SCIENTIFIC REPORTS

OPEN

Processing complexity increases in superficial layers of human primary auditory cortex

Michelle Moerel[1,2,3,4], Federico De Martino[2,3,4], Kâmil Uğurbil[4], Essa Yacoub[4] & Elia Formisano [ID] [1,2,3]

The layers of the neocortex each have a unique anatomical connectivity and functional role. Their exploration in the human brain, however, has been severely restricted by the limited spatial resolution of non-invasive measurement techniques. Here, we exploit the sensitivity and specificity of ultra-high field fMRI at 7 Tesla to investigate responses to natural sounds at deep, middle, and superficial cortical depths of the human auditory cortex. Specifically, we compare the performance of computational models that represent different hypotheses on sound processing inside and outside the primary auditory cortex (PAC). We observe that while BOLD responses in deep and middle PAC layers are equally well represented by a simple frequency model and a more complex spectrotemporal modulation model, responses in superficial PAC are better represented by the more complex model. This indicates an increase in processing complexity in superficial PAC, which remains present throughout cortical depths in the non-primary auditory cortex. These results suggest that a relevant transformation in sound processing takes place between the thalamo-recipient middle PAC layers and superficial PAC. This transformation may be a first computational step towards sound abstraction and perception, serving to form an increasingly more complex representation of the physical input.

The human neocortex consists of six layers, each with a distinct anatomical connectivity and functionality. Across the primary sensory cortices (i.e., primary visual, auditory, and somatosensory cortex) the anatomical build up and connectivity of these cortical layers is relatively similar. Input from the thalamus to the primary sensory cortex arrives in layer IV, cortico-cortical connections arise from superficial layers II-III, and efferent subcortical projections originate from deep layers V–VI[1,2]. As a result of this layer-specific connectivity, the deep, middle, and superficial layers support different functions. While sensory processing in the middle input layers may be relatively simple, inherited from thalamic processing, superficial and deep layers may comprise increasingly complex stimulus representations as a result of columnar processing, horizontal connections, and cortico-cortical feedback[3,4].

In the primary visual and somatosensory cortex, laminar processing differences have been repeatedly shown. For both sensory modalities, processing is simplest and most representative of the physical sensory input in the thalamo-recipient middle layers, and increases in complexity towards superficial and deep layers[5–7]. For the auditory cortex, consistent principles of laminar processing have been more difficult to demonstrate. While early studies reported conflicting findings[3], recent studies provide evidence for an auditory laminar organization that is similar to the other early sensory cortices. In the PAC, the fastest responses to sounds are seen in layer IV[8], followed by deep and superficial layers (but see also reports of faster sound responses in deep layers, possibly due to direct thalamic connections[9,10]). Results from cat PAC showed systematic processing variations throughout the cortical depth, where processing complexity increased with distance from granular layer IV[10]. Accordingly, recordings in the awake marmoset monkey showed that while neurons in middle thalamo-recipient layers respond well to simple sounds, many neurons in superficial PAC could not be that easily driven[11]. Instead, superficial PAC neurons displayed sustained responses to complex sounds and con-specific vocalizations[12]. Thus,

[1]Maastricht Centre for Systems Biology, Maastricht University, Universiteitssingel 60, 6229 ER, Maastricht, The Netherlands. [2]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV, Maastricht, The Netherlands. [3]Maastricht Brain Imaging Center (MBIC), Oxfordlaan 55, 6229 EV, Maastricht, The Netherlands. [4]Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, 2021 6th Street SE, Minneapolis, MN, 55455, USA. Correspondence and requests for materials should be addressed to M.M. (email: michelle.moerel@maastrichtuniversity.nl)

1
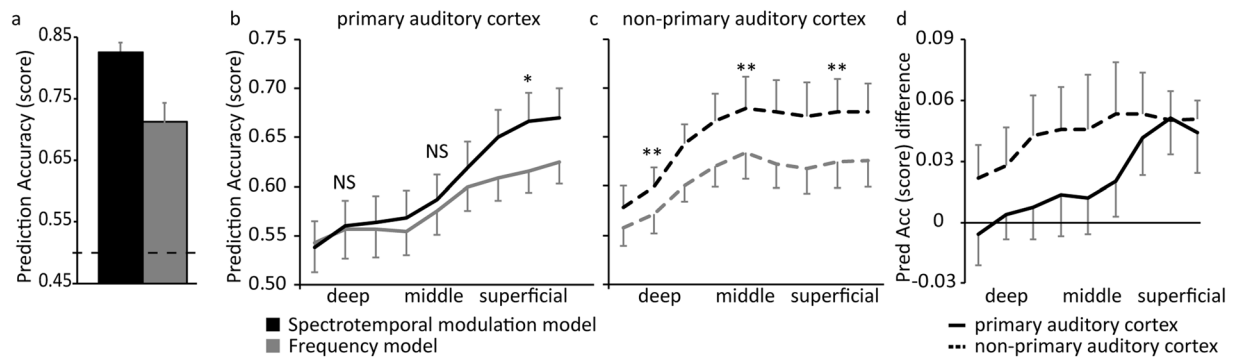
Figure 1. Model performance. Model performance is evaluated by the prediction accuracy of responses to test sounds (i.e., sound identification score). (**a**) Performance of the models on the entire dataset, which covers the majority of the supratemporal plane (STP) and parts of the superior temporal gyrus (STG). The dashed line corresponds to chance performance (score = 0.5), and the error bars indicate the standard error across subjects ($N=6$). (**b,c**) The cortical depth dependent model performance ranging across deep, middle, and superficial grey matter for the part of the grid with the highest (i.e., primary auditory cortex) and lowest (i.e., non-primary auditory cortex) 50% of myelin-related contrast in solid and dashed lines, respectively. Statistical results are indicated by NS (not significant), $*p < 0.05$, and $**p < 0.001$. (**d**) The cortical depth dependent difference in model performance (ranging across deep, middle, and superficial grey matter) between primary auditory cortex and non-primary auditory cortex. In b-d, the error bars indicate the standard error across hemispheres ($N=12$).

while layer IV may contain a relatively faithful representation of the spectrogram of incoming sounds, other cortical depths may incorporate more complex sound representations.

