavior: a trial started when the monkey had been holding the bar and staying motionless for 200ms. Then, to avoid interferences between sound stimulation and scanner noise, the scanner stopped acquisitions such that three repetitions of a 500ms stimulus (inter-stimulus interval of 250ms) were played on a silent background. Then scanning resumed and the monkey had to stay still for another 6 s period in order to receive a reward. Trials were interrupted as soon as motion was detected or a hand was released from the bar.

fMRI acquisition

Human and monkey participants were scanned using the same 3-Tesla scanner (Siemens Prisma). Human participants were scanned using a whole-head 64-channels receive coil (Siemens) in a single session including one T1-weighted anatomical scan (TR = 2.3 s, TE = 2.9ms, flip angle: 9°, matrix size = 192 × 256 × 256; resolution 1 × 1 x 1 mm3) and two functional runs (multiband acceleration factor: 4, TR = 0.945 s, TE = 30ms, flip angle = 65°, matrix size = 210 × 210 × 140, resolution of 2.5x2.5x2.5 mm3). In monkey M1 a T1-weighted anatomical image was acquired under general anesthesia (MPRAGE sequence, TE = 3.15ms, TR = 3.3 s, flip angle = 8°, matrix size: 192 × 192 × 144, resolution 0.4 × 0.4 × 0.4 mm3). During functional sessions, blood oxygen level-dependent (BOLD) EPI volumes were acquired using a single receive loop coil (diameter 11 cm) positioned around the head-post (BOLD sessions: TR = 859ms, TE = 30ms, flip angle = 56°, matrix size = 76 × 76 × 16, resolution 3 × 3 × 3 mm3; MION sessions: TR = 1437ms, TE = 20.6ms, flip angle = 70°, matrix size = 112 × 112 × 24, resolution 2 × 2 × 2 mm3). We acquired a total of 10 BOLD sessions for M1 (60 runs of 456 volumes each), plus 6 additional sessions (24 runs of 475 volumes each) using the MION contrast agent[33] for comparison with the BOLD session. For monkeys M2 and M3 a high-resolution T1-weighted anatomical volume was acquired under general anesthesia (MP2RAGE sequence, TE = 3.2ms, TR = 5 s, flip angle = 4°, matrix size = 176 × 160 × 160, resolution 0.4 × 0.4 × 0.4 mm3). MION functional volumes were acquired with an 8-channels surface coil (KU, Leuven) using EPI sequences (multiband acceleration factor: 2, TR = 0.955 s, TE = 19ms, flip angle = 65°, matrix size = 108 × 108 × 48, resolution 1.5 × 1.5 × 1.5mm3). We acquired a total of 19 sessions in M2 (79 runs of 96 stimulus presentations each) and 21 sessions in

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

M3 (72 runs of 96 stimulus presentations each). Before and after each MION run, data were collected to allow the calculation of T_2* maps by acquiring 18 volumes at the 3 gradient echo times of 19.8, 61.2 and 102.5ms.

QUANTIFICATION AND STATISTICAL ANALYSIS

fMRI data preprocessing

Preprocessing of the functional data included motion correction, spatial distortion reduction using field maps, inter-runs registration and spatial smoothing. Motion parameters were first computed to identify steady and moving periods for each run. Every functional volume was then realigned and unwarped to a reference volume taken from a steady period in the session that was spatially the closest to the average of all sessions. Spatial smoothing was done with a full-width half-maximum 3-dimensional Gaussian kernel that was twice the size of the functional voxels (i.e., 6 mm for M1 BOLD, 4 mm for M1 MION, 3mm for M2 and M3, 5mm for humans). Tissue segmentation and brain extraction was performed on the structural scans using the default segmentation procedure of SPM for human data and a custom-made segmentation pipeline for monkey data. This pipeline included the following steps: volume cropping (FSL); bias field correction (ANTS N4); denoising (spatially adaptive nonlocal means, SPM); first brain extraction (FSL); registration of the INIA19 macaque template brain (https://www.nitrc.org/projects/inia19) to the anatomical scan, in order to provide priors to SPM's old_segment algorithm; second brain extraction from the segmented tissues. Transformation matrices between anatomical and functional data were computed using boundary-based registration (FSL) for BOLD data (M1 & humans), and using non-linear registration (ANTS, SyN) for MION data in M2 & M3. These transformation matrices were used to register the tissue segmentation to the functional data and to register the functional results to the high-resolution anatomical scan. Individual human data were registered to the MNI152 ICBM 2009c Nonlinear Asymmetric template.[34] T_2* and R_2* maps were computed from the multi-echo data to assess the blood iron concentration.[33] The mean relaxation rate during a MION run was estimated from the mean R_2* across all brain voxels obtained from the multi-echo acquisitions before and after the run. MION runs with an estimated relaxation rate below 30 s^{-1} were excluded from the analysis. After this step, the analysis included all 24 MION runs of M1, 67 of the 79 MION runs of M2 and 64 of the 72 MION runs of M3.

