Attention modulates the geometry of auditory representations

Pedersini, CA[1] - Muukkonen, I [1,2] - Wikman, P [1,3,4,5]

[1] Department of Psychology, PO Box 21 FI-00014 University of Helsinki, Finland

[2] Department of Brain and Cognition, PO box 03711, KU Leuven, Leuven, Belgium

[3] Advanced Magnetic Imaging Centre, PO Box 11000 FI-00076, Aalto University, Espoo, Finland

[4] Neuroscience Center, HiLife, PO Box 21 FI-00014 University of Helsinki, Finland

[5] Helsinki Collegium For Advanced Studies, PO Box 21 FI-00014 University of Helsinki, Finland

Corresponding author: Caterina Annalaura Pedersini (cpedersini86@gmail.com)

Acknowledgment: This work was supported by The Research Council of Finland (grant number #1348353). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We want to thank M. Salmikivi, O. Varis, W. Vikatmaa, S. Rossow, and L. Lehtimäki for helping with data acquisition, as well as Ville Laaksonen and Jaakko Kauramäki for help with the study design.

Data Availability Statement: All preprocessed behavioral and fMRI data, anonymized to protect participant privacy, will be deposited on the Open Science Framework (OSF). The corresponding analysis code is available at: https://github.com/neural-review2025/SoundScapes.

Abstract

Selective attention enables the brain to prioritize behaviorally relevant information in complex environments, yet the underlying neural mechanisms remain poorly understood. Here, we combined functional magnetic resonance imaging (fMRI) with representational geometry analyses to investigate how selective attention reshapes neural population codes in naturalistic soundscapes. Participants attended to individual sound objects presented either among distractors from the same category (3OW) or from different categories (3OA). We observed that attention selectively attracted neural population responses toward the object-alone template. This effect, evident in the low-dimensional PCA space, likely reflects a reorganization of the underlying high-dimensional representational geometry. This geometric effect was modulated by the surrounding soundscape: in 3OW, where all sounds belonged to the same category, attentional modulation was minimal, whereas in 3OA, attended representations were drawn significantly closer to their template relative to distractors. Category-specific patterns were also evident: for speech, attentional modulation extended beyond the auditory cortex to left-lateralized language regions, whereas for instruments and animals, effects were largely confined to bilateral auditory areas, including core, belt regions and the superior temporal sulcus. Finally, the attraction effect observed was behaviorally meaningful: higher subjective ratings of attentional engagement predicted smaller distances between attended and object-alone representations.

Taken together, these findings demonstrate that selective attention reshapes neural population geometry in a context- and category-dependent manner, with geometric changes closely reflecting behavior.

Introduction

Imagine standing in a park: a friend speaks beside you, children play in the distance, a bird chirps overhead, and a dog barks across the street. Everyday hearing means navigating this constant mixture of overlapping sounds. A dominant view has been that the the brain transforms raw acoustic input into increasingly abstract representations along the auditory ventral pathway, from simple spectral features (e.g., frequency) in primary auditory cortex (AC), to more complex patterns in secondary regions, and to object-level and categorical representations in anteroventral AC and anterior temporal lobe (ATL) (Rauschecker, 2018; Theunissen & Elie, 2014). Although this description suggests that objects are constructed step by step along a hierarchy, the reality may be more complex as top-down cognitive processes can shape object formation at early stages, and local groupings can be reweighted by broader scene-level context (Shinn-Cunningham, 2008).

Selective attention is a top-down cognitive process that dynamically reshapes population-level representations to prioritize behaviorally relevant information. In the visual system, multivariate analyses have shown that attention sharpens feature tuning and warps representational spaces, including semantic dimensions, to emphasize task-relevant categories (Çukur et al., 2013; Engel et al., 2016; Ibos & Freedman, 2016; Jehee et al., 2011; Kay et al., 2015; Nastase et al., 2017; Sprague & Serences, 2013). These effects reflect a flexible reconfiguration of neural population geometry rather than simple gain modulation (Kriegeskorte & Diedrichsen, 2019). Parallel mechanisms have been observed in the auditory cortex (AC), where attention modulates gain (Shinn-Cunningham, 2008; Zatorre et al., 1999), or tuning of neurons representing task-relevant sounds (Da Costa et al., 2013; Kauramäki et al., 2007; Lage-Castellanos et al., 2022; Moerel et al., 2013; Schäfer et al., 2018), thereby supporting source separation (Atiani et al., 2009; Caruso et al., 2018; J. Fritz et al., 2003; J. B. Fritz et al., 2007; Lakatos et al., 2013). In the context of the "cocktail-party" effect, where multiple speakers are talking at the same time, recent studies showed that AC represents more accurately attended than ignored speech (Ding & Simon, 2012; Mesgarani & Chang, 2012; Wikman et al., 2024). While such models provide important insight into attentional modulation, they are typically studied under controlled conditions (such as the *dichotic listening task* or the *cocktail-party-like setting*) that cannot fully capture the complexity of naturalistic soundscapes, where multiple overlapping sounds compete for representation. In these natural auditory scenes, it remains debated whether selective attention primarily operates at the object- or feature-level, or whether both are adaptively engaged depending on context and task demands (S. Shamma et al., 2010; S. A. Shamma et al., 2011; Wikman et al., 2021).

In the last few years, our group has examined how top-down information shapes neural representations of naturalistic auditory scenes with overlapping sounds (Wikman et al., 2022, 2024). In the most recent study (Wikman et al., 2025), the authors examined selective attention in naturalistic auditory scenes containing overlapping sounds from the same or different categories. The main findings highlighted that attention enhances the neural representation of the attended object's identity in the auditory cortex, but that this enhancement is flexible, adapting to the structure of the surrounding scene. In this study, the authors applied searchlight multivoxel pattern analysis (MVPA), assessing whether local neural patterns elicited by multi-object stimuli resembled the patterns elicited when each object was presented in isolation. While informative about local pattern modulations, such analyses do not capture the population-level structure of responses (Ebitz & Hayden, 2021), nor do they reveal how selective attention reshapes the geometry of the population code. This limitation

arises because decoding analysis primarily relies on a single geometric property of the neural manifold—the distance between manifold centers (Sorscher et al., 2022; St-Yves et al., 2025)—and is largely insensitive to other geometric transformations such as rotations of population axes or redistribution of variance, which can occur without affecting linear separability.

Despite heterogeneous activity at the level of single neurons, empirical evidence suggests that population activity often lies on a low-dimensional neural manifold, reflecting latent dynamics that constrain responses and shape behavior (Perich et al., 2025). In this reduced space, the dominant axes of covariation across voxels (population dimensions) become explicit, allowing us to quantify how attention reshapes representational geometry. In the current study, we apply a geometry-based pipeline to the same neuroimaging data as in Wikman et al. (2025), to characterize attentional effects on population codes. As a first step, we used hyperalignment based on Generalized Procrustes Analysis (GPA) (Gower, 1975; Guntupalli et al., 2016, 2018; Haxby et al., 2011, 2020) to transform individual activation patterns from anatomical space into a shared high-dimensional common space. GPA derives orthogonal transformation matrices (rigid rotations and reflections) that minimize distances between functional patterns while preserving the representational geometry of each participant's data. This procedure functionally aligns corresponding dimensions across participants. We then applied principal component analysis (PCA) to project the aligned high-dimensional activation patterns into a low-dimensional population space. Within this space, we quantified attentional modulations using two complementary metrics: Euclidean distance, to capture changes in the spacing and clustering of object representations, and principal angles, to assess rotations of population axes toward task-relevant dimensions. This geometry-based pipeline has proven to be a powerful framework to study perceptual and cognitive processes in both non-human primates and humans (M/EEG, fMRI) (Barbosa et al., 2025; Cheng et al., 2025; Chou et al., 2025; Gao et al., 2021; Li & Curtis, 2023; Panichello & Buschman, 2021; Santo-Angles et al., 2025; Xie et al., 2022). Extending this approach, in this study we apply a geometry-based pipeline to selective auditory attention in naturalistic auditory scenes, to provide a novel window into how attention reshapes neural population geometry under ecologically realistic conditions.