To date, the spatial resolution of non-invasive measurements has prohibited investigating if similar differences in laminar processing are present in the human auditory cortex. This is a crucial omission, as human audition is unique in its flexibility and richness, e.g., for processing speech and language. Therefore, results from animal studies are not directly transferable to the human. Moreover, while the PAC was traditionally considered to be a simple feature analyser, more recent studies highlight its role in processes beyond the physical sound analysis. For example, processing in the PAC is rapidly modulated with changing task demands[13,14], and is affected by signals from other sensory cortices[15]. Furthermore, the PAC has been implicated in auditory object formation[16], and auditory category learning[17,18]. These complex auditory processes, going beyond the physical sound analysis, may occur within PAC as information flows through the layers of the canonical microcircuit. Exploring layer-dependent sound processing may shed light on how PAC computations support auditory cognition.

Ultra-high field magnetic resonance imaging (MRI) profits from an increased sensitivity and specificity to neuronal activity, and allows investigating the human brain at a sub-millimetre spatial resolution[19]. Given that the thickness of the human PAC is ~2–3 mm[20], this sub-millimetre spatial resolution enables acquiring independent signals from deep, middle (presumably reflecting thalamo-recipient layer IV), and superficial cortical depths. Here we investigate responses to natural sounds across cortical depths of the human auditory cortex. We observe that while BOLD responses in deep and middle PAC layers are equally well represented by a simple *frequency* model and a more complex frequency-specific *spectrotemporal modulation* model, responses in superficial PAC are significantly better represented by the more complex model. This suggests that the neuronal populations underlying the responses to natural sounds in superficial PAC display an increase in processing complexity compared to deep and middle PAC layers. This increased processing complexity is present throughout cortical depths in the non-primary auditory cortex as well. Thus, a transformation in sound processing takes place between middle and superficial PAC. This transformation may be a key step from a representation of the physical sensory input toward an increasingly complex, categorical, and abstract sound representation.

Results

We obtained high resolution anatomical data (0.6 mm isotropic) and functional data (0.8 mm isotropic) while subjects listened to natural sounds. We observed significant responses to the sounds throughout the supratemporal plane, including HG (the putative location of primary auditory cortex), Heschl's sulcus, planum polare (PP), planum temporale (PT), and the parts of the superior temporal gyrus and sulcus that were covered by the fMRI field of view.

Cortical responses to the sounds were analyzed with two encoding models, representing different hypotheses on sound processing. A first *frequency* model was simplest, describing sound processing in terms of the frequency preference of cortical neuronal populations. A second *spectrotemporal modulation* model described cortical sound processing as the frequency-specific tuning of neuronal populations to combined spectral and temporal modulations. The performance of the models based on the responses in the auditory cortex as a whole (including all voxels with a significant response to the sounds; $p < 0.05$ uncorrected) is shown in Fig. 1a. Both the simpler *frequency* model and the *spectrotemporal modulation* model could predict responses to testing sounds above chance (average prediction accuracy [SEM] = 0.83 [0.02] for the *spectrotemporal modulation* model, and 0.71 [0.03] for the frequency model), and thus provided a meaningful representation of sound processing in the auditory cortex. In agreement with our previous study[21], the *spectrotemporal modulation* model outperformed
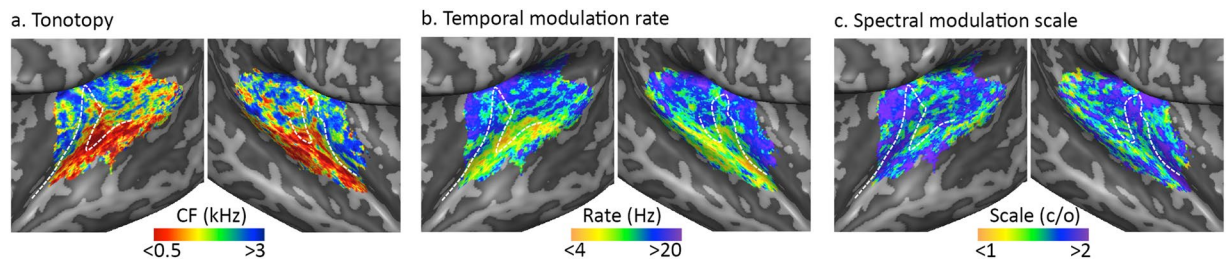
Figure 2. Group topographic maps. (**a**–**c**) CBA-based group maps for preferred frequency (tonotopy), temporal modulation rate, and spectral modulation scale, on an inflated representation of the left and right hemisphere (in the left and right column, respectively). The white dashed line outlines HG.

the *frequency* model (two-sided paired t-test after Fisher transformation of the prediction accuracy values; t(5) = 7.19; $p = 8.09 \times 10^{-4}$). This finding was not driven by a subset of sounds, but instead was observed across the large majority of sound categories (Fig. S1a).

We further examined the fitted *spectrotemporal modulation* model by creating large-scale topographic maps for each of the three acoustic features. Maps of tonotopy (i.e., best frequency [BF]), preferred temporal modulation rate, and preferred spectral modulation scale are shown in Fig. 2. The tonotopy maps were in accordance with previous reports[22–25]. That is, the tonotopy maps showed a cortical region responding best to low frequencies situated on the posterior part of HG and anterior part of Heschl's sulcus, surrounded anteriorly and posteriorly by regions preferring higher frequencies (Fig. 2a). The temporal modulation rate and spectral modulation scale maps displayed both similarities and differences compared to previous findings. We observed a significantly negative correlation between the temporal modulation rate and spectral modulation scale map (mean [SE] = −0.027 [0.002]; $p = 0.031$; two-tailed non-parametric signed rank test on Fisher transformed correlation values between individual maps prior to spatial smoothing) in agreement with previous results[21,26], yet this correlation was weaker than reported before[21]. Consistent with our previous results[21], the medial and lateral part of the STP preferred higher and lower temporal modulation rates, respectively (Fig. 2b). More specifically, the cortical region at the posterior adjacency of the medial HG preferred the highest modulation rates, while regions along the STG preferred the lowest temporal modulation rates. In the maps of spectral modulation scale preference, observations consistent with our previous results include tuning to higher spectral modulation scales along the anterior part of HG and on the PP, and a preference for lower spectral modulation scales in the majority of the PT and STG (Fig. 2c).

