Population coding for visual and auditory quantity in human numerotopic maps

Joram Soch[1,4], Anne-Sophie Kieslinger[1,4], Robert Trampel[2], Andreas Nieder[3] and Michael A. Skeide[1] ✉

[1] Research Group Learning in Early Childhood,
Max Planck Institute for Human Cognitive and Brain Sciences,
Stephanstraße 1A, 04103 Leipzig, Germany

[2] Department of Neurophysics,
Max Planck Institute for Human Cognitive and Brain Sciences,
Stephanstraße 1A, 04103 Leipzig, Germany

[3] Animal Physiology Unit, Institute of Neurobiology,
Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany

[4] These authors contributed equally: Joram Soch, Anne-Sophie Kieslinger.

✉ e-mail: skeide@cbs.mpg.de

Correspondence should be addressed to Michael A. Skeide

Abstract

Neural populations that are able to extract quantitative information from multiple sensory domains are essential for the survival of numerous species. How quantity is encoded across special senses is far from understood. Here, we identified an overarching coding scheme for visual and auditory numerosity using high-field functional magnetic resonance imaging at 7 Tesla in humans. Based on a neurobiologically plausible model informed by electrophysiological data, we discovered hemodynamic responses revealing logarithmic Gaussian tuning to numerosity in both domains. Responses turned out to be organized topographically, forming numerotopic maps. We found several visual maps scattered over the association cortices and anatomically distinct auditory maps in superior temporal and premotor cortices. The present data shed light on the multisensory foundations of numerical information processing in the human brain. These insights open avenues for future research exploring how different species detect quantity in different sensory modalities.

Main

Electrophysiological recordings from single neurons in the parietal and medial temporal lobe of human neurosurgical patients uncovered populations that selectively respond to simultaneously displayed dots of a certain numerosity. Neuronal response profiles revealed peak activity for one preferred numerosity out of the numerosities one to five and a gradual decrease of activity the more the number of dots deviated from the preferred value[1,2].

Invasive recording techniques achieve unparalleled spatial and temporal resolution but suffer from selection bias to neurological patients and sampling bias to a relatively small set of predefined regions of interest. High-field magnetic resonance imaging, however, despite measuring neuronal activity only indirectly, revealed whole-brain wide neural populations with visually and haptically elicited hemodynamic response profiles similar to the electrophysiological response profile of numerosity-tuned neurons[3–5]. It is currently unknown, however, whether this common coding scheme also governs other special senses, such as audition. Audition is a critical test case for numerosity coding since, unlike vision, it primarily draws on temporal rather than spatial numerical information. This means that auditory numerosity processing does not operate in a simultaneous, but in a sequential mode of stimulus presentation. Moreover, previous work could show that visually and haptically elicited hemodynamic responses form topographic maps in which neighboring numerosities are represented in adjacent patches of cortex[3–5]. It remains to be explored whether auditory responses also follow this functional mapping principle.

Here, we set out to explore the existence of cross-modal visual and auditory population coding for quantity. To this end, we conducted experiments in which human adult participants either viewed dot arrays or listened to beeps. Visual dot arrays were presented simultaneously with varying positions but constant luminance. Beeps were presented sequentially with randomly varying pitch. The numerosities one to five appeared 384 times in each modality while recording high-field magnetic

resonance imaging data at 7 Tesla. Following previous electrophysiological and magnetic resonance imaging work, we employed a single-subject design in which twelve individual data sets served as replication units and not as measurement units.

We consistently found topographically organized neural populations with logarithmic Gaussian tuning to numerosity across senses, presentation modes and stimulus features. While sharing coding schemes and topographic mapping principles, these populations generated anatomically dissociable responses to either visual or auditory numerosity.

Results

Hemodynamic responses display logarithmic Gaussian tuning to visual and auditory numerosity

Participants were instructed to press a button whenever visual dot arrays were shown in different color or auditory beep sequences were presented in different pitch (Fig. 1). Across all participants and runs, the hit rate for these catch trials was 99.69% for the visual and 80.38% for the auditory experiments on average (Supplementary Table S1). This behavioral performance indicates that participants were paying attention to the stimuli.

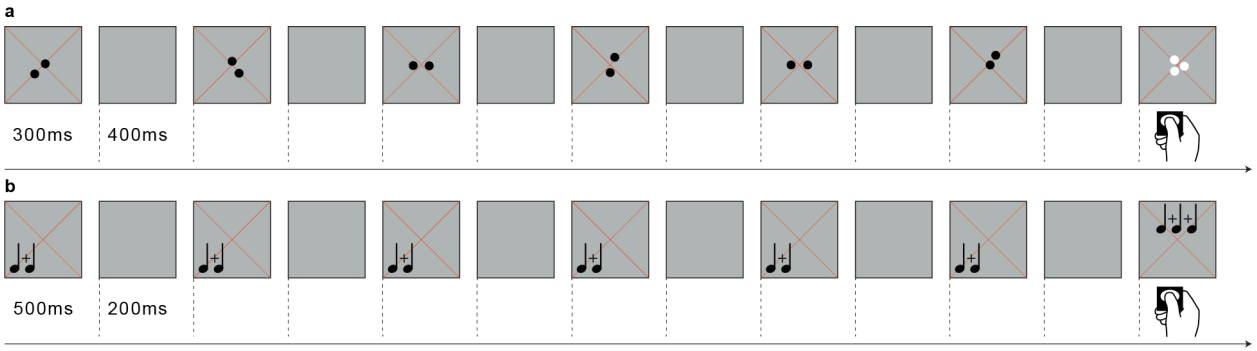

Fig. 1 | Stimuli and task. a, In the visual experiment, participants viewed six different dot arrays per numerosity that were separately displayed for 300 ms and alternating with 400 ms of gray background. Arrays appeared within a radius of 0.75° (visual angle) of a diagonal cross of thin red fixation lines. Summed surface area was kept constant and dot positions were randomly varied. Participants were instructed to press a button whenever white

instead of black dots were shown. These catch trials occurred in 10% of all stimulus events. **b**, In the auditory experiment, participants listened to sequences of beep tones played six times for 500 ms each, alternating with 200 ms of silence. Within each sequence pitch was randomly varied between 333, 359, 392 and 440 kHz. Participants were instructed to press a button whenever 1,000 kHz beeps (superscript notes) were played instead of 333 to 440 kHz beeps (subscript notes). These catch trials occurred in 10% of all stimulus events. **a**, **b**, Numerosities one through five were first presented in ascending order, followed by a block of twenty items, then in descending order, followed by another block of twenty items before this cycle was repeated.

Building on previous magnetic resonance imaging and electrophysiological work in humans and nonhuman primates, we developed numerosity tuning models describing logarithmic Gaussian functions[4,6]. These models were estimated in individual human subjects, separately for each vertex on the cortical surface. Our models captured the peak response of a neural population to a preferred numerosity and the numerosity range to which a neuronal population responds (full width at half maximum, FWHM) (Fig. 2a,c,e,g; for replication in other subjects, see Supplementary Fig. S1 for visual tuning and Supplementary Fig. S2 for auditory tuning). Summarizing the hemodynamic activation based on these two parameters, these models explained substantial amounts of the observed signal variance related to regional differences in numerosity tuning (visual: maximum $R^2 = 0.86$; auditory: maximum $R^2 = 0.71$; both $p < 0.001$) (Fig. 2b,d,f,h; for replication in other subjects, see Supplementary Fig. S1 for visual responses and Supplementary Fig. S2 for auditory responses). Fig. 2a–d depicts visual numerosity tuning in a vertex of the left parieto-occipital cortex. Fig. 2e–h depicts auditory numerosity tuning in a vertex of the left superior temporal cortex.