Building on this approach, the present study addresses three main questions: (1) how auditory representations progress along the ventral pathway, (2) how selective attention reshapes the geometry of neural population spaces and how this geometry is further modulated by the surrounding soundscape under the same attentional condition, and (3) whether neural geometry relates to behavioral performance. To do this, we asked participants to pay attention to a specific sound object (orientation of attention), while modulating the soundscape through the presentation of distractors of the same or different category. We hypothesize that each auditory object is represented as a distinct point in population space, acting as an attractor within a low-dimensional manifold embedded in the higher-dimensional neural activity (Barack & Krakauer, 2021). Attention may dynamically reshape this manifold, either by altering the spacing of object representations (quantified with Euclidean distance) or by rotating the dominant axes of population activity toward task-relevant features (quantified with principal angle). By investigating these questions, we aim to reveal how top-down and scene-level factors interact to organize auditory representations at the population level. Understanding these mechanisms can shed light on how the brain navigates complex auditory environments, such as a park where overlapping voices, birdsong, and other sounds must be segregated and prioritized for meaningful perception.

Method

Participants. Twenty healthy right-handed monolingually native Finnish speaking participants underwent an fMRI session (8 females, age range 19–32, mean age 22.7 years, standard deviation 3.68). All participants were students at the University of Helsinki and Aalto University, had normal hearing, normal or corrected-to-normal vision and no history of developmental or neurological disorders. We assessed handedness by using the Edinburgh Handedness Inventory (Oldfield, 1971) and acquired written consent from the participants before the study.

Stimuli. The experiment included auditory and visual stimuli, presented with Presentation 24.0 software (Neurobehavioral systems, Berkeley, California, USA). The auditory stimuli were presented binaurally through earphones (Sensimetrics model S14; Sensimetrics, Malden, Massachusetts, USA) and auditory volume was adjusted individually for each participant at the beginning of the session at ca. 85 dB. Visual stimuli were projected onto a mirror attached to the head coil. The participant answered the questions with an fMRI-compatible response pad with either the index-, middle- or ring finger of their right hand. The auditory stimuli comprised sound clips from three object categories: speech, animal and instrument sounds, each with six subcategories. The speech subcategories included vocalizations from two adult females, two adult males, and two children (one girl and one boy). Animal sounds were sampled from huskies, cats, whales, monkeys, birds, and seals, while instrument sounds came from a baritone guitar, a 12-string guitar, a saxophone, a trumpet, an organ, and a bebop organ. Each subcategory contained eight exemplars, resulting in 144 distinct auditory clips (3 categories × 6 subcategories × 8 exemplars). Sound durations ranged from 4 to 7 seconds, with an average of 5.3 seconds. For each exemplar sound, we extracted six acoustic features with MIRtoolbox (v 1.8.1.) running in Matlab (2023b), obtaining one value for each feature and clip. Pitch was computed with *mirpitch* using the default autocorrelation method, selecting the dominant periodic component within 75–10,000 Hz. The spectral centroid (*mircentroid*) was defined as the mean frequency of the stimulus spectrum. Harmonics-to-noise ratio (HNR) was derived as the ratio of the dominant periodic component (maximum autocorrelation) to the aperiodic signal component (Boersma & Weenink, 2001; (Leaver & Rauschecker, 2010). Amplitude modulation standard deviation (AMSD) was calculated as the standard deviation of the rectified waveform after resampling to 60 Hz, quantifying amplitude fluctuations over time. Frequency content standard deviation (FCSD) was defined as the standard deviation of spectral power across frequencies. Spectral entropy (*mirentropy*) was computed as the relative Shannon entropy of the stimulus spectrum. Visual stimuli were created from all auditory clips used in the study, including both training items and non-targets from match-to-target tasks, totaling 214 sounds. For each sound, a scrolling waveform display was generated using MATLAB (R2020a, MathWorks Inc., Natick, MA, USA), with a one-second window shown in blue on a white background. To ensure the video always exceeded the auditory clip duration during synchronized presentation, the waveform was duplicated and followed by one second of silence. Video frames were cropped to 876 × 656 pixels and compiled into silent 25 fps videos using ffmpeg, encoded with the MPEG4 codec (2000 kbps bitrate) in AVI format for compatibility. After manual screening to remove visually redundant items, 180 visually distinct control videos were selected and presented independently of the auditory stimuli. These stimuli, along with the video generation procedure, are described in greater detail in our previous work (Wikman et al., 2025).

Experimental session. Each participant completed three experiments within the same session, all of which followed an identical trial structure consisting of four phases (Fig. 1). First, an instruction was presented for 1 second, indicating whether the participant should perform an auditory or visual task. In auditory trials, the instruction also specified the sound object to attend to (e.g., "attend to the man"). This was followed by the stimulus presentation phase, during which the designated auditory or visual object was presented. In experiments with overlapping sounds presentation, the attended sound object was always presented 250-500ms before the distractors. Simultaneously, a visual video clip (randomly sampled without replacement) was shown for the same duration as the auditory stimulus (4–7 seconds). After the stimulus presentation, participants provided a subjective rating indicating how well they were able to attend to the designated object. Responses were given within a 1.5 second window by pressing a button to select one of three options: 1 (less than 33% of the time), 2 (33–66% of the time), or 3 (more than 66% of the time). In 25% of the auditory attention trials, a match-to-sample task followed, driven by the change to green of the visual display background. In this task, participants heard a 2-second sound sample and indicated whether it matched the attended sound object. The sample was either an exemplar of the attended sound object (target trial, 50% probability) or a different exemplar of the same object category that had not been used as the attended stimulus (nontarget trial). In 100% of the visual attention trials, the participant was presented with a short sample (2 sec.) of either the video clip they had been attending to (target trial, 50% probability) or another video clip (nontarget trial). The intertrial interval was randomized to last from 201 to 5500 msec and the response time window was 1.5 seconds. After answering, the participant received feedback on their performance (0.5 sec.; correct/incorrect). See figure 1 for a representation of the trial structure.

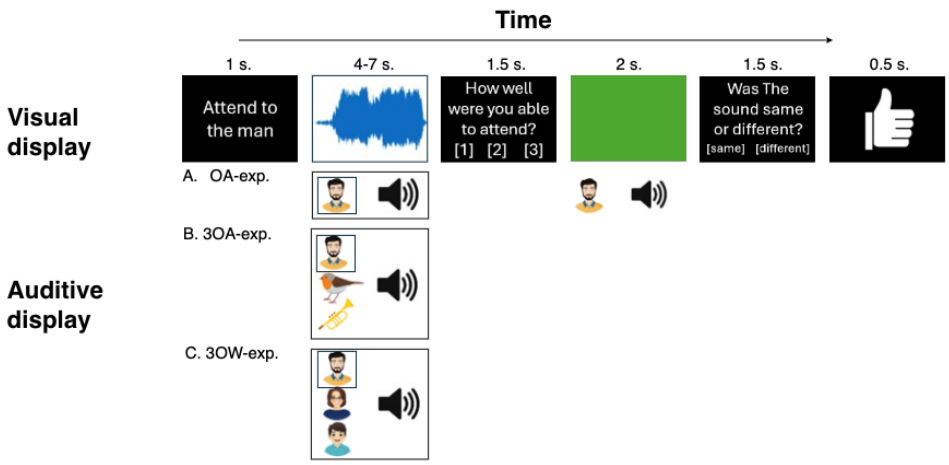

Fig. 1. Behavioral paradigm in the three experiments (OA, 3OA, 3OW). Each trial began with an instruction to attend to either an auditory object or the visual display, followed by stimulus presentation (one or three concurrent sounds with a scrolling waveform). After each trial, participants rated their attentional success, and on some trials performed a match-to-sample task with feedback.

In the **Object alone (OA) experiment,** participants were instructed to attend to the single auditory object presented on each trial. Although visual videos were displayed concurrently, they were never

task-relevant, and participants were instructed to ignore them. Each run consisted of 36 trials, including two repetitions of the same sound object within the run. The experiment comprised four runs in total, allowing for the presentation of all 144 sound object exemplars once, 48 for each category. Trial order was randomized for each participant. For each sound object type, one match-to-sample task featured a sample that matched the attended sound (target trial), and one featured a non-matching exemplar of the same object type (nontarget trial). As the match-to-sample task was presented in 25% of trials per run, each run included nine auditory match-to-sample trials.

In the **3OA experiment**, each auditory scene consisted of three sound objects belonging to different categories (e.g., boy, husky, and trumpet). Within each run, the participant performed two tasks: the *attend sound object task*, including 75% of the trials, and *the visual control task*, including the remaining 25% of the trials. In the former, participants had to attend to a pre-specified sound object, while the other two objects acted as distractors and had to be actively blocked out. The combinations of sound stimuli were randomized for each participant and systematically rotated to ensure balanced pairings. Each sound object type appeared twice as attended across two different scenes, using a different exemplar each time. This resulted in 36 auditory attention trials per run (12 scene types × 3 attention conditions), with 108 unique exemplars presented once per run. In the latter, participants were instructed to fully ignore all auditory stimulation and attend to the visual video. As in the auditory task, the same 12 scene types were used for the visual control task, but with different exemplars, bringing the total to 48 trials per run (around 8 minutes). While the scenes remained the same across runs for the same participant, their order was randomized.