fMRI data analysis

General linear model estimates of responses to all sounds versus silence (all > silence) and to conspecific vocalizations versus all other sound categories (CV > non CV) were computed using fMRISTAT.[27] The general model included several covariates of no interest: the first 7 and 4 principal components of a principal component analysis performed on an eroded mask of the functional voxels identified as containing white matter and cerebrospinal fluid respectively; one vector for each functional volume belonging to a "moving period," as identified during the first preprocessing step (these vectors contained zeros for every time step except for the moving volume). For BOLD data (M1 & humans), a hemodynamic response function (HRF) with a peak at 4 s and an undershoot at 10 s was used during analysis.[35] For MION data a MION-based response, manually designed to have a reversed sign, a long tail and no undershoot, similar to previous descriptions,[33] was used. Voxel significance was assessed by thresholding T-maps at $p < 0.05$, corrected for multiple comparisons using Gaussian Random Field Theory.[36] The quality of monkey functional data depends on numerous factors that cannot be assessed quantitatively (e.g., coils and insert earphones placement, monkey engagement). In these conditions, the global fMRI response to sound can be used to assess the quality of a run[5] and reject poor quality runs. To assess the contribution of each run to the global sound response, we computed the spatial extent of the significant voxels, as well as the maximum t-value, elicited in the all > silence contrast using a jackknife procedure that systematically leaved out each run from the entire dataset. Only the runs showing a positive sound response contribution were kept. We kept at this stage 33 runs out of 51 for M1, 26 out of 48 for M2 and 25 out of 42 for M3.

Representational similarity analysis (RSA)

We investigated cortical representations with RSA in two regions of interest (ROI): primary auditory cortex (A1) and anterior voice area (aTVA) in each species and hemisphere. In each subject and hemisphere, the center of the A1 ROI was defined as the maximum value of the probabilistic map (non-linearly registered to each subject functional space) of Heschl's gyri provided with the MNI152 template[8] for human subjects, and of Macaque A1s as identified in an earlier study[7] for monkeys. In each human subject and hemisphere, the center of the aTVA region corresponded to the local maximum of the CV > non CV t-map whose coordinates were the closest to the aTVAs reported in an earlier study.[37] M1 was not included in the RSA analysis as it had not been scanned with the marmoset vocalizations. In M2 and M3, aTVAs were identified bilaterally as the local maximum of the individual CV > non-CV t-map that was in the most anterior portion of the STG. Once a center of a ROI was defined, the 19 voxels in the functional space that were the closest to the center of the ROI and above 50% in the probabilistic maps or above significance threshold in CV > non CV t-maps, constituted the ROI (in most cases the ROI was a sphere). Note that the voxel number of the ROIs was the same for each species. ROI volume was 297 mm^3 (diameter 7.5 mm) in humans and 64 mm^3 (diameter 4.5 mm) in monkeys.

Brain RDMs were generated for each ROI (A1 and aTVA regions of both species) by computing the Euclidean distance between stimulus subcategories in multi-voxel activity space. These 16 × 16 brain RDMs were averaged across subjects to obtain one mean brain RDM per ROI and per species (8 brain RDMs in total). Acoustical RDMs were generated for each of the three measures by computing for each pair of stimulus subcategory the difference of the measure averaged across stimuli of each subcategory. Loudness and Spectral center of gravity (SCG, an acoustical correlate of timbre brightness) were estimated by modeling each sound using

Please cite this article in press as: Bodin et al., Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.043

the time-varying loudness model by Glasberg and Moore.[38] Pitch was estimated by modeling each sound using the YIN pitch extraction model by De Cheveigné and Kahawara.[39]

Planned comparisons between pairs of brain RDMs as well as between brain RDMs and Acoustical RDMs were performed using bootstrapped Spearman's rho correlation value (100,000 iterations, one-tailed) with Bonferroni correction for multiple comparisons (eight comparisons performed), resulting in a corrected p value threshold of $p = 0.05 / 8$. Planned comparisons between brain RDMs and the 3 Categorical RDMs were performed by comparing the within versus between portions of the brain RDMs predicted by each model using 2-sample t tests (100,000 iterations, one-tailed) with Bonferroni correction for multiple comparisons (twelve comparisons performed), resulting in a corrected p value threshold of $p = 0.05 / 12$. Visual representation of the pattern of correlations between RDMs in Figure 3D was performed via multidimensional scaling using the RSA toolbox.[15]