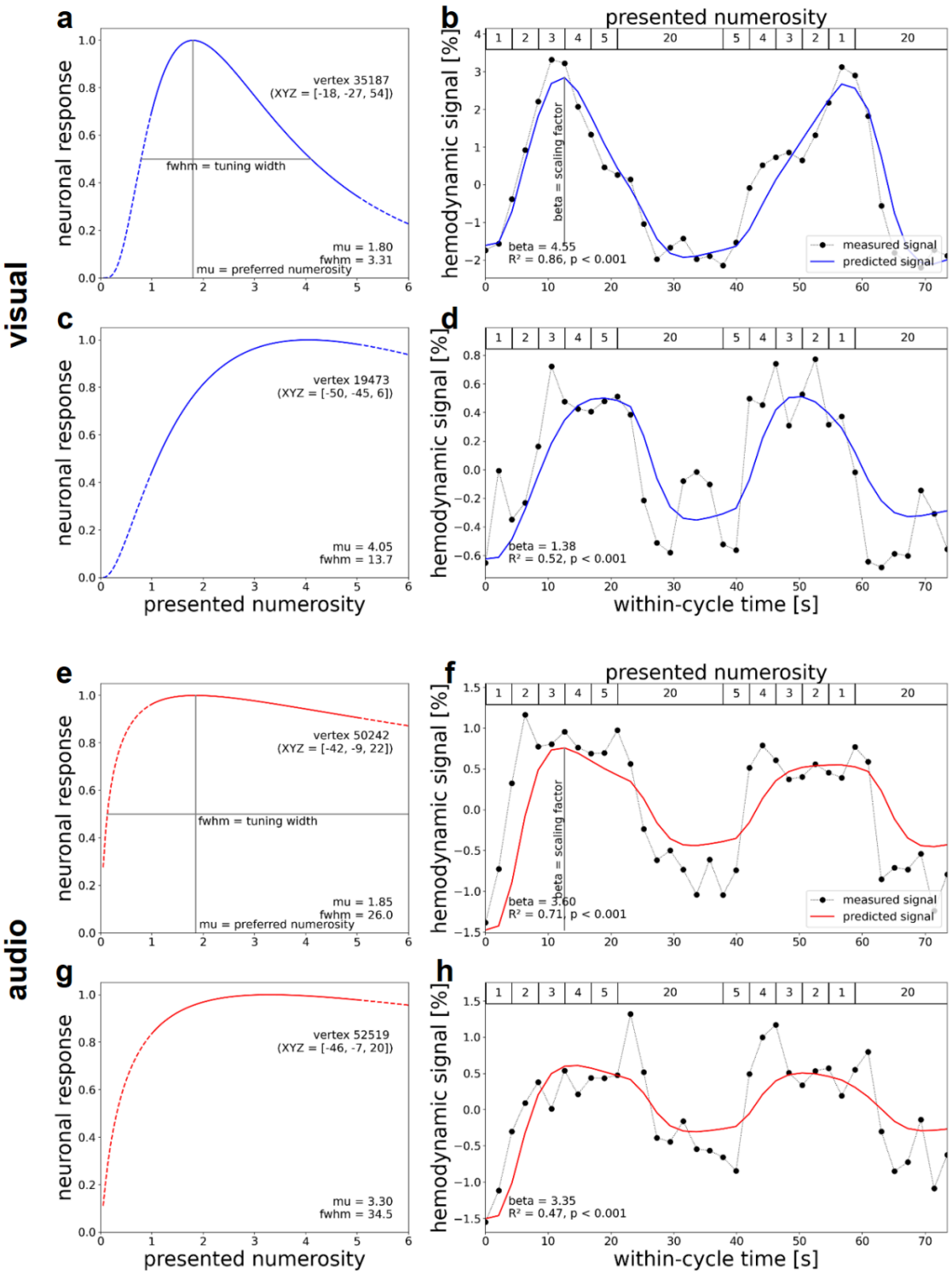

Fig. 2 | Neural numerosity tuning and hemodynamic activation time course. a–h, Results of numerosity population receptive field analyses for an exemplary subject seeing visual stimuli (top, blue) and another subject hearing auditory stimuli (bottom, red). **a–d** Responses in the left parieto-occipital cortex were elicited by visual numerosities. **e–h** Responses in the left superior temporal cortex were elicited by auditory numerosities. **a,c,e,g**,

Logarithmic Gaussian tuning functions are shown for **a,e** a small-numerosity vertex and **c,g** a larger-numerosity vertex per subject (XYZ = coordinates in FreeSurfer fsnative space). These functions are described by a preferred numerosity (mu) and the full width at half maximum (FWHM). **b,d,f,h**, Combining neuronal tuning models with a hemodynamic forward model generated predicted time courses (solid lines) for **b,d** visual numerosity and **f,h** auditory numerosity presentation. Comparing predicted to measured time courses (dotted lines, averaged across runs and cycles) yielded the coefficient of determination (R^2) quantifying the variance explained by the model (calculated across an averaged run).

Visual and auditory presentation activates distinct numerosity maps

In each hemisphere, we identified six visual regions and two auditory regions in which the numerosity tuning models explained more than 20% of the measured signal variance (Fig. 3a,b; for replication in other subjects, see Supplementary Fig. S3). In accordance with the literature[3,5], we denote these numerosity-selective maps using abbreviations starting with "N" for visual numerosity and starting with "Na" for auditory numerosity. Visual numerosity tuning was evident in temporo-occipital (NTO), parieto-occipital (NPO), posterior parietal (NPC1), dorsal anterior parietal (NPC2), ventral anterior parietal (NPC3) and superior frontal (NF) cortices (Fig. 3c). Auditory numerosity tuning was evident in superior temporal (NaT) and premotor (NaF) cortices (Fig. 3d). None of the regions with numerosity tuning in the visual modality revealed numerosity tuning in the auditory modality and vice versa. These numerosity-specific regions were found in a majority of subjects (Fig. 4a,b), but revealed considerable individual differences (Fig. 3c,d). Within numerosity-selective regions, numerosities preferred by vertices, i.e. stimulus intensity for which vertices revealed peak response, were topographically organized, forming numerotopic maps (Fig. 3e,f; for replication in other subjects, see Supplementary Fig. S4).