Next, we explored the performance of the sound representation models at a smaller spatial scale. That is, as the primary auditory cortex (PAC) is more densely myelinated than surrounding auditory regions[27,28], we used myelin-related contrast (MRC) maps to obtain a non-invasive estimate of the PAC in each hemisphere (Fig. 3). Throughout hemispheres we localized PAC to the postero-medial part of HG (Figs 3 and S2). The region identified as PAC moved posteriorly when incomplete or complete duplications of HG were present (e.g., the left hemisphere of S1 in Fig. S2). Our definition of PAC covered approximately half of the HG.

Finally, we analysed model performance in the PAC and the non-PAC, at deep, middle, and superficial cortical depth (Fig. 1b–d). The GE-EPI signal is strongest at the cortical surface, and this has been shown to correlate with an overall prediction accuracy increase towards superficial cortical depths[29]. The improved performance of both models towards the cortical surface, as can be observed in Fig. 1b,c, should therefore not be interpreted as a neuroscientific result. On the other hand, we previously confirmed experimentally that the cortical depth dependent *difference* in model performance removes the effect of the signal strength (and thus the bias towards the surface)[29], making the depth dependent effects interpretable (Fig. 1d). In the PAC only, we observed a significant interaction such that the *difference* in model prediction accuracy varied with cortical depth (two-way RM ANOVA after Fisher transformation of the prediction accuracy values; significant interaction between the factors 'Computational Model' and 'Cortical Depth'; F[2,22] = 5.31; $p = 1.31 \times 10^{-3}$). Only at a superficial depth of the PAC, did the *spectrotemporal modulation* model significantly outperform the *frequency* model (one-sided paired t-test; t(11) = 3.01; p(corrected) = 1.79×10^{-2}). In the non-PAC, there was a significant difference in model performance (main effect of 'Computational Model'; F[1,11] = 7.12; $p = 2.18 \times 10^{-2}$) which did not vary with cortical depth (no significant interaction between the factors 'Computational Model' and 'Cortical Depth'; main effect of 'Cortical Depth; F[2,22] = 5.77; $p = 9.60 \times 10^{-3}$). In other words, the *spectrotemporal modulation* model significantly outperformed the frequency model in these non-primary auditory regions uniformly throughout cortical depth. The improved performance of the *spectrotemporal modulation* model in the non-PAC compared to the PAC was not confined to a subset of sounds, but was observed for all sound categories (Fig. S1b).

Discussion

To date, it remains unclear if the laminar anatomical connectivity pattern of the human PAC is paralleled by layer-specific variations in sound processing. Using ultra-high field MRI, we observed that BOLD responses in middle PAC layers were equally well represented by a simple *frequency* model and the *spectrotemporal modulation* model. Instead, BOLD responses in superficial PAC were significantly better represented by the *spectrotemporal modulation* model, suggesting that neuronal populations in superficial PAC underlying the BOLD responses displayed an increased processing complexity. This increased complexity of sound processing was propagated throughout cortical depths in non-primary auditory cortex, and in accordance with our previous report at lower
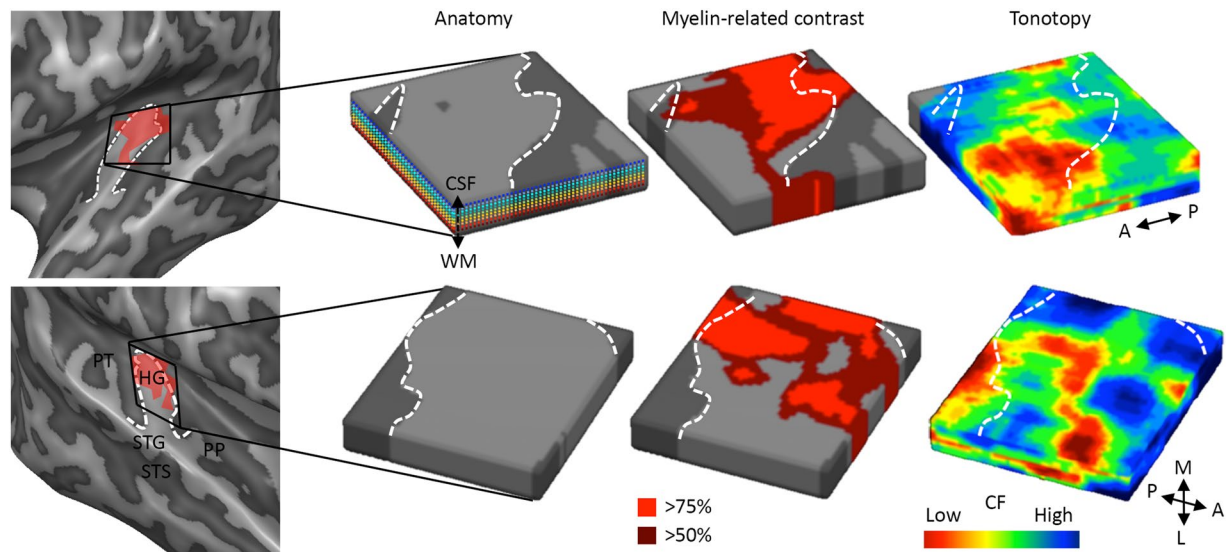
Figure 3. Analysis pipeline. The left and right temporal lobe (top and bottom row, respectively) of a representative subject (S2; see Supplementary Fig. 2 for the data of all individuals). Light and dark grey regions represent gyri and sulci respectively. The white dashed line outlines Heschl's gyrus (HG). Major anatomical landmarks are shown in (**b**; first column), and include HG, planum polare (PP), planum temporale (PP), the superior temporal gyrus (STG) and the superior temporal sulcus (STS). The red shaded region in the first column approximates the region with highest myelin-related contrast (MRC; >50%). For each subject and hemisphere a cortical depth-dependent grid, situated on the postero-medial part of HG, is defined. The cortical ribbon is sampled at nine cortical depths, range from approaching the white matter (WM) to approaching the cerebrospinal fluid (CSF; coloured lines in second column of **a**). The cortical depth dependent grids are used to sample the MRC maps, which are thresholded to define the primary auditory cortex (PAC; MRC > 50%, bright and dark red colours in the third column), and non-PAC (MRC < 50%; light and dark grey colours in the third column). The cortical depth dependent grids are furthermore used to sample the functional data, allowing for cortical depth dependent fitting of the computational models and resulting in cortical depth dependent maps of feature preference (e.g., tonotopy; fourth column).

field strength and spatial resolution[21], was preserved throughout the supratemporal plane. The observed increase in processing complexity in superficial PAC and throughout non-PAC may reflect neuronal sensitivity to acoustic feature combinations, such as previously described in the auditory cortices of echo-locating bats[30], songbirds[31], and the marmoset monkey[12]. Tuning to acoustic feature combinations may serve as a stepping stone towards the creation of an auditory object, and provide a mechanism by which those sounds that are ecologically important can be more efficiently (i.e., sparsely) represented[12]. Overall, the modified sound representation in superficial PAC may be a first computational step in the transformation of the physical input towards sound abstraction and perception[32].