In the **3OW experiment**, each auditory scene consisted of three sound objects belonging to the same category (e.g., Male1, Boy, Female2 or Bird, Husky, Whale). As in the 3OA experiment, each sound object type was presented in two different auditory scenes. For each participant, the sound object types within a category were first randomly shuffled, then grouped so that each appeared in exactly two scenes. This resulted in four unique scene subsets per category. Each sound object could serve in three roles—attended, distractor 1, or distractor 2—across the four scene types and three categories, yielding 36 auditory attention trials per run (3 attention conditions × 4 scene types × 3 categories). As in the 3OA experiment, the same 12 scene types were also used for the visual control task, but with different exemplars. See figure 1 (bottom section) and our previous work (Wikman et al., 2025), providing more details on experiments and stimuli used.

Each participant completed a training session outside the scanner on the same day as the fMRI acquisition, during which each experiment was practiced for a total of two runs. The fMRI session always began with the OA experiment, while the order of the remaining tasks was counterbalanced across participants, resulting in a total of 12 runs per subject. Short breaks were provided between runs, during which participants could rest and communicate with the researchers if needed.

fMRI data acquisition. Data was collected from a 3T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil, with 2 coils removed to make the stimuli visually accessible. We used T2* echo planar imaging which collected 45 continuous oblique axial slices (TR 1050 msec, TE 30 msec., flip angle 75°, field of view 21 cm, slice thickness 2.7 mm, isotropic in-plane resolution of 2.7 mm). Each session consisted of 12 runs. For the first 4 runs (OA experiment) we collected approximately 360 volumes per run, whereas for the last 8 runs (3OA/3OW-experiments) we collected approximately 570 volumes per run. At the end of each session, we collected a high-resolution T1 anatomical image (TE 3.3 msec., TR 2530 msec., voxel matrix 256×256, isotropic in-plane resolution of 1 mm). Each session lasted approximately 100 minutes.

Preprocessing and analysis of fMRI data. Functional MRI data were preprocessed using FSL Version 6.00, part of FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl), including using MCFLIRT (Jenkinson et al., 2002), interleaved slice-timing correction, non-brain removal with BET (Smith, 2002), and high-pass temporal filtering (with a cut-off of 100 Hz). Functional images were registered to the individual structural (T1-weighted) images using FLIRT (Jenkinson et al., 2002; Jenkinson & Smith, 2001). All further fMRI analyses were conducted in surface space. Functional data were projected to the FreeSurfer average cortical surface (fsaverage7) using *mri_vol2surf* (Fischl, 2012), and all analyses were performed separately for each hemisphere on the surface.

For the first-level analysis, we used the preprocessed functional images as input and included nuisance regressors to account for motion and physiological noise. Specifically, we included the six standard motion parameters (three translations: *trans_x, trans_y, trans_z*; and three rotations: *rot_x, rot_y, rot_z*), framewise displacement (FD), and the first five components from the anatomical CompCor method (aCompCor) derived from signals in white matter and cerebrospinal fluid. These confounds were extracted from the output of fMRIPrep 20.2.5 (Esteban et al., 2019)and entered as additional nuisance regressors in the general linear model using FSL FEAT.

In the first-level analyses, we fitted a general linear model (GLM) to the time series data of each vertex in each run, using FEAT at the subject level. All GLMs included separate boxcar regressors, defined with the onset at the beginning of the first sound and a duration corresponding to the entire auditory stimulation (all three sounds). For each of the three sound categories (speech, instrument, and animal), six distinct objects were used, and each was repeated twice within a run. This resulted in a total of 36 regressors (3 categories × 6 objects × 2 repetitions). Nuisance regressors were included to account for all timepoints associated with the match-to-sample task. In addition to the attended sound object regressors, the GLM included 12 regressors for the visual control task, modeling each visual condition separately. All further fMRI analyses were conducted at the level of individual ROIs, using the Glasser Atlas (Glasser et al., 2016). We first performed the main geometric analysis of neural subspaces at the whole-brain level, across all 360 regions, to localize the overall effect of attention on brain manifold geometry. Subsequent analyses focused on a subset of 35 regions per hemisphere, covering auditory, temporal, parietal, and frontal areas (see Fig. 5, E panel for a spatial map of these regions).

Construction and Alignment of Representational Spaces. To assess attention-driven differences in brain space geometry, we extracted one input matrix per experiment and condition, for each ROI. For each subject, the COPEs (contrasts of parameter estimates), which represent the estimated effect size at each vertex for a given contrast, were first averaged across repetitions of the same sound object within each run and then across runs. In the attended condition, each sound object was presented twice, while in the distractor condition, each object appeared four times. To balance the number of samples across conditions, we randomly selected two COPEs without replacement from the distractor presentations for each object in each run. As a result, for each subject, ROI and hemisphere, we obtained five data matrices. Each matrix contained 18 rows (one per unique sound object; 6 objects per category across 3 categories) and a number of columns equal to the number of vertices in the ROI. The five matrices corresponded to the following conditions: experiment 3OW attended sound objects, experiment 3OA attended sound objects, experiment object-alone, experiment 3OW distractor sound objects, and experiment 3OA distractor sound objects. In addition to these matrices extracted at object-level, for the OA-experiment, we constructed the input matrices so that each row

corresponded to a specific exemplar, resulting in a single matrix with 144 rows, as each exemplar was presented once across different runs.

These input matrices were then hyperaligned across subjects, to align representational spaces across individuals (see figure 2A for a graphical representation). The data were first mean centered by subtracting the mean activation across conditions for each vertex. To align the representational spaces across subjects, we applied the *Generalized Procrustes Analysis* (Gower, 1975; Haxby et al., 2011), without scaling. This procedure returns transformation matrices that optimally align each subject's ROI-level representational space to a common model space. To reduce the dimensionality of the representational brain space and to quantify the structure of condition-specific response patterns for each subject and ROI, we applied Principal Component Analysis (PCA; *pca.m* function in matlab), on the aligned data from each experiment separately (see figure 2B for a graphical representation). We extracted the PCA scores and computed the cumulative explained variance across components. Once the PCA scores were extracted, we applied *Procrustes alignment* without scaling to align the representational spaces of experiments 3OW and 3OA (attended and distractors) to that of experiment OA (reference). Scaling was omitted in both alignment steps to preserve the intrinsic geometry of the neural representational space, ensuring that distances between conditions reflect meaningful patterns of neural activity.

Comparison of Acoustic and Neural Manifolds. To examine the spatial distribution of similarity between the neural and the acoustic feature manifold, we implemented the following analysis pipeline. First, we derived the acoustic manifold by performing PCA on the six standardized acoustic features extracted from each exemplar sound. This yielded principal component scores representing the acoustic space. We then aligned these acoustic scores (3 PCs) to the neural scores (hyperaligned across subjects) obtained from the object-alone experiment at exemplar level (144 exemplars), using Procrustes alignment with scaling. In this case, scaling was applied to account for differences in units and overall magnitude between the two spaces. From the aligned matrices, we extracted the Procrustes Distance - which represents the sum of squared differences between two sets of points after optimal scaling, rotation, and translation - between the acoustic and the neural manifold, separately for each ROI (35) and hemisphere (Barbosa et al., 2025; Ilomäki et al., 2025; Schönemann, 1966). Finally, we created a null distribution of distances and tested in which brain regions the neural representation geometry is closer to the acoustic feature manifold than would be expected by chance ($p<0.05$, 10000 permutations). To visualize the spatial distribution of the significant results, we plotted the significant Procrusted Distances on a lateral brain surface.

Geometric analysis of neural subspaces. To quantify differences between conditions in the geometry of the low-dimensional representation for each category, we extracted two metrics from the data projected to the low-dimensional PCs: Euclidean Distance (ED) and Principal Angle (PA).
ED was calculated for each object belonging to the same category, between the data projected on the first PCs (scores) of experiment OA (used as a reference) and the aligned PCA representation (scores) of experiment 3OW or 3OA. These object-wise Euclidean distances were then averaged across objects of the same category, to yield a mean distance representing the overall distance of that category's representations (see Fig. 2 for a graphical representation of the pipeline). To assess whether the observed differences across conditions were driven by within-category representational changes rather than global shifts in category position, we performed a centroid alignment. For each subject

and sound category, neural patterns from the comparison experiment were aligned to the reference experiment (Experiment OA) by computing the vector connecting the category centroids and shifting all object-level representations, thereby preserving relative exemplar geometry within each category. ED for each object representations were then computed using either the centroid-aligned or unaligned data. In the main text, we present results using the first three PCs. However, we also performed the same analysis optimizing the number of PCs included by identifying the number of components required to explain at least 70% of the variance in each dataset. To ensure a consistent representation across experiments, we then selected the maximum number of components needed across all matrices. This approach ensured that the chosen dimensionality captured sufficient variance in every experiment's data. Results from this variance-optimized analysis are reported in the supplementary material and are largely consistent with the main findings.