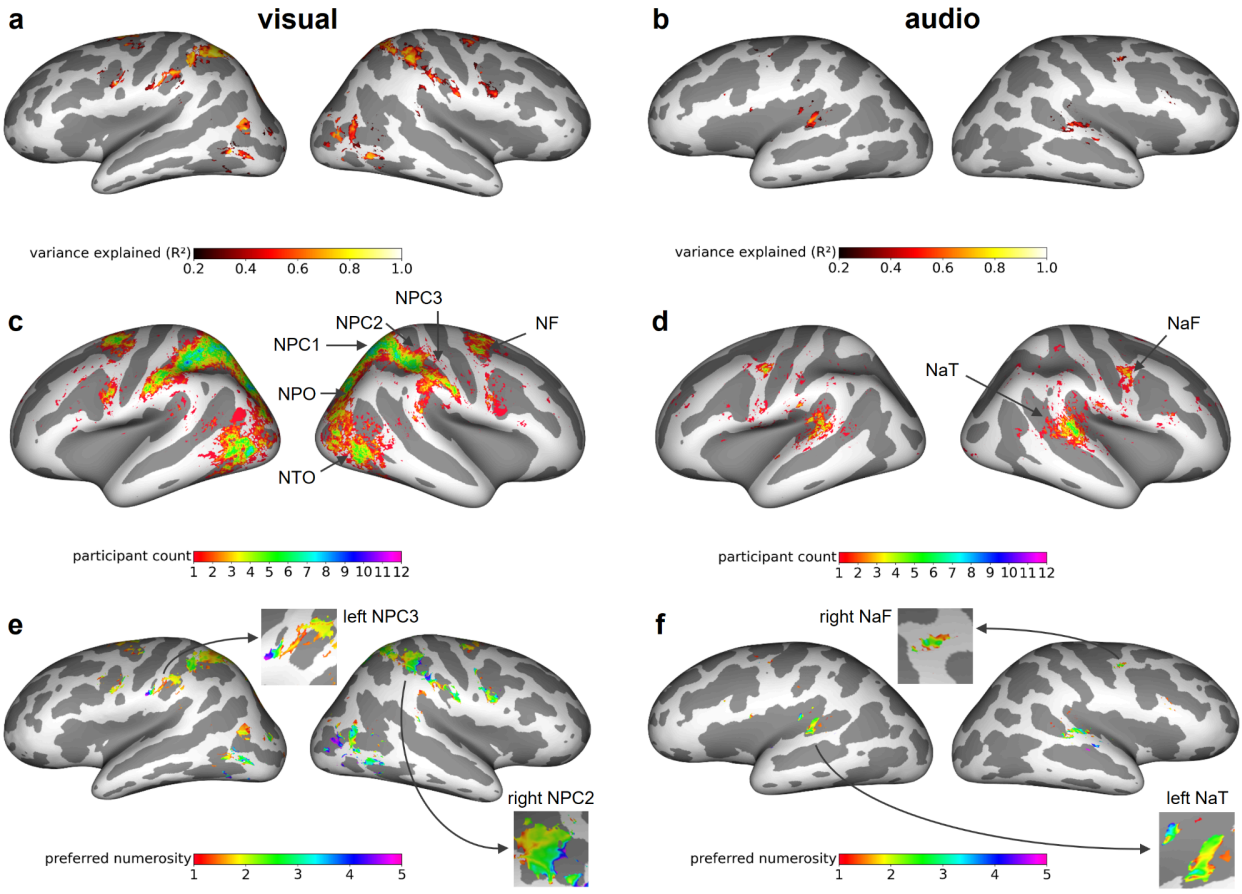

Fig. 3 | Spatially separated neural responses to visual and auditory numerosity. a,b, Inflated surface maps showing regions in which neural tuning models explain at least 20% of the variance for **a** an exemplary subject seeing visual stimuli and **b** another subject hearing auditory stimuli (same subjects as in Fig. 2). The colorbar indicates the variance explained by the model. **c,d**, Participant count maps for **c** visual numerosity and **d** auditory numerosity in standard space (FreeSurfer fsaverage). The colorbar indicates how many subjects exhibit tuning to numerosity according to the variance explained criterion (R^2 > 20%). **e,f**, Preferred numerosity maps of single subjects obtained from **e** visual and **f** auditory experiments (same subjects as in a,b). The colorbar indicates the estimated preferred numerosity. Inset maps show exemplary visual numerotopic organization in one numerosity map per hemisphere and subject. NTO = temporo-occipital visual numerosity field; NPO = parieto-occipital visual numerosity field; NPC1,2,3 = parietal visual numerosity fields; NF = frontal visual numerosity field; NaT = superior temporal auditory numerosity field; NaF = superior frontal auditory numerosity field.

Cortical geometry of visual and auditory maps reveals numerotopic organization

For each single subject, supra-threshold clusters were assigned to numerotopic maps if their cortical surface area was larger than a minimum area (visual: $A_{min} = 50\ mm^2$; auditory: $A_{min} = 25\ mm^2$) and if their distance from previously reported center coordinates[3] was smaller than a maximum distance ($d_{max} = 1.7\ mm$). Whenever multiple maps formed one single cluster, like NPC1,2,3, this cluster was assigned to the map closest to it. Analyzing the cortical surface extent of those maps, we found that most numerotopic maps can be identified in the majority of subjects. It turned out that visual numerotopic maps are larger than auditory numerotopic maps (Fig. 4a,b).

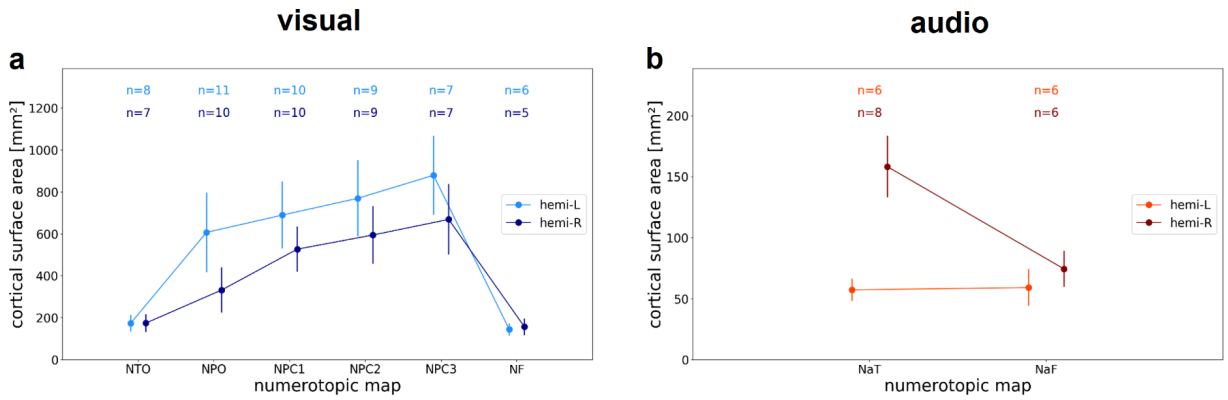

Fig. 4 | Surface area of numerotopic maps. a,b, Cortical surface area differentiated by modality, hemisphere and numerotopic map for **a** visual and **b** auditory numerosity. Numbers on top of the graph (n) indicate the number of subjects which exhibit each map, based on minimum cluster size and maximum distance from the center of the map across the group (see Methods). NTO = temporo-occipital visual numerosity field; NPO = parieto-occipital visual numerosity field; NPC1,2,3 = partietal visual numerosity fields; NF = frontal visual numerosity field; NaT = superior temporal auditory numerosity field; NaF = superior frontal auditory numerosity field. L = left hemisphere; R = right hemisphere.

For each hemisphere, we then partitioned the preferred numerosity of all vertices into bins of width 0.5 (eight bins: 1-1.5, 1.5-2, …, 4-4.5, 4.5-5) and then calculated how much cortical surface is devoted

to representing each level of preferred numerosity. We found a negative correlation between cortical surface area and preferred numerosity for both modalities (Fig. 5a,b; for replication in other subjects, see Supplementary Fig. S5), indicating that fewer cortical surface area represents higher numerosities. We also extracted vertex-wise tuning width, measured as full width at half maximum (FWHM), and analyzed its relationship with binned preferred numerosity in each hemisphere. This analysis revealed a positive correlation between preferred numerosity and FWHM tuning width for both modalities (Fig. 5c,d; for replication in other subjects, see Supplementary Fig. S6), indicating that the larger the numerosity to which a population responds most strongly, the higher the extent to which it also responds to neighboring numerosities.