Sound processing complexity in superficial PAC may increase through processes operating at a local spatial scale, resulting from the flow of information throughout the canonical microcircuit[3]. Previous electrophysiological studies showed that processing of spectral and temporal modulation information undergoes a radical change from thalamus to the granular PAC input layers[33] and, within PAC, with increased distance from middle input layers[10] or in superficial layers[12]. Our results show an increase in processing complexity when moving from middle to superficial cortical depths. However, we did not observe an increased processing complexity in infragranular layers, as may have been expected based on cat recordings[10] and laminar processing principles in other early sensory cortices[5,7]. This discrepancy could represent a species-specific difference. While processing complexity increases in infragranular cat PAC, this may not be present in the marmoset monkey[12] or in the human. An exclusive change in supragranular PAC processing complexity may result through influences originating within the auditory cortex, through lateral connections of pyramidal cells in superficial PAC layers[6] or as a result of short-distance feedback projections from secondary auditory cortical areas[34]. Alternatively, the reason for not observing increased processing complexity in infragranular PAC could be methodological. BOLD fMRI is an indirect measure of neuronal activity, which represents neuronal input (i.e., LFP) rather than spiking activity[35]. Input to middle PAC is thalamic, and also input to deep PAC layers may be partially thalamic[9]. The thalamic input to deep and middle layers could have driven the BOLD responses that we obtained from these cortical depths, while the BOLD responses collected from superficial cortical depths may have reflected cortical auditory processing[33].

Two methodological issues related to the execution of the analysis and the interpretation of our findings require further discussion. First, the non-invasive identification of the PAC is still a matter of debate, largely resulting from the fact that tonotopic maps alone are insufficient to delineate the PAC[36]. Non-invasive functional measures beyond tonotopy have been proposed to identify the PAC[37–39], but none of these have been consistently adopted. The PAC is more densely myelinated than surrounding auditory regions[27,28] and, consequently, maps of

4

myelin related contrast (MRC) were shown to be valuable for its localization[40–42]. Myelination and, correspondingly, the MRI-based MRC, varies gradually throughout the STP. We considered the half of the grid with the highest MRC as the PAC. While this threshold is necessarily arbitrary, it resulted in plausible PAC definitions. That is, the localization of the PAC to the postero-medial part of HG, as well as the posterior movement of PAC with (in)complete HG duplications, is consistent with results from cytoarchitecture[27,43–45]. While the size of our PAC, covering approximately half of the entire HG, is a smaller cortical area than the definition in a number of previous fMRI studies[23,46], this size concurs with cytoarchitectonic PAC definitions[47]. Overall, while it is possible that our definition of PAC did in fact not cover the complete PAC, it is highly likely that the region we refer to as PAC is indeed fully situated within the primary auditory cortex. The remainder of the grid, which we refer to as non-PAC, covers regions anterior and posterior to HG. In most hemispheres, it included the anterior part of HG and the first transverse sulcus, and posteriorly extended to Heschl's sulcus, the lateral part of a second HG when applicable, and the anterior part of the PT. These regions can be considered secondary auditory cortex (medial and lateral belt respectively), and may reflect part of regions 'MBelt' and 'LBelt' as recently described by Glasser *et al.*[48].

The second methodological issue that requires additional discussion concerns the spatial specificity of our fMRI measurements. Using GE-EPI, the acquired fMRI signal contains contributions from large pial veins situated on top of the grey matter[49], diving veins penetrating the grey matter, and the microvasculature[50]. As a result of the pial vein contribution, the observed signal changes are strongest at the cortical surface (high sensitivity) and spatially less specific than those at deeper cortical depths (low specificity)[51–53]. A first implication of these cortical depth-dependent changes in sensitivity and specificity is that GE-EPI cortical depth-dependent topographic maps may be biased in the assignment of the voxels' preferred feature[29,51,52]. We therefore do not report cortical depth-dependent maps of tonotopy and spectrotemporal modulation preference. These cortical depth-dependent maps were previously reported based on a more spatially specific, predominantly T_2-weighted, dataset[54]. A second implication is that an alternative explanation for the difference in model performance between PAC and non-PAC observed in our data may exist, namely that the higher performance of the *spectrotemporal modulation* model in superficial PAC is merely a vascular artefact. For example, the higher *spectrotemporal modulation* model performance may have 'bled in' from non-PAC, and did so only in superficial layers as this cortical depth has the lowest spatial specificity due to its proximity to pial veins. While it is difficult to exclude this scenario altogether, the results reported in our previous study make this alternative explanation unlikely[29]. If the increased performance of the *spectrotemporal modulation* model compared to the *frequency* model in superficial PAC was a large vein artefact, we would expect this relative increase in model performance to be strongest in superficial locations *closest* to veins. Instead, after quantification of each gridpoint's distance to a vein we observed that while the overall model performance was dependent on proximity to veins, the *difference* in model performance was *unrelated* to vein proximity (see Fig. 5f in[29]). Thus, in spite of the contribution of large veins to the signal, we argue that through model comparison and careful evaluation of the data, the cortical depth dependent effect is reliable and interpretable.

Feature-selective auditory attention, auditory category learning, and input from other sensory modalities have been shown to modify primary auditory processing[14,15,18,55]. While our current results characterized fundamental processing differences across the layers of the human primary auditory cortex, the major functions of the laminar architecture of the neocortex may only emerge fully when it is engaged in such behaviourally relevant tasks. Our results, describing stimulus-evoked laminar PAC processing, provide a framework and the methodology that can be employed for exploring how laminar processing differences in PAC support auditory behaviour throughout increasingly complex environmental settings and demands.