PA was extracted to quantify how attention may affect the orientation and the alignment of neural manifolds (Gallego et al., 2018; Panichello & Buschman, 2021; Xie et al., 2022). Specifically, we adapted the pipeline described by Panichello et al. (2021), to calculate the principal angle between category-specific subspaces, at subject-level. In brief, we applied a second PCA separately for each stimulus category and experiment (input matrix = 6 objects x 3PCs). This yielded eigenvectors (basis functions) of size 3x3. The first two eigenvectors (2x2) were used to define the plane-of-best-fit to the points defined by the rows of the 6 objects belonging to the same category and they explained more than 93% of the variance of each set of points in the 3D space. Given two vectors (*v1,v2*) defining the first plane and (*v3,v4*) defining the second plane, the cosine of the angle θ between the planes was computed from the determinant of the dot-product matrix between the spanning vectors, normalized by the determinants within each subspace (see (Panichello & Buschman, 2021) for details). This yielded a measure ranging from 0° (identical subspaces) to 90° (orthogonal subspaces).

To quantify attentional shifts on the geometry of brain representations, we compared ED and PA between attended and distractors, including all ROIs (180 for each hemisphere) using a non-parametric sign-flipping permutation test. For each category, we computed the observed difference in representational dissimilarity (e.g., experiment 3OW attended vs. distractors), then generated a null distribution by randomly flipping the difference signs across 50,000 permutations. A permutation-based t-statistic was obtained as the ratio of the observed mean difference to the standard deviation of the null distribution, and two-tailed p-values were calculated as the proportion of permuted values exceeding the observed difference. P-values were extracted for ROIs (180 in each hemisphere) and metric (ED and PA) and corrected for multiple comparisons using FDR. We plotted significant results on a medial and lateral brain surface, overlapped with the Glasser Atlas.

To test whether the observed differences in representational geometry exceeded what would be expected by chance, we applied our full analysis pipeline to surrogate data (see figure 2C for a graphical representation). We generated these surrogates by concatenating stimuli of the same category across the three input matrices (i.e., attended 3OW, distractors 3OW and experiment OA), shuffling the rows, and then splitting them back into separate matrices. This procedure preserves the overall covariance structure within each category while disrupting the specific arrangement of stimuli in the low-dimensional manifold. Applying the pipeline to the shuffled matrices produced null distributions of differences between attended and distractor conditions over 1,000 iterations, for each ROI and hemisphere. We then compared the observed differences to these null distributions to compute permutation-based p-values and z-scores. Finally, we extracted p-values at the whole-brain level, applied false discovery rate (FDR) correction across ROIs and hemispheres, and visualized the

significant effects on the cortical surface, highlighting regions with reliably altered representational geometry.

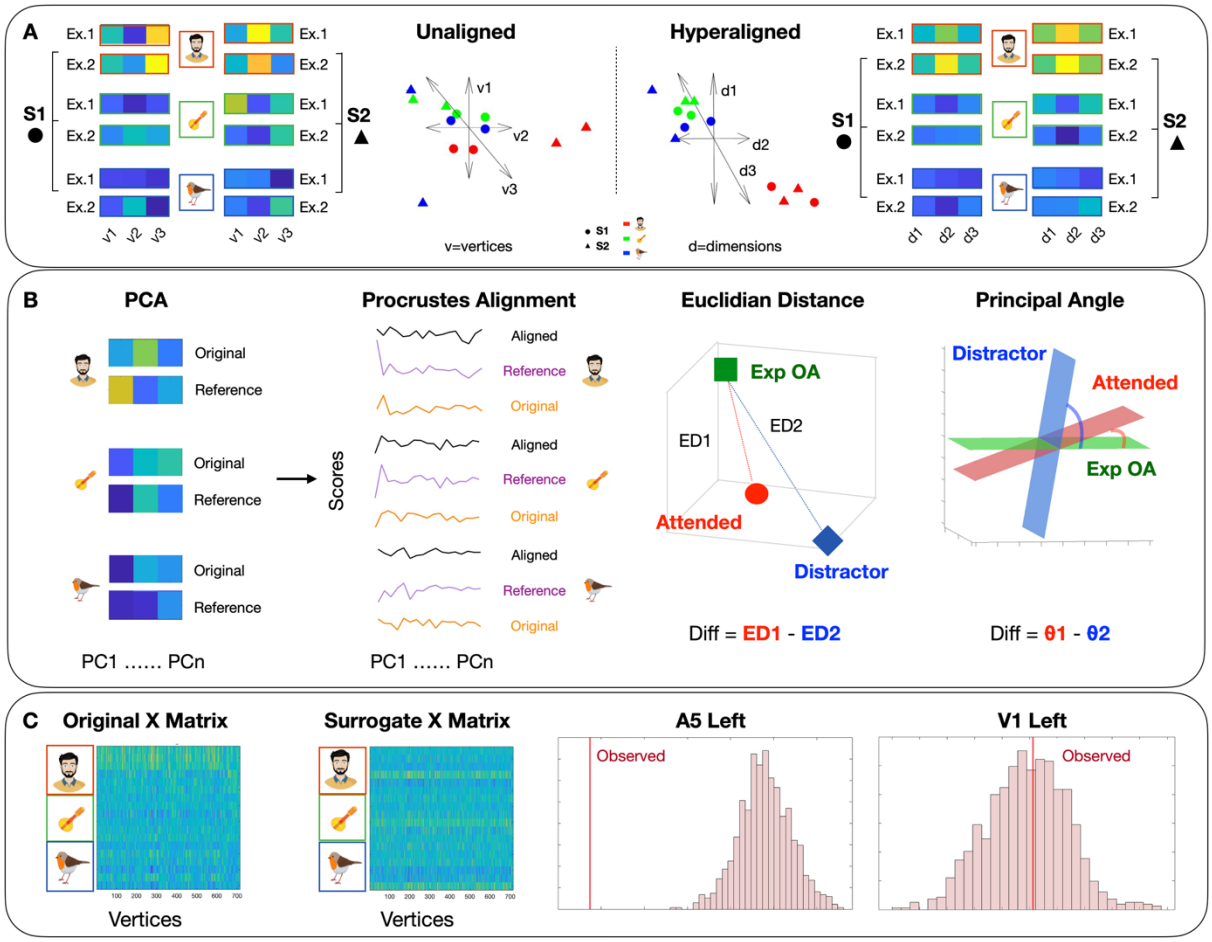

Fig. 2. Overview of the geometry analysis pipeline. A. Hyperalignment of response matrices (X) across subjects using Generalized Procrustes Analysis. *Left:* Beta matrices from two subjects for six stimuli (two per category) and their corresponding geometric representations in a three-vertex space. *Right:* Geometric representations in the aligned model space and the aligned beta matrices for the same subjects and stimuli. Hyperalignment creates a shared representational space across participants. **B.** Dimensionality reduction and comparison of representational geometry. PCA is applied to the hyperaligned data to reduce dimensionality. The resulting PC scores are then aligned to scores from experiment OA (reference) using Procrustes analysis without scaling. These aligned scores are used to quantify attention-related changes in representational geometry by computing Euclidean distances (ED) and Principal Angles (PA). Specifically, ED and PA are computed for each condition of experiments 3OW and 3OA relative to experiment OA, and the differences between them are extracted. **C**. Surrogate null distribution pipeline. To estimate null distributions, stimuli of the same category across the three input matrices (attended 3OW, distractors 3OW, and experiment OA) were concatenated, their rows shuffled and then split back into separate matrices to generate structureless surrogate data. The same analysis pipeline as in B was applied to these surrogate matrices to compute null distributions of differences between attended and distractor stimuli. Observed differences were then compared to these null distributions to derive permutation-based p-values.