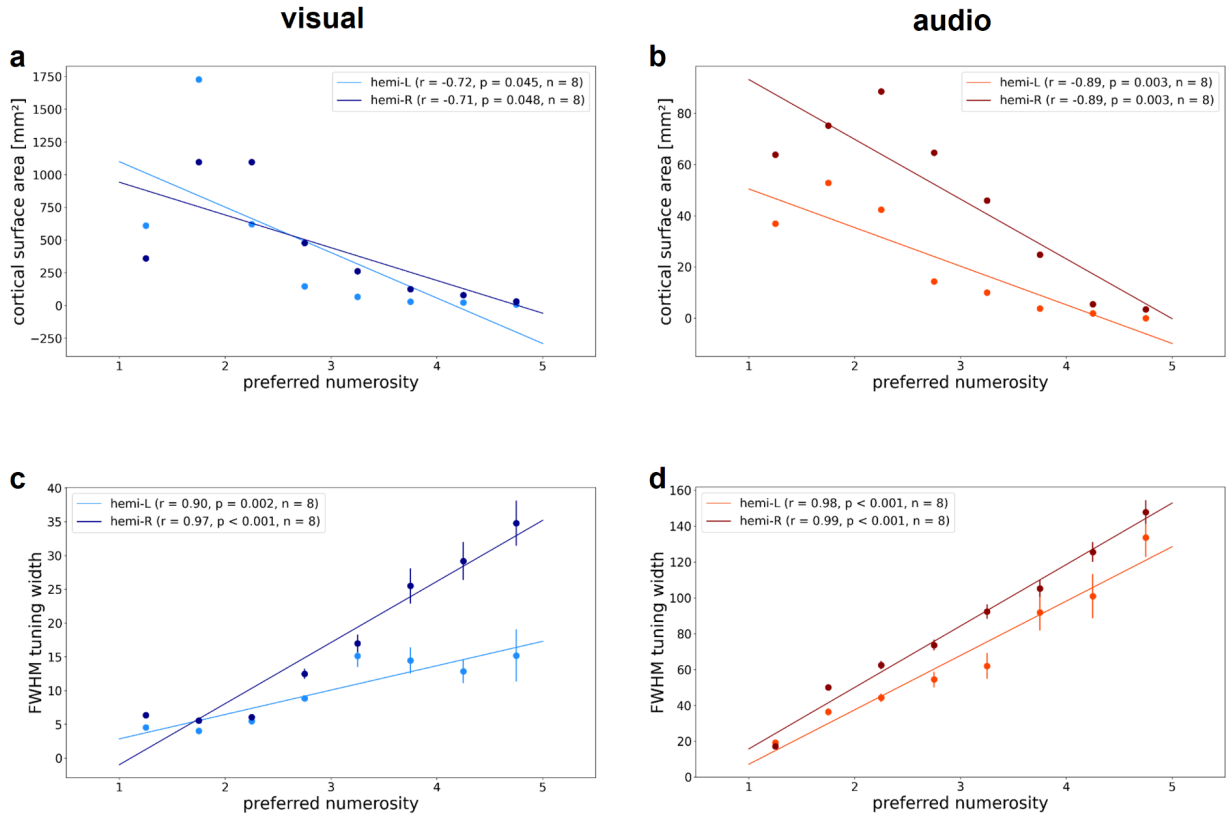

Fig. 5 | Associations between preferred numerosity and cortical surface area. a,b, Associations between preferred numerosity and cortical surface area in all supra-threshold vertices for **a** an exemplary subject seeing

visual stimuli and **b** another subject hearing auditory stimuli (same subjects as in Fig. 2 and Fig. 3). Points correspond to the summed cortical area of all surface triangles of vertices with an average preferred numerosity falling in the respective range (bin width = 0.5). **c,d**, Relationship between preferred numerosity and tuning width in all supra-threshold vertices for **c** visual numerosity and **d** auditory numerosity (same subjects as in **a,b**). Points correspond to mean FWHM in each preferred numerosity bin. Error bars correspond to the standard error of the mean. Solid lines correspond to simple linear regression fits. L = left hemisphere; R = right hemisphere. r = Pearson correlation coefficient. n = number of numerosity bins.

Next, we ran linear mixed-effects models, using summed cortical surface area or average FWHM tuning width as the dependent variable, binned preferred numerosity and brain hemisphere as independent variables while modeling subject as a random effect. We detected significant negative effects of preferred numerosity on surface area (all $p < 0.001$) and significant positive effects of preferred numerosity on tuning width (all $p < 0.001$) in both hemispheres and modalities (Table 1), indicating that the larger the numerosity the smaller the cortical surface area covered and the wider the tuning. There were no significant effects of hemisphere on tuning width (visual: $p = 0.728$; auditory: $p = 0.370$), but significant effects on surface area (visual: $p = 0.042$; auditory: $p = 0.001$), with more cortical surface devoted to auditory numerosity in the right hemisphere and more cortical surface devoted to visual numerosity in the left hemisphere (Table 1; also see Fig. 4).

Table 1 | Associations of surface area and tuning width with preferred numerosity.

dependent variable	independent variable	regression coefficient	z-value	p-value	95% CI (lower)	95% CI (upper)
Session: visual						
area	hemi	-210.149	-2.033	0.042	-412.778	-7.520
	mu [L]	-162.459	-7.137	<0.001	-207.077	-117.842
	mu [R]	-95.716	-4.205	<0.001	-140.334	-51.099
fwhm	hemi	1.728	0.348	0.728	-8.016	11.472
	mu [L]	9.881	9.027	<0.001	7.736	12.027
	mu [R]	9.635	8.801	<0.001	7.489	11.780

Session: auditory						
area	hemi	55.991	3.337	0.001	23.102	88.881
	mu [L]	-12.864	-3.505	<0.001	-20.057	-5.671
	mu [R]	-25.504	-6.720	<0.001	-32.942	-18.066
fwhm	hemi	-7.638	-0.896	0.370	-24.353	9.077
	mu [L]	11.685	6.264	<0.001	8.029	15.341
	mu [R]	16.534	8.574	<0.001	12.754	20.313

area = cortical surface area measured in mm^2. fwhm = tuning width measured as full width at half maximum. hemi = hemisphere. mu = preferred numerosity. CI = confidence interval. L = left hemisphere, R = right hemisphere.

Previously, topographic organization within numerotopic maps was determined by manually delineating the borders of those maps and correlating average preferred numerosity with normalized cortical distance along the principal axis of each map[3,4]. Here, we instead fitted observed preferred numerosities based on each vertex's pial surface coordinates in standard space. This approach resulted in predicted preferred numerosities which were compared to actual preferred numerosities. For all maps, there were significant positive correlations between predicted and actual numerosities (Fig. 4g,h; for replication in other subjects, see Supplementary Fig. S7), indicating that there is a relationship between cortical surface coordinates and represented numerosity. Finally, we analyzed the range of numerosities within each numerotopic map on the cortical surface in both hemispheres and for both modalities. The results obtained suggest that visual NTO and auditory NaT are representing larger numerosity ranges than the respective other maps (Fig. 4i,j; for replication in other subjects, see Supplementary Fig. S8).