Methods

Ethics statement. The Institutional Review Board (IRB) for human subject research at the University of Minnesota approved the experimental procedures. The experiment was performed in accordance with the approved guidelines and the Declaration of Helsinki. We obtained written informed consent from each participant before the measurements were started.

Participants. The participants consisted of six healthy volunteers (four females and two males; mean age [SD] = 28.5 [7.8]) who did not have a history of neurological disease or hearing disorders.

Experimental stimuli and design. We described the data analysed in this study in an earlier methodological study[29] that examined how a T_2- vs. a T_2*-weighted dataset, both acquired at submillimetre spatial resolution, performed across various fMRI analyses. In that study we observed that the T_2-weighted data was preferable for the examination of cortical depth dependent preference maps requiring high specificity, while the T_2*-weighted data was preferable for model-based comparisons. Therefore, we now compare the performance of computational models based on an optimized subset of the dataset. The anatomical data reported in this study was previously used in a study that examined the columnar stability of preference maps in auditory cortex[54].

In the current study we used the data of the first two of the original three scanning sessions. In the first session, we collected high resolution anatomical data for segmenting the white matter (WM) – grey matter (GM) boundary and GM – cerebrospinal fluid (CSF) boundary. We furthermore used the anatomical data to sample the cortical layers[56] and to compute a myelin related contrast (MRC) map[41,42]. In the second session, we collected GE-EPI data at high spatial resolution while complex, natural sounds were presented to the participants. We presented a total of 144 sounds that covered six categories: speech utterances, voice sounds, animal cries, musical instruments, tool sounds, and scenes of nature (24 sounds per category). The sampling rate of the sounds was 16 kHz and they were 1000 ms in duration. We equalized the energy (RMS) across sounds, and ramped their onset and offset with a 10 ms linear slope. We used MRI-compatible S14 model earbuds of Sensimetrics Corporation (www.sens.com) to present the sounds in the scanner.

We divided the 144 sounds into 4 non-overlapping sets (36 sounds per set). This division was pseudorandom as we ensured that each set contained an equal number of sounds of each semantic category (i.e., 6 category exemplars per set). Data in the second scanning session was collected in 12 fMRI runs. We presented one set of 36 sounds per run, and each sound was repeated once in that run. A set of sounds was presented in three runs, such that each sound was presented a total of three times across the whole experiment. We jittered the inter-stimulus interval with 2, 3, or 4 TRs, and played the sounds with an additional random jitter in the silent gaps between the functional data acquisition. We also added 8% of silent trials to increase the variability in the acquired BOLD signal. Participants performed a one-back task, and pressed a button if they heard the same sound in two consecutive trials. This occurred in 6% of the trials, and we excluded these "repeat trials" from the analysis.

Acquisition of MRI data.   We performed the measurements using a 7 Tesla whole body magnet (Magnex Scientific, Abingdon, UK) that was driven by a Siemens console (Siemens Medical Systems, Erlangen). Specifically, a 32 channel loop transceiver (custom-built, whole head) along with a high performance head gradient insert was used. In the first session we collected the following anatomical datasets: Two T_1 datasets, a proton density [PD] dataset, and a T_2* weighted dataset at a high spatial resolution (voxel size = 0.6 mm isotropic). The T_1 weighted data was acquired using a modified magnetization-prepared rapid gradient-echo (MPRAGE) sequence (repetition time [TR] = 3100 ms; time to inversion [TI] = 1500 ms; time echo [TE] = 3.45 ms; flip angle = 4°; generalized autocalibrating partially parallel acquisition [GRAPPA] = 3; matrix size = 384 × 384; 256 slices). PD images were acquired with the same MPRAGE sequence as used for the collection of the T_1 weighted images, but without the inversion pulse (TR = 2160 ms; TE = 3.45 ms; flip angle = 4°; GRAPPA = 3; matrix size = 384 × 384; 256 slices; pixel bandwidth = 200 Hz/pixel). A T_2* dataset was acquired using a modified MPRAGE sequence that allows freely setting the TE (TR = 3700 ms; TE = 16 ms; flip angle = 4°; GRAPPA = 3; matrix size = 384 × 384; 256 slices). Acquisition time for the T_1, PD, and T_2* datasets were ~14, 5, and 8 minutes, respectively.

We acquired GE-EPI data during the second scanning session (voxel size = 0.8 mm isotropic; TR = 2400 ms; TE = 22.8 ms; time of volume acquisition [TA] = 1200 ms; silent gap = 1200 ms; GRAPPA = 3; multiband = 2; slices = 36). Slice placement included the bilateral superior temporal plane (STP) and parts of the superior temporal gyrus (STG), covering the auditory cortices. This session comprised 12 runs, each ~5 minutes in duration. A T_1 weighted scan was acquired for the purpose of realignment across sessions and for slice placement.

Anatomical data analysis.   We divided the T_1 dataset by the PD images to minimize inhomogeneities induced by the receive coil profile (Van de Moortele et al.[57]; left column of Fig. S3). We further corrected the resulting dataset for residual inhomogeneities, down-sampled it to the spatial resolution of the functional data, and brought it to anterior and posterior commissural (ACPC) space. Then, we used BrainVoyager QX 2.8 (Brain Innovation, Maastricht, Netherlands) to identify the WM – GM boundary and the GM – CSF boundary. These boundaries were manually edited to correct segmentation mistakes (i.e., general corrections throughout the brain and detailed editing of both boundaries within the supratemporal plane).

Next, for each subject and hemisphere, we measured cortical thickness using a procedure based on the Laplace equation[52,56,58] (as implemented in BrainVoyager QX) and defined a cortical depth-dependent grid that covered HG and cortical regions in its immediate vicinity (Figs 3 and S3). These grids contained nine cortical depths ranging from 0.1 (corresponding to deep grey matter) to 0.9 (corresponding to superficial grey matter) cortical depth. The WM-GM and GM-CSF boundaries themselves were not part of the cortical depth-dependent grids.