Relating Neural Geometry to Subjective Attention. As part of this study, we tested whether the effect of attention on the geometry of brain representations could be predicted by how well participants reported being able to attend to the designated object (*Subjective Rating*). To do so, we quantified the effect of attention as the Euclidean distance between neural representations of the same exemplar across experiments. Specifically, experiments 3OA and 3OW consisted of 36 trials

composed of sound triplets, either from different categories (3OA) or the same category (3OW). Within a subject, the same exemplars were repeated across runs, but exemplar sets differed across subjects. For the Euclidean distance analysis, we focused on the exemplars that were common between experiment OA and experiments 3OA/3OW. Using the hyperaligned data from experiment OA as the reference space, we computed Euclidean distances between the principal component (PC) scores of the same stimuli across experiments in two ways: (i) using the raw, uncentered PC scores, and (ii) after centering the PC scores by subtracting the category-specific centroid from experiment OA, thereby aligning each stimulus representation to its OA category centroid. These trial-level distances served as dependent variables in our linear models, used to relate trial-level Euclidean distance measures to participants' subjective behavioral ratings. Data were filtered by condition (attended), and analyses were performed separately for each ROI (35) and hemisphere. For each subject, we estimated the slope of the relationship between subjective ratings and Euclidean distance using a linear regression model with Subjective Rating (predictor of interest) and Category (predictor of no interest) as predictors:

$$ED \sim \text{Subjective Rating} * \text{Category}$$

These slopes were then tested against zero using one-sample, one-tailed *t*-tests, reflecting the a-priori hypothesis of a negative relationship between attention and representational distance (higher rating corresponding to lower distance). Resulting *p*-values were corrected for multiple comparisons across ROIs and hemispheres using the false discovery rate (FDR). To visualize the results, we plotted individual subject slopes as well as model-based mean regression lines with 95% confidence intervals.

Results

In this study, we designed three experiments to provide complementary insights into how attention shapes the geometry of brain spaces and interacts with a naturalistic soundscape. The 3OA experiment examined how selective attention modulates processing when attending to objects from different categories within a scene. The 3OW experiment tested how attention affects processing in scenes composed of objects from the same category. Finally, the OA experiment, in which objects were presented alone and thus required no selective attention, served as a reference to (1) align the data across experiments (Procrustes alignment) and (2) compute the distance between attention-modulated and object-alone representations.

Comparison of Acoustic and Neural Manifolds. The OA experiment offers the opportunity to trace the progression of auditory representations along the auditory ventral pathway. We specifically asked whether neural manifolds evolve from acoustically grounded geometries in early auditory regions toward more compact and clustered configurations in higher-level auditory cortex and semantic networks. To address this, we quantified the alignment between acoustic and neural manifolds using the Procrustes Distance as a measure of alignment between matrices (Barbosa et al., 2025; Schönemann, 1966). A systematic pattern emerged in the distribution of Procrustes Distance, with smaller distances in primary and secondary auditory regions compared to temporal, parietal, and frontal areas (fig. 3E). This indicates that neural representational geometries only in primary and secondary auditory regions converge more closely toward the acoustic feature manifold than would

be expected by chance (fig. 3C). To illustrate these effects, we visualized the data projected into a 3D PCA space (PC1–PC3) for a few representative regions (Fig. 3D): **MBelt, A4, and STSva** as examples of regions showing a similar pattern between acoustic and neural manifolds (lower Procrustes Distance), and **TPOJ2, 45** and **TE2p** as examples of regions exhibiting more divergent patterns (higher Procrustes Distance). These regions were chosen to exemplify the two main effects observed: (1) **TE2p**, representing inferior temporal cortex, showed extensive overlap among stimulus representations, with both acoustic and categorical distinctions largely lost; and (2) **TPOJ2 and 45**, representing parietal and frontal areas, showed a categorical structure, primarily separating speech from the other categories.

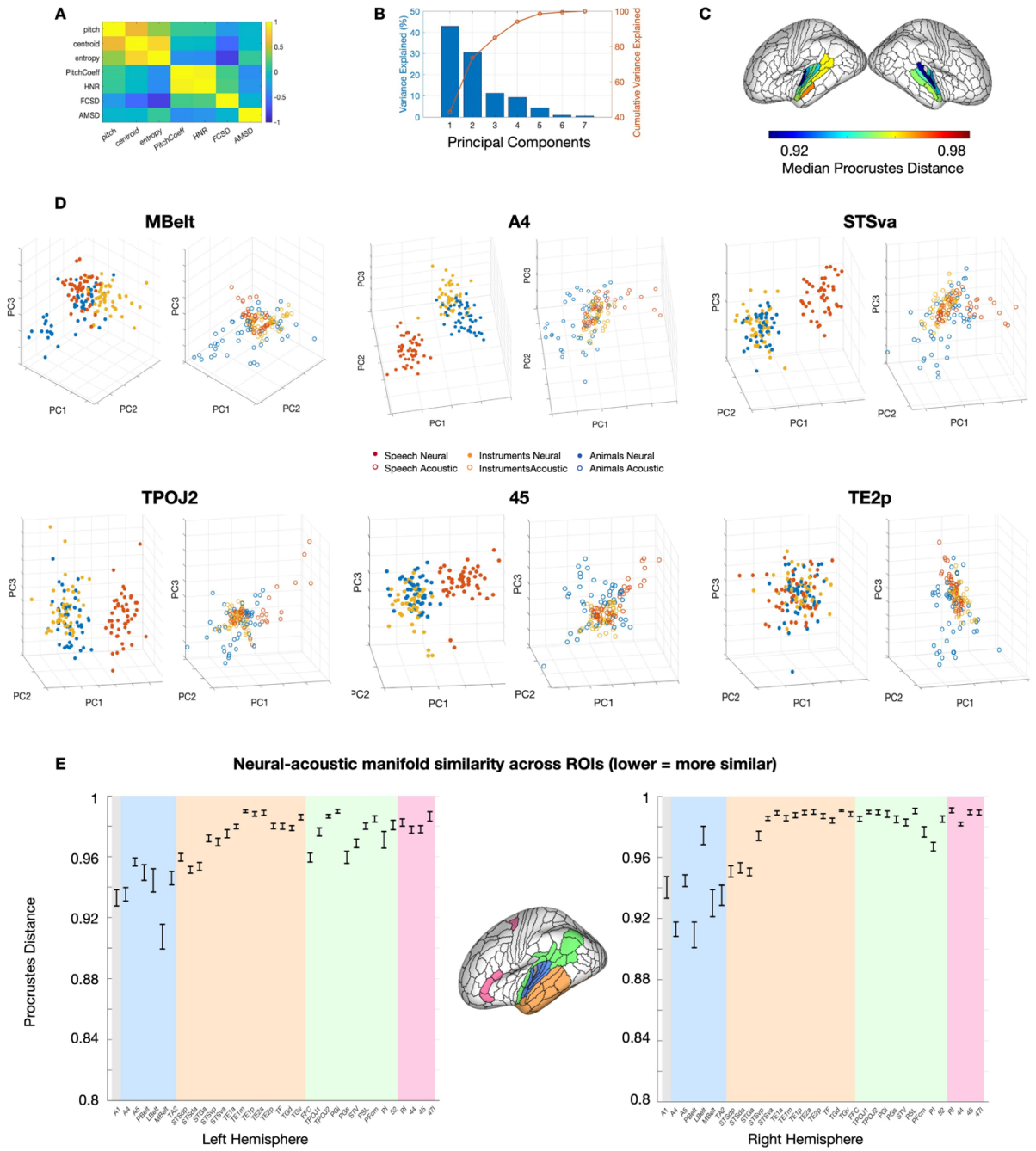

Fig. 3. Similarity between acoustic and neural manifolds. A. Correlation matrix between acoustic features of the 144 auditory stimuli. **B**. Variance Explained and Cumulative Variance Explained across Principal Components. **C.** Spatial map showing the distribution of median Procrustes Distance between the acoustic and neural manifold, significantly different from chance ($p<0.05$ - 10000 permutations). **D**. Projection of the 144 stimuli onto the first three principal components, color-coded by category, for the neural manifold of experiment OA and the aligned acoustic manifold (Procrustes alignment with scaling), shown as the median across participants. Represented regions are selected from the left hemisphere as examples of low (MBelt, A4, STSva) and high (TPOJ2, 45, TE2p) Procrustes Distances. **E.** Neural–acoustic manifold similarity (median ± SEM) across 35 regions in the left and right hemispheres. Regions are grouped into five clusters (also shown on the left lateral brain surface): primary auditory cortex A1 (grey), secondary auditory regions (blue), temporal regions (orange), parietal regions (green), and frontal regions (pink).

Geometric analysis of neural subspaces. This section addresses two questions: (1) How does selective attention reshape the geometry of neural population spaces? and (2) How does the surrounding soundscape modulate this geometry? We quantified differences between attended and distractor conditions in terms of Euclidean distance and Principal angle, separately for experiments 3OW and 3OA.