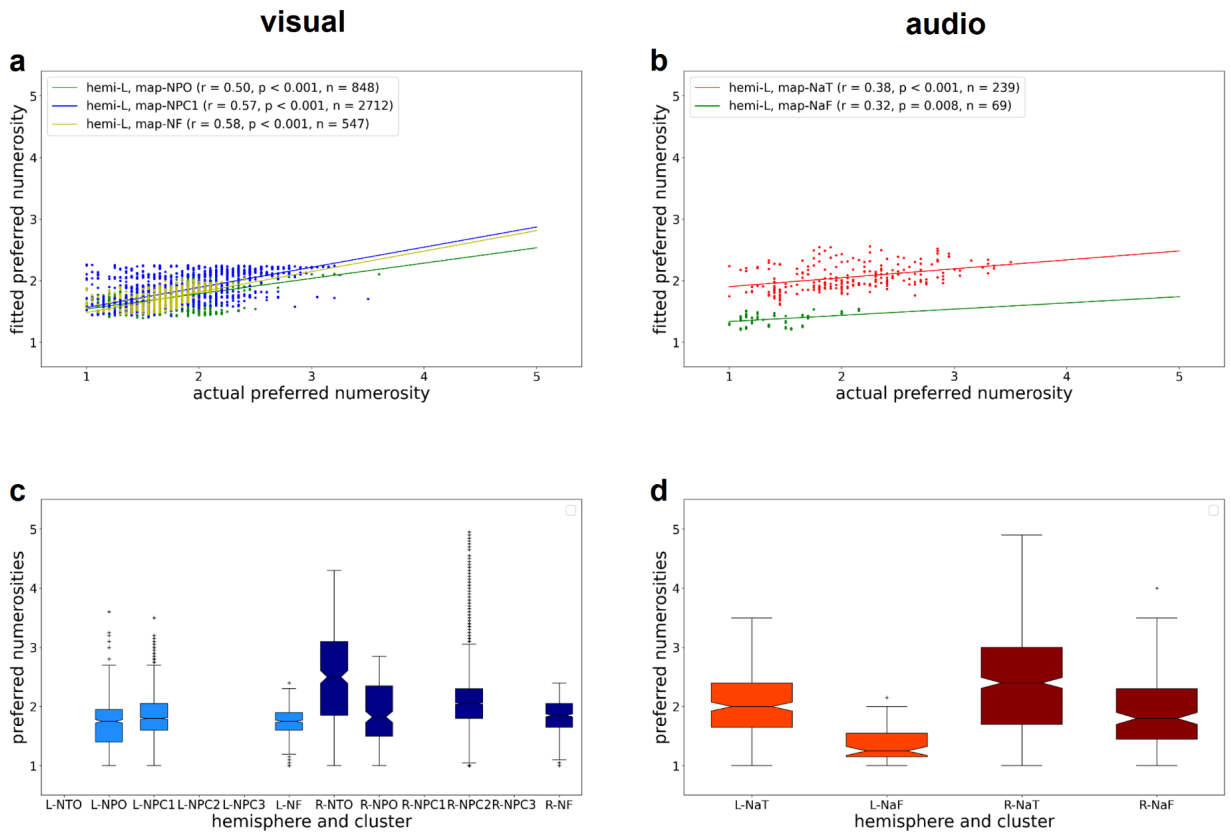

Fig. 6 | Associations between preferred numerosity and cortical geometry. a,b, Plots of preferred numerosity, predicted from cortical surface coordinates, against actual **a** visual and **b** auditory numerosity for numerotopic maps in the left hemisphere for two example subjects (same single subjects as in Fig. 2-5). Significant effects indicate that preferred numerosity can be predicted from surface coordinates, suggesting a spatially systematic organization of numerosity selectivity. **c,d**, Ranges of represented numerosity for all observed **c** visual and **d** auditory numerotopic maps of two example subjects. Whenever a cluster covered multiple maps in one hemisphere (e.g. NPC1,2,3), it was assigned to the center of the map it was closest to.

Discussion

We identified a common neural population coding scheme underlying visual and auditory quantity detection in the human brain. Populations tuned to numerosity were revealed by electrophysiologically inspired models reconstructing hemodynamic response profiles from high-field magnetic resonance imaging data. These populations were specifically activated by either visual or

auditory numerosity and formed anatomically separated numerotopic maps with similar tuning geometries.

The numerosity tuning models used here are based on population receptive field models originally developed for visual cortical responses[7]. Our results go beyond the visual system and demonstrate receptive field-like tuning to auditory numerosities in superior temporal and premotor cortices. Quantity detection thus follows a coherent overarching coding scheme that generalizes across sensory modalities and cortical areas. This coding scheme encompasses temporal and spatial sources of numerical information as well as sequential and simultaneous modes of stimulus presentation. Accordingly, our observation corroborates the vital role of the ability to extract quantitative information from multiple sources of sensory input[5].

While it is well-established that there are topographically organized responses to visual and tactile numerosity in several cortical regions[3-5], we show that topographically organized responses to numerosity also exist in human auditory and premotor cortices. Compared to visual maps, auditory maps turned out to be fewer in number (Fig. 3d), by an order of magnitude smaller (Fig. 4), and larger in the right than in the left hemisphere (Fig. 4b). Moreover, responses to auditory numerosity displayed larger tuning widths compared to visual responses (Fig. 5d), suggesting an inverse relationship between map size and tuning width known from the literature on visual population receptive fields[8]. One possible explanation for these differences between visual and auditory maps might be that it is impossible to simultaneously deliver auditory numerosity stimuli, unlike in vision or somatosensation[5]. Another explanation might be that auditory signals could potentially be noisier than visual signals due to residual background scanner sound. This difference would also be consistent with lower response accuracy in auditory versus visual catch trials.

Electrophysiological recordings in frontal and parietal cortices of macaques uncovered single neurons exhibiting tuned responses to the number of sounds in a sequence which the monkeys matched to the numerosity of a subsequently shown visual dot array[6]. Lack of evidence for auditory numerosity tuning in corresponding human cortices might point to a lack of comparability between tasks. Our passive unisensory perception task differs fundamentally from the reward-based working memory-intensive multisensory integration task used in the nonhuman primate study. Other factors that are beyond experimental control relate to further differences between species and recording techniques[3,9].

Numerosity co-varies with non-numerical physical features of a stimulus. Varying the number of items in a set changes spatial or temporal features while fixing one feature causes another to vary with the number of items[4]. Accordingly, responses to numerosity should be comparable across systematically varied spatial or temporal features to be considered as abstract[6]. In line with this criterion for abstract representation, responses to visual numerosity have been shown to be equal across co-varying spatial features, such as total area, size, density, circumference and shape[4]. To which degree auditory numerosity responses are robust to co-varying temporal features, such as slower and faster rates and longer and shorter durations, remains to be investigated in future experiments. According to a previous auditory event timing study, superior temporal and premotor cortices respond to rates as fast as about 200 ms (1/frequency) and durations as short as about 100 ms[12]. It can be expected that there are yet to be determined lower and upper thresholds with respect to the rate and duration of presentation, at which stimuli will no longer be perceived as a coherent set of discrete items.

It remains an open question how the brain integrates unimodal visual and auditory numerosity information. Integration could be achieved in principle by exchanging information about

modality-specific representations or by transforming numerosity information from different sensory sources into a supramodal neural representation, for example in parietal and frontal cortices[10,11]. Future studies employing multi-unit electrophysiology and stimulation techniques are needed to test these mutually non-exclusive hypotheses.

How numerotopic maps emerge in the course of brain development is currently unknown. It is possible that these maps already form in the first months of life given the behavioral evidence for numerosity discrimination in infants[13]. At the same time, the observation that numerosity discrimination ability refines substantially towards adulthood suggests that numerotopic maps may undergo developmental reorganization, such as increasingly sharper tuning[14].

The experimental approach used here can be adapted to several stimulation and recording techniques beyond magnetic resonance imaging. This versatility of our approach paves the way for future studies tackling the question of how different species detect quantity in different sensory modalities[15–18]. In principle, this line of research could be extended to any senses that are able to process sequential stimuli, even to senses that humans do not have, such as echolocation in bats and electroreception in sharks[19,20].

Methods

Participants

Following previous electrophysiological and magnetic resonance imaging work, we employed a single-subject design focused on individual data sets[1–4]. Twelve human participants (aged 20–34 years, all right-handed, six males; see Supplementary Table S1) were recruited to demonstrate reproducibility and generalization across subjects. Given that the datasets were in agreement and led

to similar conclusions with high statistical confidence, we did not consider it as scientifically necessary, economically and ecologically reasonable and ethically plausible to include more participants. All participants had normal or corrected to normal visual acuity, normal hearing ability and were well educated (eleven college degrees, one vocational degree). All participants gave written informed consent to participate. The study was approved under the reference 317/19-ek by the Ethics Committee at the Medical Faculty of the University of Leipzig, Germany (IRB00001750).

Sensory stimulation

Visual stimuli. In the visual experiment, stimuli were projected on a 34 × 25 cm screen inside the MRI bore with a resolution of 1,024 × 768 pixels. Participants viewed this screen through a mirror attached to the head coil. Stimuli were black dots on a middle gray background presented at the center of the display. To guide fixation and maintain attention, dots appeared within a radius of 0.75° (visual angle) of a diagonal cross of thin red lines covering the entire display throughout the experiment. To control for low-level non-numerical visual features, summed surface area (and thus luminance) was kept constant and dots were distributed homogeneously (preventing perceptual grouping) while randomly varying the positions of the dots. All stimuli were created using PsychoPy version 2021.2.3[21]. Other low-level features have been repeatedly and consistently shown not to influence numerosity-specific hemodynamic time courses[3,4].