Myelination is more dense in primary sensory cortices, including in the PAC, than in the surrounding cortex[27,28], and therefore MRC maps can be used to non-invasively define the PAC in individuals. We created MRC maps by dividing the T_1 weighted dataset by the T_2* weighted dataset, which reduced receive coil inhomogeneities and enhanced the intracortical anatomical contrast (see De Martino et al.[41] for details; right column of Fig. S3). Next, the T_1/T_2* dataset was corrected for residual inhomogeneities and brought to ACPC space. We limited the MRC maps to the supratemporal plane as segmentations of sufficient precision existed only for this brain region, and created the maps by color coding the intensity of the T_1/T_2* dataset within the cortical ribbon. We used the T_2* weighted dataset to define and remove veins from the MRC maps (see[29] for details). In each individual hemisphere, the PAC was identified by dividing the cortical depth dependent grid in two parts of equal size based on the MRC map (see Fig. 3 for a representative subject, and Fig. S2 for the data of all subjects). We refer to the half of the grid with higher MRC as PAC, and to the half of the grid with lower MRC as non-PAC.

Functional data analysis.   We analysed the functional data with BrainVoyager QX and custom MATLAB code (The MATHWORKS Inc., Natick, MA, USA). The functional data were pre-processed by correcting for slice scan time, correcting for motion, applying a temporal high pass filter (removing drifts of 2 cycles and less per run), and temporally smoothing the time series (2 data points). We then co-registered the functional to the anatomical data and transformed the dataset to ACPC space.

We estimated the response in each voxel to the natural sounds following the procedure outlined in[37,59]. In short, we first denoised the data[60] (using the toolbox available at http://kendrickkay.net/GLMdenoise/) and then estimated the hemodynamic response function (HRF) separately for each voxel but common to all sounds[61]. This HRF was used to, per voxel, compute a response estimate (beta weight) for each sound.

Cortical responses to the sounds were analysed with two encoding models that represent different hypotheses on sound processing. A first *frequency* model was simplest, describing sound processing in terms of the frequency preference of cortical neuronal populations. We represented sounds in the space of the *frequency* model by using a biologically-inspired model of auditory cochlear to midbrain processing[62] (i.e., the first stage of the NSL Tools package; available at http://www.isr.umd.edu/Labs/NSL/Software.htm). That is, the sounds were passed through a filterbank that produces sound spectrograms as output (128 overlapping bandpass filters equally spaced along a

logarithmic frequency axis covering 5.3 octaves; minimum frequency = 180 Hz; maximum frequency = 7040 Hz). We averaged these spectrograms over time, which resulted in 128 model parameters to be estimated. A second *spectrotemporal modulation* model described cortical sound processing as the frequency-dependent neuronal processing of spectrotemporal modulations. We represented sounds in the space of the *spectrotemporal modulation* model by passing the output of the first stage of the NSL model to a second model stage that represents cortical auditory processing. That is, a set of modulation filters (temporal modulation rates of $\omega = [1, 3\ 9, 27]$ Hz; spectral modulation scales of $\Omega = [0.5, 1, 2, 4]$ cycles/octave) was applied to the sounds' spectrograms to extract their modulation content. We averaged the resulting sound representation over time, divided the frequency axis in 8 bins with equal bandwidth in octaves, and averaged the modulation energy within each frequency bin, which resulted in 128 model parameters to be estimated (8 frequencies x 4 temporal modulation rates x 4 spectral modulation scales). We refer to the representation of the sounds in the space of the computational models as feature matrix $\mathbf{W}$ (of size $[S \times F]$, where S is the number of natural sounds and F is the number of features to estimate).

For each voxel that significantly responded to the training sounds ($p < 0.05$, uncorrected), we estimated the model parameters in 4-fold cross-validation. Each cross-validation used three out of four sound sets for model training (108 training sounds S_{train}), and the remaining sound set was used for model testing (36 testing sounds S_{test}). As these cross-validations followed the sound set division as implemented in the experimental design, in each cross-validation the responses to the training and testing sounds were estimated on completely distinct sets of fMRI data collection runs. We modelled the response to training sounds in each voxel i as a linear combination of the features in matrix $\mathbf{W_{train}}$ $[S_{train} \times F]$:

$$Y_{train,i} = \mathbf{W_{train}} R_i \qquad (1)$$

where R_i $[F\mathrm{x}1]$ quantifies the contribution of each of the model features to the response of voxel i. We used ridge regression[63] to solve equation 1, where we determined regularization parameter λ for each voxel independently by automatic inspection of the stability of the ridge trace[21,64]. We assessed model performance with a sound identification analysis[21,65,66], where the model was evaluated by its ability to correctly predict responses to sounds not used in model training (i.e., testing sounds). That is, in each voxel the response to all testing sounds was estimated in the same manner as done for training sounds, with the exception that the voxel-specific HRF was based on the training sounds. Next, we predicted the response to the testing sounds in voxel i based on the estimated feature preference R_i $[F\mathrm{x}1]$ and the representation of the testing sounds in the model space $\mathbf{W_{test}}$ $[S_{test} \times F]$:

$$\hat{Y}_{test,i} = \mathbf{W_{test}} R_i \qquad (2)$$

For each sound k, we horizontally concatenated the predicted responses to test sounds $\hat{Y}_{test,k}$ $[1 \times V]$ across the four cross-validations, and computed a *sound identification score* by correlating $\hat{Y}_{test,k}$ to the measured fMRI responses $\mathbf{Y_{test}}$ $[S_{test} \times V]$. For each sound k, we then noted the rank of the correlation between predicted response $\hat{Y}_{test,k}$ and measured response of test sound $Y_{test,k}$. If the model was able to correctly match the predicted sound k with measured sound k, this rank would be equal to 1. Instead, a rank equal to S_{test} represents the worst prediction accuracy, and a rank equal to $S_{test}/2$ (S_{test}, the number of testing sounds, was equal to 36) represents chance. We defined the prediction accuracy P_k as 1 - the normalized rank:

$$P_k = 1 - \frac{r_k - 1}{S_{test} - 1} \qquad (3)$$

Values of P_k range between 0 (i.e., consistently wrong predictions) and 1 (i.e., perfect prediction), with chance = 0.5. We averaged the prediction accuracy across all testing sounds to obtain the model's overall accuracy. Additionally, we obtained the prediction accuracy per sound category by averaging over the subsets of testing sounds belonging to the same category (i.e. separately for speech, voice, animals, music, tools, and nature sounds). We statistically compared the performance of the models, based on sound responses in the STP as a whole, by Fisher transformation of the prediction accuracy values followed by a two-sided paired t-test.