Regarding the **Euclidean Distance**, we did not observe any significant difference between attended and distractors in experiment 3OW (Fig. 4A), where attended and distractors belonged to the same category. This result suggests that selective attention does not substantially modulate the geometry of neural spaces in this specific soundscape. Indeed, the visualization of category centroids in a 3D PCA space (PC1–PC3) shows that both attended and distractors lie close to the object-alone configuration, with no systematic separation (Fig. 4A). It is important to remember that these distances are calculated at the object level (e.g., saxophone), not at the category level (e.g., instrument).

On the contrary, in experiment 3OA, where attended and distractors belonged to different categories, attention shifted manifold geometry toward the object-alone configuration. This effect was observed bilaterally in auditory and ventro-temporal regions for instruments and animals and extended to left frontal regions for speech (Fig. 4B). Surrogate analyses confirmed that these effects were significantly different from chance (Fig. 4C). The visualization in Fig. 4D illustrates the displacement of attended representations toward the object-alone configuration, whereas distractor representations remained more distant. Together, these findings indicate an interaction between attention and soundscape: attentional attraction toward the object-alone configuration occurred only when attended and distractor sounds belonged to different categories. Furthermore, when the categorical effect is removed by aligning the centroids of attended and distractor conditions to the object-alone centroid, all significant differences disappear.

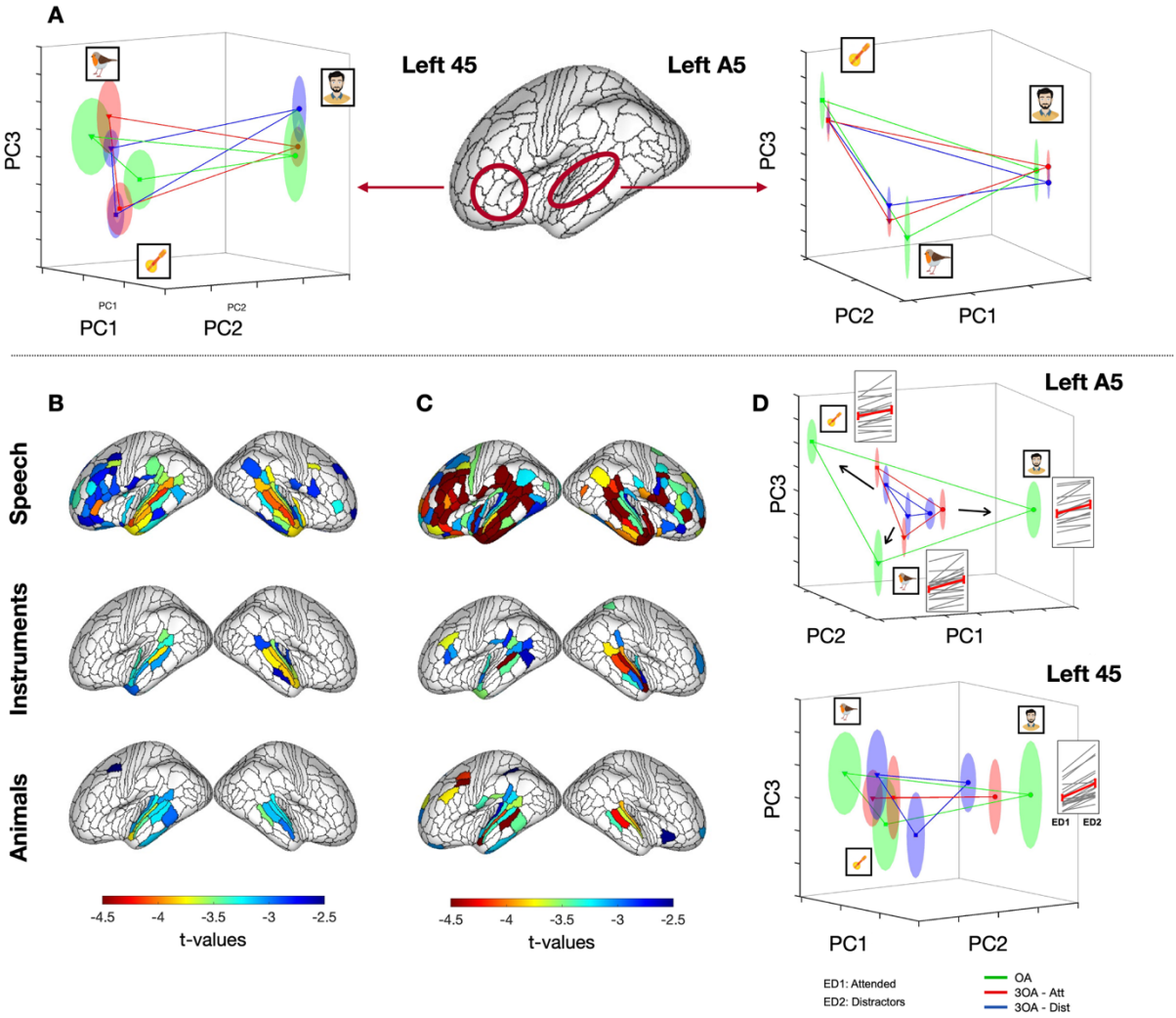

Figure 4. Effect of attention on the geometry of neural representations. A. Effect of attention on Euclidean distance in the 3OW experiment. Visualization of the mean centroids of three stimulus categories across experiments in a 3D PCA space (PC1–PC3) for two regions: Left 45 and A5. Ellipsoids around each centroid represent the standard error of the mean (SEM), and colored lines connect category centroids within each experiment (green = object alone; red = attended; blue = distractors). The lateral brain surface shows the location of the ROIs. **B. Effect of attention on Euclidean distance in the 3OA experiment.** Whole-brain t-values for speech, instrument, and animal sounds are displayed on the lateral surface. Negative t-values indicate that attended sounds are represented closer to the object-alone configuration than distractor sounds. **C. Statistical significance relative to chance.** Whole-brain z-values show regions where the attentional effect on Euclidean distance differs significantly from a surrogate null distribution. Larger negative values indicate that the observed difference is significantly more negative than expected by chance. **D. Attentional effect of attraction.** Visualization of the mean centroids of three stimulus categories across experiments in a 3D PCA space (PC1–PC3) for two regions. The black arrows in the Left A5 panel indicate the displacement of attended representations toward the object-alone configuration. For each significant effect, Euclidean distances ED1 (attended vs. object alone) and ED2 (distractors vs. object alone) are plotted for individual subjects (grey lines) and group averages (red line with SEM error bars) to illustrate the consistency of the effect.

Regarding the **Principal Angle**, no significant differences were observed in the 3OW experiment (fig. 5A), consistent with the lack of attentional effects on Euclidean distance. In 3OA, a few left-hemisphere regions showed category-specific effects: FOP2 (frontal operculum) for speech and PBelt (posterior belt of the auditory cortex) for instruments, where attention reduced the angle with the object-alone representation, and 6v (ventral part of area 6), where attention increased the angle (fig. 5B). FOP2 is implicated in higher-order auditory and speech-related processing, PBelt in processing

complex sounds such as music and environmental stimuli, and 6v in sensorimotor integration and auditory-motor interactions. These results suggest that, while attention consistently shifts neural representations toward the object-alone configuration in terms of Euclidean distance, effects on the orientation of the neural subspace are more limited and region- or category-specific.

Notably, including a higher number of PCs in the analysis did not substantially alter the Euclidean distance results, which remained similar with only a slight increase in the number of involved regions. For Principal Angle, using more PCs revealed effects in different auditory (A1), parietal (LIPv), and frontal (6a and IFSp) regions, only for instruments and animals, confirming that the orientation of neural subspaces is region- and category-specific (see Fig S1).