Auditory stimuli. In the auditory experiment, participants listened to beep tone sequences via magnetic resonance compatible MR Confon headphones (http://www.mr-confon.de/) at an individually comfortable sound pressure level. To control for low-level non-numerical auditory features, the auditory frequency (pitch) within each sequence was randomly varied between 333, 359, 392 and 440 kHz. All stimuli were created using Audacity 3.0.2 (https://www.audacity.de/).

Stimulation procedure. Numerosities one through five were used to capture both the subitizing range (one through three) and the estimation range (four to five)[22,23]. They were first presented in ascending order in blocks changing every 4,200 ms. Within each block of the visual experiment, a dot array was shown six times for 300 ms each, alternating with 400 ms of gray background. Within each block of the auditory experiment, a tone sequence was played six times for 500 ms each, alternating with 200 ms of silence. Short stimulus durations were chosen to prevent participants from counting and to avoid adaptation effects. Short equal pauses were introduced to prevent participants from keeping numerosities in working memory. After the ascending numerosity sequence, twenty items (visual dots and audio beeps, respectively) were presented 24 times in the visual experiment and 8 times in the auditory experiment for 16.8 s including pauses. Next, numerosities one through five were presented again, but in descending order, followed by another block of twenty items. This stimulation cycle (1 – 2 – 3 – 4 – 5 – 20 – 5 – 4 – 3 – 2 – 1 – 20) was repeated four times in each run. The rationale for the long baseline period was that hemodynamic responses could return to baseline for numerosity-selective regions with small preferred numerosities and little neural responses to 20 items. The rationale for the ascending-descending numerosity presentation was to avoid adaptation effects, because each stimulus is preceded by both lower and higher numerosities in different blocks of presentation, thereby counterbalancing potential effects of previously presented numerosity[4]. Additionally, keeping this fixed and smooth order of presentation is known to result in more homogeneous topographic maps than a pseudorandom stimulation sequence. Alternating between ascending and descending presentation served to minimize neural adaptation by ensuring each stimulus was preceded by a stimulus inducing both a lower and higher response. Stimuli were presented with PsychoPy version 2021.2.3[21]. There were eight runs in each session (visual and auditory) for a total experiment duration of 41 min per session.

Target detection task. In 10% of all stimulus events, white instead of black dots were shown during the visual experiment and 1,000 kHz beeps were played instead of 333 to 440 kHz beeps during the auditory experiment. Participants were instructed to respond to these events by pressing a button. This simple target detection task was introduced to ensure participants were paying attention and following the task.

Data acquisition

Magnetic resonance imaging. T2*-weighted functional magnetic resonance imaging data were acquired with a 1-channel transmit / 32-channel receive coil (NOVA Medical, Wilmington MA, USA) on a 7 Tesla Magnetom TERRA scanner (Siemens Healthineers, Erlangen, Germany). We used a 2D gradient-echo echo-planar imaging (2D-EPI) sequence for acquiring 41 slices (thickness = 1.75 mm) with a field of view FOV = 192 mm x 192 mm and a matrix size = 110 x 110 resulting in an isotropic voxel size of 1.75 mm x 1.75 mm x 1.75 mm. The slice package had a transversal orientation, but was tilted towards coronal orientation in order to spare anterior frontal and temporal lobes where pronounced B0 and B1 inhomogeneities degrade image quality at 7T. To achieve a repetition time TR of 2.1 s, the acquisition was accelerated using GRAPPA with an iPAT factor of 3. An echo time TE of 24 ms, a flip angle FA of 70° and a receiver bandwidth of 1684 Hz/Px was used. 145 volumes were acquired per run resulting in a single run time of 05:04,50 min. For distortion correction, an EPI data set using the same parameters, but with the phase-encoding direction flipped from anterior-posterior to posterior-anterior was acquired. For anatomical referencing, a T1-weighted image was acquired using an MP2RAGE sequence[24] with the following parameters: TR = 5.000 ms, TE = 2.01 ms, $TI_{1/2}$ = 900 ms / 2.750 ms, $FA_{1/2}$ = 5° / 3°, isotropic voxel size of 0.7 mm x 0.7 mm x 0.7 mm.

Data preprocessing

Preprocessing was performed using FreeSurfer 7.4.1[25] and fMRIprep 23.1.4[26], based on Nipype 1.8.6[27]. The following description of fMRIPrep's preprocessing is based on a boilerplate distributed with the software covered by a 'no rights reserved' (CC0) license. For more details about the pipeline, see the section corresponding to each workflow in the fMRIPrep documentation.

Structural preprocessing. Before running fMRIPrep, we segmented the MP2RAGE UNI images using a custom workflow implemented in the python package fmritools 1.0.3 (https://doi.org/10.5281/zenodo.10573042). MP2RAGE image backgrounds were denoised as described in O'Brien et al. 2014[28] and the resulting images were bias-corrected with SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12). Then the first five Freesurfer cortical reconstruction steps were applied before computing a skullstrip mask based on the INV2 of the MP2RAGE image. This skullstrip mask was applied to the remaining Freesurfer cortical reconstruction pipeline. As a first step in fMRIPrep, the denoised and thresholded structural images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection[29], distributed with ANTs[30]. Images were then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (implemented in ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted images using FSL FAST[31]. In the case that two T1w images were available for a subject, a structural T1w reference map was computed after co-registration of the INU-corrected T1w images using mri_robust_template implemented in FreeSurfer 7.3.2[32]. Otherwise, the INU-corrected T1w image was used as T1w reference throughout the pipeline. Brain surfaces were reconstructed using the recon-all procedure of FreeSurfer 7.3.2[25]. Brain masks estimated before were refined based on Mindboggle's method to integrate ANTs-derived and FreeSurfer-derived segmentations of the cortical gray matter. This way, the best

results of the ANTs-derived and FreeSurfer-derived cortical gray-matter segmentations could be included in the final masks[33].

Functional preprocessing. Before functional data preprocessing, image quality was assessed using MRIQC and high-motion runs (framewise displacement >1.7mm for >10% of volumes) were excluded from further analysis[34]. Using fMRIPrep 23.1.4, for each of the up to 16 functional runs per subject (across both tasks), the following preprocessing steps were taken: First, a reference volume for estimating head-motion parameters and its skull-stripped version were generated. Head motion parameters (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using FSL MCFlirt[35]. B0-nonuniformity maps (fieldmaps) were generated from two (or more) EPI references with FSL topup[36]. Fieldmaps were then aligned with rigid registration to the target EPI reference run. Field coefficients were mapped onto the reference EPI using the transformation matrices generated when aligning the fieldmaps with the reference EPI. Next, functional runs were slice-time-corrected to 1.025 s (which is equivalent to 0.5 of the slice acquisition range of 0-2.05 s) using AFNI's 3dTshift[37].

Functional reference images were then co-registered to the T1w reference images by applying boundary-based registration with bbregister in FreeSurfer[38]. Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed functional images: framewise displacement (FD), DVARS (D referring to temporal derivative of time courses, VARS referring to root mean square variance over voxels) and three region-wise global mean signals extracted from within the CSF, the WM, and the whole-brain masks. FD was computed using two formulations based on the absolute sum of relative motions and the relative root mean square displacement between affines[35,38]. FD and DVARS were calculated for each functional run using their implementations in Nipype which follow the definitions by Power and colleagues[39].

These nuisance variables and head motion estimates were returned as a confounds file for each individual run. Specifically, the confound time series derived from head motion estimates and global signals were expanded by including temporal derivatives and quadratic terms for each[40]. Frames that exceeded a threshold of 1.7 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance time series were calculated by performing principal components analysis for the signals found within a thin band (crown) of voxels around the edge of the brain, as proposed by Patriat and colleagues[41].

The functional time-series were resampled onto the fsnative (individual participant) and fsaverage (standard space) surfaces following the FreeSurfer reconstruction nomenclature. All resamplings were performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction if available, and co-registrations to anatomical and FreeSurfer output spaces). Non-gridded (surface) resamplings were performed using mri_vol2surf in FreeSurfer. No spatial or temporal smoothing was applied[42].