We assessed the cortical depth-dependent model performance by sampling the responses to the sounds (matrices $\mathbf{Y}$) and the trained encoding models (matrices $\mathbf{R}$) on the cortical depth-dependent grids and repeating the analysis procedure for deep, middle, and superficial cortical depths. Statistical differences in model performance were assessed separately for the half of the grid with highest and lowest MRC (i.e., the primary and non-primary auditory cortex, respectively). Per MRC-defined dataset, a two-way repeated measures (RM) ANOVA with the factors 'Computational Model' and 'Cortical Depth' after Fisher transformation of the prediction accuracy values was performed, followed by a multiple comparison corrected paired t-test comparing the performance of the two models per cortical depth.

Maps of acoustic feature preference. We created maps of feature preference by color-coding each voxel or grid point according to the frequency, temporal modulation rate, or spectral modulation scale with the highest weight in the trained *spectrotemporal modulation* model. The voxels' best frequency (BF; tonotopy maps) was mapped using a red-yellow-green-blue colour scale, where voxel tuning to low frequencies was indicated in red colours and tuning to high frequencies was indicated in blue. A yellow-green-blue-purple colour scale was used for the maps of temporal modulation rate and spectral modulation scale, where low and high rates and scales were assigned with yellow and purple colours, respectively.

We created group maps by bringing the individual hemispheres in cortex-based aligned (CBA)[67] space. Each individual map was smoothed (FWHM = 2.4 mm), and sampled in this CBA space. Group maps were computed as the median of the individual subject maps at each vertex that was included in at least 4 individual maps.

Data Availability

The data will be made available by the corresponding author upon reasonable request.

References

1. Mitani, A. *et al.* Morphology and laminar organization of electrophysiologically identified neurons in the primary auditory cortex in the cat. *J. Comp. Neurol.* **235**, 430–447 (1985).
2. Douglas, R. J. & Martin, K. A. C. Neuronal Circuits of the Neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004).
3. Linden, J. F. Columnar Transformations in Auditory Cortex? A Comparison to Visual and Somatosensory Cortices. *Cereb. Cortex* **13**, 83–89 (2003).
4. Hirsch, J. A. & Martinez, L. M. Laminar processing in the visual cortical column. *Current Opinion in Neurobiology* **16**, 377–384 (2006).
5. Brumberg, J. C., Pinto, D. J. & Simons, D. J. Cortical columnar processing in the rat whisker-to-barrel system. *J. Neurophysiol.* **82**, 1808–1817 (1999).
6. Atencio, C. A. & Schreiner, C. E. Laminar diversity of dynamic sound processing in cat primary auditory cortex. *J. Neurophysiol.* **103**, 192–205 (2010).
7. Heimel, J. A., Van Hooser, S. D. & Nelson, S. B. Laminar organization of response properties in primary visual cortex of the gray squirrel (Sciurus carolinensis). *J. Neurophysiol.* **94**, 3538–54 (2005).
8. Phillips, D. P. & Irvine, D. R. F. Responses of single neurons in physiologically defined area AI of cat cerebral cortex: sensitivity to interaural intensity differences. *Hear. Res.* **4**, 299–307 (1981).
9. Huang, C. L. & Winer, J. A. Auditory thalamocortical projections in the cat: Laminar and areal patterns of input. *J. Comp. Neurol.* **427**, 302–331 (2000).
10. Atencio, C. A., Sharpee, T. & Schreiner, C. E. Hierarchical computation in the canonical auditory cortical circuit. *Proc. Natl. Acad. Sci.* **106**, 21894–21899 (2009).
11. Sadagopan, S. & Wang, X. Level invariant representation of sounds by populations of neurons in primary auditory cortex. *J. Neurosci.* **28**, 3415–3426 (2008).
12. Sadagopan, S. & Wang, X. Nonlinear Spectrotemporal Interactions Underlying Selectivity for Complex Sounds in Auditory Cortex. *J. Neurosci.* **29**, 11192–11202 (2009).
13. Fritz, J. B. Differential Dynamic Plasticity of A1 Receptive Fields during Multiple Spectral Tasks. *J. Neurosci.* **25**, 7623–7635 (2005).
14. Fritz, J., Shamma, S., Elhilali, M. & Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **6**, 1216–1223 (2003).
15. Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A. & Schroeder, C. E. Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron* **53**, 279–292 (2007).
16. Nelken, I., Fishbach, A., Las, L., Ulanovsky, N. & Farkas, D. Primary auditory cortex of cats: feature detection or something else? *Biol. Cybern.* **89**, 397–406 (2003).
17. Ley, A. *et al.* Learning of new sound categories shapes neural response patterns in human auditory cortex. *J. Neurosci.* **32**, 13273–80 (2012).
18. Ohl, F. W., Scheich, H. & Freeman, W. J. Change in pattern of ongoing cortical activity with auditory category learning. *Nature* **412**, 733–736 (2001).
19. Yacoub, E. *et al.* Imaging brain function in humans at 7 Tesla. *Magn. Reson. Med.* **45**, 588–94 (2001).
20. Meyer, M., Liem, F., Hirsiger, S., Jancke, L. & Hanggi, J. Cortical surface area and cortical thickness demonstrate differential structural asymmetry in auditory-related areas of the human cortex. *Cereb. Cortex* **24**, 2541–2552 (2014).
21. Santoro, R. *et al.* Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *PLoS Comput Biol* **10**, e1003412 (2014).
22. De Martino, F. *et al.* Spin echo functional MRI in bilateral auditory cortices at 7T: An application of B1 shimming. *Neuroimage* **63**, 1313–1320 (2012).
23. Da Costa, S. *et al.* Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. *J. Neurosci.* **31**, 14067–14075 (2011).
24. Langers, D. R. M. Assessment of tonotopically organised subdivisions in human auditory cortex using volumetric and surface-based cortical alignments. *Hum. Brain Mapp.* **35**, 1544–1561 (2014).
25. Humphries, C., Liebenthal, E. & Binder, J. R. Tonotopic organization of human auditory cortex. *Neuroimage* **50**, 1202–1211 (2010).
26. Schönwiesner, M. & Zatorre, R. J. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc. Natl. Acad. Sci. USA* **106**, 14611–14616 (2009).
27. Hackett, T. A., Preuss, T. M. & Kaas, J. H. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* **441**, 197–222 (2001).
28. Nieuwenhuys, R. The myeloarchitectonic studies on the human cerebral cortex of the Vogt-Vogt school, and their significance for the interpretation of functional neuroimaging data. *Brain Struct. Funct.* **218**, 303–352 (2013).
29. Moerel, M. *et al.* Sensitivity and specificity considerations for fMRI encoding, decoding, and mapping of auditory cortex at ultra-high field. *Neuroimage* **164**, 18–31 (2018).
30. Suga, N., O'Neill, W. E. & Manabe, T. Harmonic-sensitive neurons in the auditory cortex of the mustache bat. *Science* **203**, 270–274 (1979).
31. Lewicki, M. S. & Konishi, M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl. Acad. Sci. USA* **92**, 5582–5586 (1995).
32. Griffiths, T. D. & Warren, J. D. What is an auditory object? *Nat. Rev. Neurosci.* **5**, 887–92 (2004).
33. Miller, L. M., Escabí, M. A., Read, H. L. & Schreiner, C. E. Functional convergence of response properties in the auditory thalamocortical system. *Neuron* **32**, 151–160 (2001).
34. Clavagnier, S., Falchier, A. & Kennedy, H. Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. in. *Cognitive, Affective and Behavioral Neuroscience* **4**, 117–126 (2004).
35. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–7 (2001).
36. Rauschecker, J. P. & Tian, B. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J. Neurophysiol.* **91**, 2578–2589 (2004).
37. Moerel, M., De Martino, F. & Formisano, E. Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. *J. Neurosci.* **32**, 14205–14216 (2012).
38. Wessinger, C. M. *et al.* Hierarchical Organization of the Human Auditory Cortex Revealed by Functional Magnetic Resonance Imaging S7. **13**, 1–7 (2001).
39. Barton, B., Venezia, J. H., Saberi, K., Hickok, G. & Brewer, A. A. Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc. Natl. Acad. Sci.* **109**, 20738–20743 (2012).
40. Glasser, M. F. & Van Essen, D. C. Mapping human cortical areas *in vivo* based on myelin content as revealed by t1- and t2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011).
41. De Martino, F. *et al.* High-Resolution Mapping of Myeloarchitecture *In Vivo*: Localization of Auditory Areas in the Human Brain. *Cereb. Cortex* **25**, 3394–3405 (2015).