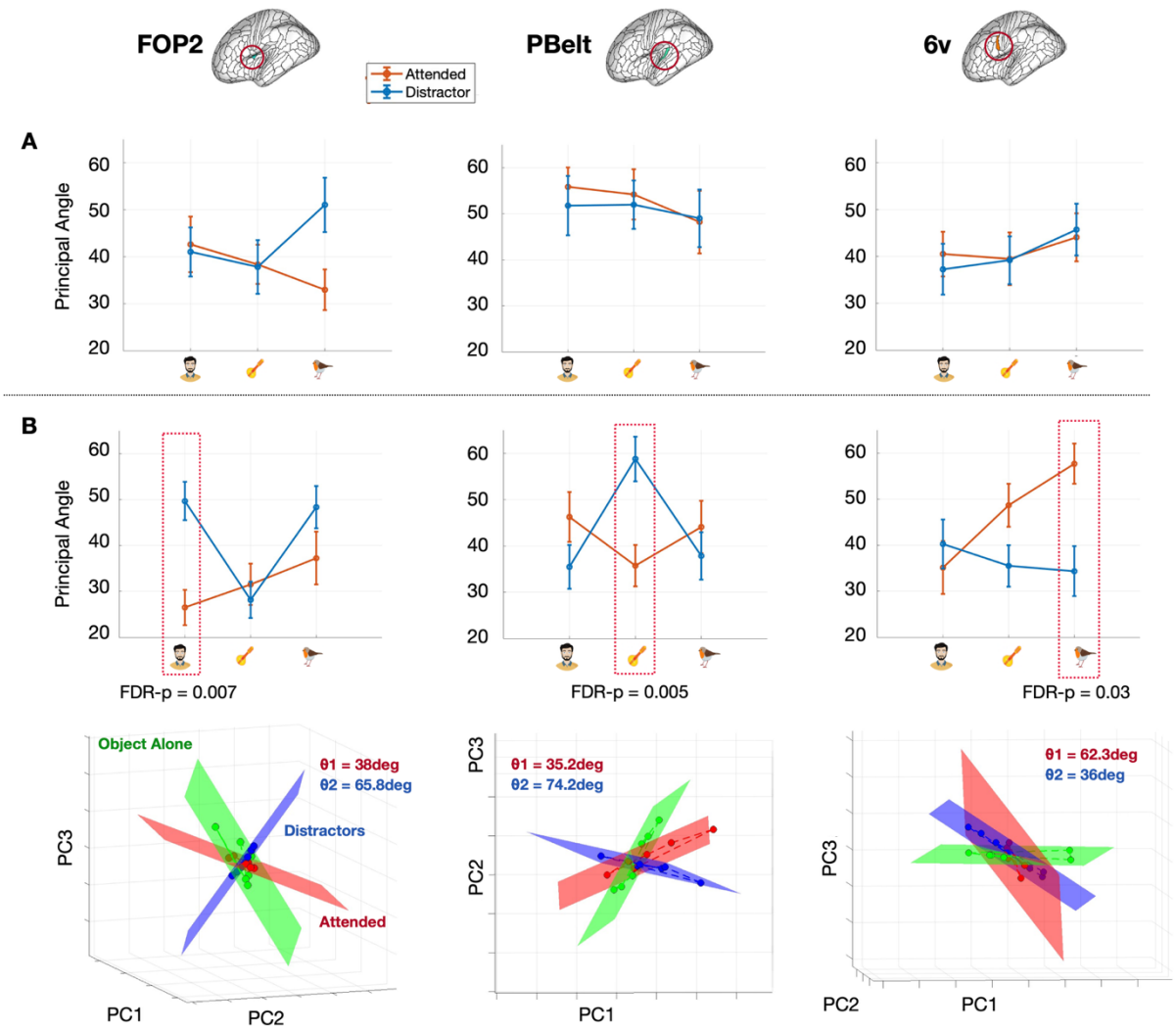

Figure 5. Effect of attention on the geometry of neural representations. A. Effect of attention on Principal Angle in the 3OW experiment. Principal angle is shown as a function of category (speech, instruments, and animals) for attended (red) and distractor (blue) conditions. The lateral brain surface indicates the spatial location of each ROI. **B. Effect of attention on Principal Angle in the 3OA experiment.** Upper panel: Principal angle as a function of category (speech, instruments, and animals) for attended (red) and distractor (blue) conditions, in ROIs showing a significant attentional effect relative to chance. Bottom panel: Visualization of the three main planes extracted from the object-alone (green), attended (red), and distractor (blue) conditions in a 3D PCA space (PC1–PC3) for a representative subject. Θ_1 indicates the angle between attended and object-alone, while Θ_2 indicates the angle between distractor and object-alone. Dots represent stimulus exemplars projected onto each condition's 2D subspace, and dashed lines connect exemplars of the same category.

Relating Neural Geometry to Subjective Attention. Behavioral performance (match-to-sample task) and subjective ratings of the ability to selectively attend to the specified stimulus were analyzed in our previous work (Wikman et al., 2025). Briefly, both measures showed significant effects of Experiment and Category, with the lowest accuracy and ratings observed in experiment 3OW and for instrument sounds. Building on these findings, the present study asks whether attentional effects observed at the behavioral level can predict the brain representational geometry.

When analyzing the raw data without category centroid alignment, we observed a significant effect of Subjective Rating in the 3OA experiment (Fig. 6A), primarily in the primary and secondary auditory cortex and in some temporal regions belonging to the superior temporal sulcus (STS). Euclidean distance between the attended and object-alone representations decreased as subjective rating increased, with most participants showing negative slopes for the regressor of interest. Because subjective rating reflects participants perceived ability to attend to the stimulus, this result suggests that scene representations aligned more closely with the attended exemplar when attentional engagement was stronger. Notably, this effect emerged only when the attended and distractor stimuli belonged to different categories.

When considering centroid-aligned data (removing the effect of sound category), we found a significant effect of Subjective Rating in both experiments (Fig. 6B). In this case, the effect was stronger and more widespread in 3OW, encompassing bilateral primary and secondary auditory cortex as well as multiple temporal regions. By contrast, in 3OA the effect was largely restricted to the same regions identified in the raw-data analysis, with FDR-corrected p-values close to the significance threshold ($p = 0.047$). In both experiments, the trend was consistently negative, with nearly all participants showing smaller distances for higher subjective ratings (negative slopes).

Together, these results indicate a systematic relationship between behavioral and neural data: in auditory and ventro-temporal regions, higher subjective ratings of attentional engagement were associated with smaller distances between attended and object-alone representations.