Numerosity analysis

Numerosity population receptive field models. Task-based fMRI data were analyzed using a well-established forward modelling approach[3–5,7,43]. Each vertex is thought to represent the presented numerosities using a logarithmic Gaussian tuning curve characterized by two parameters: the preferred numerosity μ (the numerosity to which the population responds strongest) and a tuning width ω (the full width at half maximum (FWHM) width of the response function). Given presented numerosity x_t at continuous time t, the neuronal response at t is given by

$$z_t = exp\left[-\frac{1}{2}\left(\frac{\ln x_t - \mu_{log}}{\sigma_{log}}\right)^2\right]$$

where $\mu_{log} = \ln \mu$ and σ_{log} is the standard deviation of the response function in logarithmic space from which the FWHM tuning width can be calculated:

$$\omega = exp\left[\mu_{log} + \sqrt{2 \ln 2}\,\sigma_{log}\right] - exp\left[\mu_{log} - \sqrt{2 \ln 2}\,\sigma_{log}\right]$$

Given an assumed hemodynamic response function (HRF) $h(t)$, the hemodynamic response is given by the convolution of the neuronal response with the HRF:

$$s_t = z_t * h(t)$$

Finally, a linear relationship is assumed between this hypothesized hemodynamic response, governed by the neural tuning parameters μ and ω, and the preprocessed fMRI signal

$$y_t = \beta \cdot s_t + \beta_0 + \varepsilon_t$$

where β is a scaling factor of the hypothesized signal, β_0 is the baseline level of hemodynamic activity and ε_t are noise terms normally distributed with zero mean and unknown variance. Before estimation of tuning parameters μ, ω and β, preprocessed fMRI signals were standardized to units of percent signal change and 12 confound variables were regressed from standardized signals (six motion parameters; WM, CSF and global signal; three cosine regressors modelling temporal drifts). Then, signals were averaged across runs (but not cycles) and numerosity analysis was performed for a single averaged run.

Estimation of numerosity tuning parameters. The numerosity pRF model defines a likelihood function of the measured data y, given the tuning parameters μ and ω. Some combinations of tuning parameters make the measured signals more likely than others and the goal is to find the tuning which is most compatible with the measured fMRI data. To this end, we specified a large grid of plausible values: μ from 0.8 to 5.2 in steps of 0.05, σ_{log} from 0.05 to 30 in steps of 0.05 (such that

$0.12 \leq \omega \leq 34.17$ for the lowest presented numerosity $\mu = 1$ and $0.59 \leq \omega \leq 170.9$ for the highest presented numerosity $\mu = 5$; cf. equation above). For each possible combination μ and ω, the ensuing neuronal responses z were calculated using the sequence of presented numerosities x and the ensuing hemodynamic signals s were generated using the canonical HRF, as implemented in SPM. Then, β was estimated from the measured fMRI signal y using ordinary least squares, resulting in a single log-likelihood value for each combination of μ and ω:

$$LL(\mu, \omega) = \ln p\left(y|\mu, \omega, \hat{\beta}\right)$$

Finally, the optimal tuning parameters are those which maximize the log-likelihood function across all possible combinations:

$$\left(\hat{\mu}, \hat{\omega}\right) = argmax\, LL(\mu, \omega)$$

When not accounting for serial correlation in fMRI signals, this is equivalent to estimating tuning parameters by minimizing the residual sum of squares (RSS), as was done previously[4,7]. We did not account for serial correlations using e.g. restricted maximum likelihood and weighted least squares, since this did not improve tuning parameter estimates in a pre-analysis simulation study, even if serial correlations were present in the simulated signals. We also did not include hemodynamic derivatives to account for possible variation of HRF between subjects and vertices, because this was found to generate rather noisy numerosity-selective clusters on the cortical surface in a single-subject pilot analysis.

Statistical analyses

Extraction of parameter estimates. The numerosity pRF model returns four quantities for each vertex: estimated preferred numerosity $\hat{\mu}$, estimated FWHM tuning width $\hat{\omega}$, estimated scaling

factor $\hat{\beta}$ and the coefficient of determination R^2 ("R-squared"). Only vertices that exhibited a positive scaling factor ($\hat{\beta} > 0$), an estimated preferred numerosity inside the presented stimulus range ($1 \leq \hat{\mu} \leq 5$) and more than 20% of variance explained ($R^2 > 0.2$) were included in the following analyses. The rationale for excluding vertices with negative scaling factor was that the preferred numerosity could not be interpreted as the stimulus generating the maximum response. The rationale for excluding vertices with preferred numerosity outside the stimulus range is that their tuning functions will be monotonically increasing or decreasing throughout the stimulus range and thus not display actual tuning.

Inferential statistics for variance explained. To assess statistical significance of variance explanation by the tuning model, we performed an F-test of the model including the numerosity regressor against a null model including only the baseline regressor, accounting for additional parameters in the degrees of freedom:

$$F = \frac{\left(RSS_0 - RSS\right)/(p-1)}{RSS/(n-p)}$$

where $n = 145$ is the number of scans per run and $p = 4$ is the number of free parameters in the numerosity pRF model ($\beta_0, \beta, \mu, \omega$). Under the null hypothesis H_0: $R^2 = 0$, this test statistic is following an F-distribution with $p - 1$ numerator and $n - p$ denominator degrees of freedom. The F-statistic can also be expressed in terms of variance explained:

$$F = \frac{R^2/(p-1)}{\left(1-R^2\right)/(n-p)}$$

With the n and p given above, this implies that a variance explained of $R^2 = 0.2$ corresponds to an uncorrected p-value of $p = 6.4 \cdot 10^{-7}$.

Calculation of cortical surface area. Alongside with estimated tuning parameters, the coordinates on the pial surface in native subject space were extracted for each supra-threshold vertex. We here chose pia mater rather than white matter or midthickness surface coordinates, as these most precisely correspond to actual cortical surface extents in native subject space. After filtering numerosity maps with the criteria mentioned above, AFNI's SurfClust function was used to form clusters of vertices which an edge distance of at most 1 on the cortical surface (i.e. vertices needed to be directly connected to belong to the same cluster). For each cluster, the total area on the cortical surface was calculated as the summed area of all triangles for which all three nodes belonged to the cluster. For analyses of cortical geometry, only clusters having a minimum cortical surface area (visual: $A_{min} = 50\ mm^2$; auditory: $A_{min} = 25\ mm^2$) were retained.

Identification of numerotopic maps. In order to investigate topographic organization in larger numerosity-selective clusters, supra-threshold clusters were assigned one of the previously reported (Harvey & Dumoulin, 2017) six numerotopic maps for the visual modality (temporo-occipital, NTO; parieto-occipital, NPO; in parietal cortex, NPC1/2/3; in prefrontal cortex, NF; N = numerosity) by being in less than some distance ($d_{max} = 1.7\ mm$) from previously reported center coordinates[3] with at least one vertex in standard space. For the auditory modality, two numerotopic maps were identified from participant count maps as being located in primary auditory cortex (temporal, NaT) and pre-motor regions (frontal, NaF; Na = auditory numerosity)[5] and clusters were assigned to these auditory numerosity maps in the same way.

Visualization of numerosity selectivity. The estimated tuning parameters $\hat{\mu}$ and $\hat{\omega}$ describe a logarithmic Gaussian tuning function and an expected time course of hemodynamic activity in a single cycle of numerosity presentation. These are plotted for a low-numerosity vertex ($1 \leq \hat{\mu} \leq 2$) and a high-numerosity vertex (visual: $4 \leq \hat{\mu} \leq 5$; auditory: $3 \leq \hat{\mu} \leq 5$) from exemplary single subjects in each

sensory modality (see Figure 2 and Supplementary Figures S1/S2). In addition to this, vertex-wise variance explained and vertex-wise preferred numerosity are visualized in native subject space for all vertices passing the above filtering criteria (see Figures 3A and 3C and Supplementary Figures S3/S4). For each vertex in standard space, the number of subjects showing numerosity selectivity according to these criteria are shown as participant count maps (see Figure 3B).