42. Dick, F. *et al. In vivo* functional and myeloarchitectonic mapping of human primary auditory areas. *J Neurosci* **32**, 16095–16105 (2012).
43. Morosan, P. *et al.* Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* **13**, 684–701 (2001).
44. Rivier, F. & Clarke, S. Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage* **6**, 288–304 (1997).
45. Galaburda, A. & Sanides, F. Cytoarchitectonic organization of the human auditory cortex. *J. Comp. Neurol.* **190**, 597–610 (1980).
46. Striem-Amit, E., Hertz, U. & Amedi, A. Extensive Cochleotopic Mapping of Human Auditory Cortical Fields Obtained with Phase-Encoding fMRI. *PLoS One* **6**, e17832 (2011).
47. Rademacher, J. *et al.* Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* **13**, 669–83 (2001).
48. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
49. Duvernoy, H. M., Delon, S. & Vannson, J. L. Cortical blood vessels of the human brain. *Brain Res. Bull.* **7**, 519–579 (1981).
50. Uludağ, K., Müller-Bierl, B. & Uğurbil, K. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *Neuroimage* **48**, 150–165 (2009).
51. Yacoub, E., Shmuel, A., Logothetis, N. & Uğurbil, K. Robust detection of ocular dominance columns in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. *Neuroimage* **37**, 1161–1177 (2007).
52. De Martino, F. *et al.* Cortical Depth Dependent Functional Responses in Humans at 7T: Improved Specificity with 3D GRASE. *PLoS One* **8**, e60514 (2013).
53. Polimeni, J. R., Fischl, B., Greve, D. N. & Wald, L. L. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage* **52**, 1334–1346 (2010).
54. Moerel, M., De Martino, F., Uğurbil, K., Formisano, E. & Yacoub, E. Evaluating the columnar stability of acoustic processing in the human auditory cortex. *J. Neurosci.* **38**, 7822–7832 (2018).
55. O'Connell, M. N., Barczak, A., Schroeder, C. E. & Lakatos, P. Layer Specific Sharpening of Frequency Tuning by Selective Attention in Primary Auditory Cortex. *J. Neurosci.* **34**, 16496–16508 (2014).
56. Zimmermann, J. *et al.* Mapping the Organization of Axis of Motion Selective Features in Human Area MT Using High-Field fMRI. *PLoS One* **6**, e28716 (2011).
57. Van de Moortele, P.-F. *et al.* T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2N contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *Neuroimage* **46**, 432–446 (2009).
58. Jones, S. E., Buchbinder, B. R. & Aharon, I. Three-dimensional mapping of cortical thickness using Laplace's equation. *Hum. Brain Mapp.* **11**, 12–32 (2000).
59. Moerel, M., De Martino, F., Santoro, R., Yacoub, E. & Formisano, E. Representation of pitch chroma by multi-peak spectral tuning in human auditory cortex. *Neuroimage* **106**, 161–9 (2015).
60. Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F. & Wandell, B. A. GLMdenoise: A fast, automated technique for denoising task-based fMRI data. *Front. Neurosci.* 1–15, https://doi.org/10.3389/fnins.2013.00247 (2013).
61. Kay, K. N., David, S. V., Prenger, R. J., Hansen, K. A. & Gallant, J. L. Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum. Brain Mapp.* **29**, 142–156 (2008).
62. Chi, T., Ru, P. & Shamma, S. A. Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**, 887–906 (2005).
63. Hoerl, A. E. & Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12**, 55–67 (1970).
64. Hoerl, A. E. & Kennard, R. W. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **12**, 69–82 (1970).
65. Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).
66. Moerel, M. *et al.* Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. *J. Neurosci.* **33**, 11888–98 (2013).
67. Goebel, R., Esposito, F. & Formisano, E. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* **27**, 392–401 (2006).

Acknowledgements

Author Contributions

M.M., F.D.M. and E.F. designed the experiment. E.Y. and M.M. collected the data. M.M. and F.D.M. analysed the data. All authors contributed to writing the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-41965-w.

Competing Interests: The authors declare no competing interests.

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