Statistical analysis of cortical geometry. For each each numerotopic map (six visual, two auditory; see above) in each hemisphere (left, right), we report the average cortical surface area (see Figure 4), as well as the represented range of preferred numerosities (see Figure 6c,d and Supplementary Figure S8). Additionally, we investigated whether there is a topographic organization within those maps, i.e. progression of preferred numerosity along the cortical surface. Previously, this was done by manually delineating start, end and borders of numerotopic maps, establishing distance along the so-defined principal axis of a numerotopic map and calculating average preferred numerosity as a function of binned cortical distance[3,4]. As manual delineation was not feasible for our amount of data (twelve participants, two sessions, six/two maps) and might lead to non-reproducible results, we instead predicted preferred numerosity based on pial x-, y- and z-coordinates in standard space (see Figure 6a,b and Supplementary Figure S7). If there is a significant correlation between actual numerosities and predicted numerosities, this indicates that preferred numerosity changes along spatial dimensions and is thus topographically organized.

Statistical analysis of tuning parameters. As further correlational analyses, attempting to replicate previously reported effects[3,4], we analyzed relationships of preferred numerosity with cortical surface area (see Figure 5a,b and Supplementary Figure S5) and FWHM tuning width (see Figure 5c,d and Supplementary Figure S6) for each hemisphere. Additionally, we ran a linear mixed-effects model

with either cortical surface area or FWHM tuning width as the dependent variable, preferred numerosity and hemisphere as independent variables and subject as a random effect (see Table 1).

References

1. Kutter, E. F., Bostroem, J., Elger, C. E., Mormann, F., & Nieder, A. Single neurons in the human brain encode numbers. *Neuron* **100**, 753–761.e4 (2018).
2. Eisenkolb, V. M. *et al.* Human acute microelectrode array recordings with broad cortical access, single-unit resolution, and parallel behavioral monitoring. *Cell Rep.* **42**, 112467 (2023).
3. Harvey, B. M. & Dumoulin, S. O. A network of topographic numerosity maps in human association cortex. *Nat. Hum. Behav.* **1**, 1–9 (2017).
4. Harvey, B. M., Klein, B. P., Petridou, N. & Dumoulin, S. O. Topographic representation of numerosity in the human parietal cortex. *Science* **341**, 1123–1126 (2013).
5. Hofstetter, S., Cai, Y., Harvey, B. M. & Dumoulin, S. O. Topographic maps representing haptic numerosity reveals distinct sensory representations in supramodal networks. *Nat. Commun.* **12**, 221 (2021).
6. Nieder, A. Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11860–11865 (2012).
7. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *Neuroimage* vol. 39 647–660 (2008).
8. Harvey, B. M. & Dumoulin, S. O. The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *J. Neurosci.* **31**, 13604–13612 (2011).

9. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
10. Arrighi, R., Togoli, I. & Burr, D. C. A generalized sense of number. *Proc. Biol. Sci.* **281**, 20141791 (2014).
11. Tsouli, A., Harvey, B. M., Hofstetter, S., Cai, Y., van der Smagt, M. J., Te Pas, S. F. & Dumoulin, S. O. The role of neural tuning in quantity perception. *Trends Cogn. Sci.* **26**, 11–24 (2022).
12. van Ackooij, M., Paul, J. M., van der Zwaag, W., van der Stoep, N. & Harvey, B. M. Auditory timing-tuned neural responses in the human auditory cortices. *Neuroimage* **258**, 119366 (2022).
13. Xu, F. & Spelke, E. S. Large number discrimination in 6-month-old infants. *Cognition* **74**, B1–B11 (2000).
14. Halberda, J., Mazzocco, M. M., & Feigenson, L. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* **455**, 665–668 (2008).
15. Rose, G. J. The numerical abilities of anurans and their neural correlates: insights from neuroethological studies of acoustic communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20160512 (2017).
16. Skorupski, P., MaBouDi, H., Dona, H. S. G. & Chittka, L. Counting insects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20160513 (2018).
17. Ditz, H. M. & Nieder, A. Neurons selective to the number of visual items in the corvid songbird endbrain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, (2015).
18. Nieder, A., Diester, I. & Tudusciuc, O. Temporal and spatial enumeration processes in the primate parietal cortex. *Science* **313**, 1431–1435 (2006).
19. Sulser, R. B., Patterson, B. D., Urban, D. J., Neander, A. I. & Luo, Z.-X. Evolution of inner ear neuroanatomy of bats and implications for echolocation. *Nature* **602**, 449–454 (2022).

20. Bellono, N. W., Leitch, D. B. & Julius, D. Molecular tuning of electroreception in sharks and skates. *Nature* **558**, 122–126 (2018).

21. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).

22. Starkey, P. & Cooper, R. G., Jr. Perception of numbers by human infants. *Science* **210**, 1033–1035 (1980).

23. Burr, D. C., Turi, M. & Anobile, G. Subitizing but not estimation of numerosity requires attentional resources. *J. Vis.* **10**, 20 (2010).

24. Marques, J. P. *et al.* MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage* **49**, 1271–1281 (2010).

25. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

26. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).

27. Gorgolewski, K. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* **5**, 13 (2011).

28. O'Brien, K. R. *et al.* Robust T1-weighted structural brain imaging and morphometry at 7T using MP2RAGE. *PLoS ONE* **9**, 1–7 (2014).

29. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).

30. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).

31. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).

32. Reuter, M., Rosas, H. D. & Fischl, B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage* **53**, 1181–1196 (2010).

33. Klein, A. *et al.* Mindboggling morphometry of human brains. *PLoS Comput. Biol.* **13**, e1005350 (2017).

34. Esteban, O. *et al.* MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* **12**, e0184661 (2017).

35. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).

36. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).

37. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed.* **10**, 171–178 (1997).

38. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).

39. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).

40. Satterthwaite, T. D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).

41. Patriat, R., Reynolds, R. C. & Birn, R. M. An improved model of motion-related signal changes in fMRI. *Neuroimage* **144**, 74–82 (2017).

42. Hoffmann, M. B. *et al.* Plasticity and stability of the visual system in human achiasma. *Neuron* **75**, 393–401 (2012).

43. Cai, Y. *et al.* Topographic numerosity maps cover subitizing and estimation ranges. *Nat. Commun.* **12**, 1–10 (2021).

Data availability

After publication, the data for this study will be made available through a public link on the Edmond repository (https://edmond.mpg.de/). During the peer review phase, data are available from the corresponding author upon request.

Code availability

All data analysis was performed with Spyder 5.4.3 using Python 3.10 and various third-party packages (numpy 1.24.3, scipy 1.10.1, pandas 2.1.1, statsmodels 0.14.0, nibabel 5.1.0), except for edge-based surface clustering which was run using AFNI's SurfClust. Brain surfaces were visualized with surfplot 0.2.0 and other plots were generated with matplotlib 3.7.1. All code for data analysis is available from GitHub (https://github.com/SkeideLab/EMPRISE).

Acknowledgements

We thank our lab manager Micha Vollmann for recruiting participants, coordinating logistics, handling forms, and organizing the measurements. This work was supported by the German Research Foundation (DFG Research Grant 433715509 and DFG Heisenberg Program Grant 433758790 awarded to M.A.S.) and the Jacobs Foundation (Research Fellowship awarded to M.A.S.).

Author contributions

M.A.S. conceived the study. A.K. and R.T. performed experiments. J.S. and A.K. analyzed data. J.S. visualized the results with feedback from A.K. and M.A.S. J.S. and M.A.S. wrote the manuscript with feedback from all other authors. M.A.S. supervised the study.

Ethics declarations

The authors declare no competing interests.