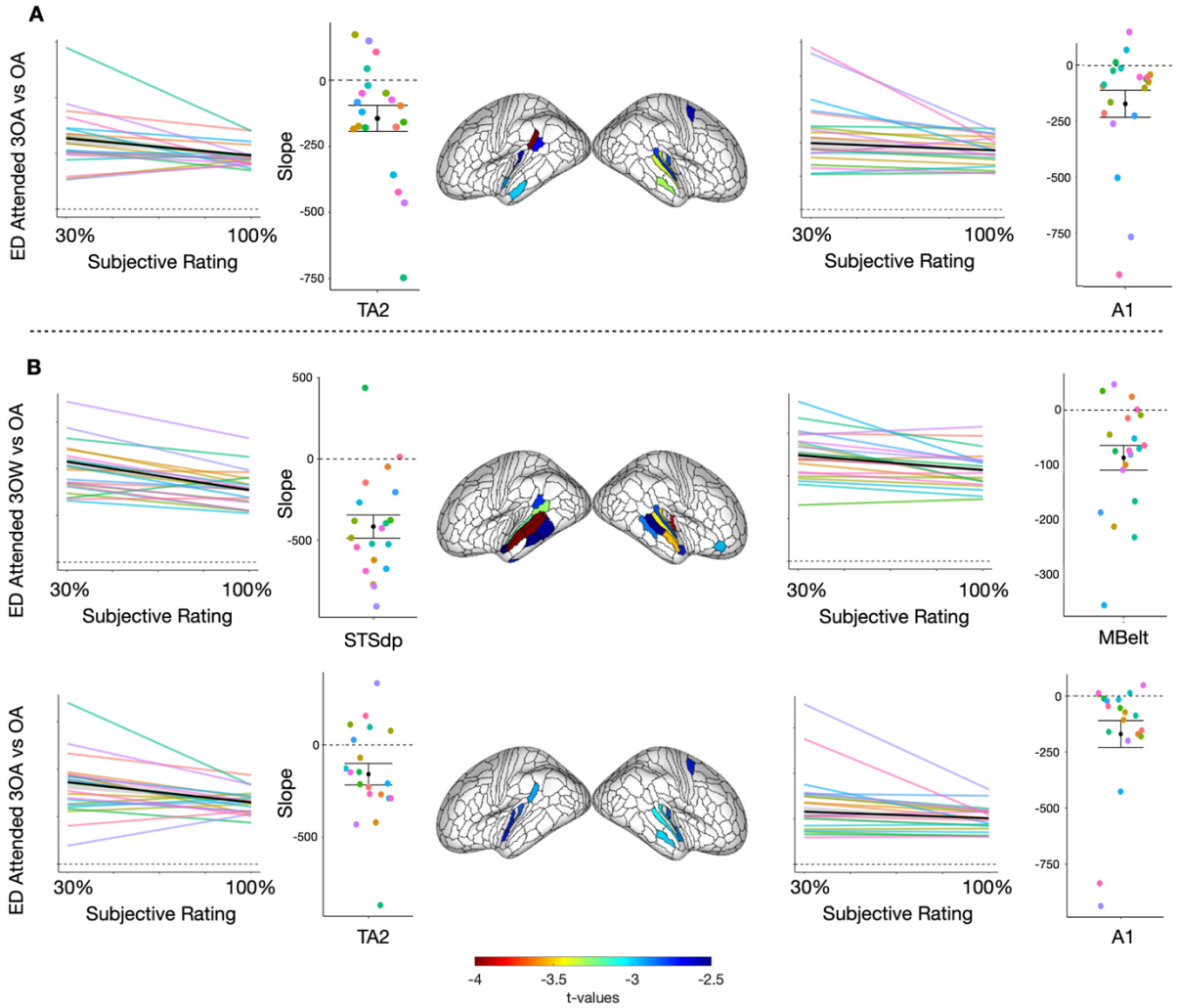

Fig. 6. Subjective rating predicts Euclidean Distance (ED). Spatial map of brain regions where subjective ratings of attention significantly predicted ED, showing a general trend: the distance between the attended and the object-alone representations decreased as subjective rating increased. The central panel highlights significant effects on the lateral brain surface, while the side panels show ROI-specific results. For each ROI we display: (1) regression lines for individual subjects together with the group mean ± 95% Confidence Interval, and (2) the mean slope ± Standard Error with individual subject values. Panel A shows results for experiment 3OA using the raw ED. Panel B shows results for experiments 3OW (top) and 3OA (bottom), after centering the PC scores by subtracting the category-specific centroid, thereby aligning scores to the OA centroid.

Discussion

In this study, we examined how acoustic structure, selective attention, and behavioral relevance shape neural population geometry in the brain, focusing in particular on the human auditory ventral pathway. The main findings can be organized into three key points. First, comparing acoustic and neural manifolds revealed a systematic representational progression: primary and secondary auditory regions closely reflected acoustic structure, whereas higher-level temporal and frontal regions showed increasingly divergent and abstract geometries. Second, geometric analyses showed that attention reshaped neural spaces depending on the soundscape, shifting representations toward the object-alone configuration only when attended and distractor sounds belonging to different categories. Third, stronger subjective attentional engagement was associated with closer alignment between

attended and object-alone representations, particularly in auditory and ventro-temporal regions. Together, these findings highlight the dynamic interplay between sensory structure, attention, and behavior in shaping auditory population geometry.

Comparison of Acoustic and Neural Manifolds. By comparing acoustic and neural manifolds, we directly addressed how auditory representations transform along the ventral pathway. Our findings reveal a systematic transformation of representational geometry, from early acoustic encoding to more abstract, higher-order representations. Lower Procrustes distances in primary and secondary auditory regions, as well as in the superior temporal sulcus (STS), indicate that neural manifolds in these areas partially preserve the geometry of the acoustic feature space, reflecting the encoding of spectral and temporal features (Moerel et al., 2012; Santoro et al., 2014). Notably, in the acoustic manifold, animal sounds are more widely dispersed along the principal components, whereas speech and instrument sounds occupy more overlapping regions. This broader acoustic spread is progressively diminished along the cortical hierarchy, indicating that these fine-grained acoustic distinctions are lost as neural representations become more abstract.

Higher-order temporal, parietal, and frontal regions exhibit progressively larger Procrustes distances, reflecting a reorganization of neural manifolds away from the low-dimensional acoustic representation. Crucially, in our results, higher Procrustes distances do not necessarily indicate stronger categorical coding; rather, they reflect greater divergence from the acoustic geometry. Categorical organization emerges in secondary auditory regions (A4) and STS, where distances remain relatively low and stimuli show a clearer separation mainly between speech and other categories (instruments and animals). Parietal and frontal areas show partial categorical structure, while ventro-temporal regions (inferior temporal cortex) exhibit extensive overlap among stimulus representations, with both acoustic and categorical distinctions largely lost. These results align with previous findings that anterior and posterior superior temporal regions differentiate speech, music, and other sound categories (Giordano et al., 2013; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015; Overath et al., 2015), highlighting a progression from feature-based to more abstract population codes along the auditory hierarchy.

A similar hierarchical transformation has been observed in the visual system along the ventral pathway for object recognition. Early sensory regions encode detailed stimulus features in high-dimensional population activity, whereas downstream areas reduce dimensionality and cluster responses into more abstract, robust representations (DiCarlo et al., 2012). This framework combines the efficient coding hypothesis (Simoncelli & Olshausen, 2001) and the observation that higher-order neural codes occupy low-dimensional subspaces, supporting reliable computations despite noise (Chung et al., 2018). High-dimensional population codes also exhibit a characteristic power-law structure, ensuring smooth and robust encoding of stimulus details (Stringer et al., 2019). These findings suggest that hierarchical representational transformations are a general principle of cortical processing. Our geometric approach extends this framework to audition, providing a quantitative characterization of how auditory representations evolve along the ventral pathway, from acoustically grounded population codes to more abstract manifolds.

Geometric analysis of neural subspaces. In this study, we addressed two central questions: how selective attention reshapes the geometry of neural population spaces, and how this geometry is

further modulated by the surrounding soundscape under the same attentional condition. On the same neuroimaging data, we previously demonstrated that attended objects can be decoded more accurately from AC than distractor objects (Wikman et al., 2025). Extending these findings, our representational geometry analyses reveal that attention reshapes the population space by attracting neural responses toward the object-alone template. Crucially, the **surrounding soundscape modulates this effect**: in 3OW scenes, where all overlapping sounds belong to the same category, both attended and distractor objects' representations are close to the object-alone template of their sound category, yielding no significant attentional modulation in Euclidean distance. By contrast, in 3OA scenes, where attended and distractor sounds belong to different categories, population responses are overall more dispersed, but attention selectively attracts the attended representation toward its template relative to distractors. Importantly, this effect appears to reflect categorical differences. Indeed, when we align the centroids of attended and distractor representations to the object-alone template across experiments, the attraction effect disappears, and the pairwise Euclidean distances are no longer significant. This indicates that attention reshapes neural geometry primarily relative to category-level structure, rather than on the level of individual sound objects. This **attention- and context-dependent "geometric attraction"** complements the MVPA results: while MVPA quantifies increased discriminability along readout dimensions, geometry reveals how attention reorganizes the entire representational landscape, drawing population activity toward the object-alone representation. Together, these converging findings support a population-level interpretation of attentional selection in auditory scenes (Barack & Krakauer, 2021), whereby attention enhances separability and simultaneously reshapes representational geometry to prioritize behaviorally relevant sound objects.

Beyond context-dependent effects, attention also operates differently across sound categories. For speech, attentional modulation was observed not only in auditory and temporal regions (core, belt, Superior Temporal Sulcus - STS) but also in extensive left-lateralized frontal areas, consistent with higher-order language processing and the reshaping of neural manifolds in regions supporting speech comprehension (Fedorenko et al., 2024; Friederici & Gierhan, 2013). In contrast, for instruments and animals, attentional effects were largely restricted to auditory and temporal regions (core, belt, STS), reflecting modulation of spectral and intermediate-level features rather than higher-order category representations (Hausfeld et al., 2024; Norman-Haignere et al., 2015). This pattern suggests that selective attention reshapes neural population geometry in a category- and context-dependent manner, primarily engaging lower- and intermediate-level auditory areas for non-speech objects, while additionally recruiting higher-order frontal regions for speech.

Relating Neural Geometry to Subjective Attention. There is longstanding interest in linking neural changes to behavioral performance. In the context of brain manifolds, this has motivated efforts to examine whether variations in geometric properties—such as distances or alignments within neural representational spaces—systematically relate to behavior (Chou et al., 2025; Fascianelli et al., 2024; Santo-Angles et al., 2025). Building on this framework, we tested whether the observed attraction of attended stimuli toward the object-alone representation is behaviorally meaningful. To address this, we tested whether subjective ratings of attentional engagement predict the Euclidean distance between brain manifolds, while controlling for category-level differences.

Our results highlight attentional engagement as a key factor shaping neural representational geometry, predicting the distance between attended and object-alone representations. In particular, higher subjective ratings of attentional engagement predict smaller Euclidean distances between attended and object-alone representations. In Experiment 3OA, this effect was localized mainly in primary and secondary auditory cortex as well as in superior temporal regions and was evident in both raw distances and after centroid alignment. Notably, the effect was stronger in the raw data, suggesting that attentional engagement in this experiment primarily reflects coarse, global differences in neural geometry between attended and distractor sounds. In Experiment 3OW, the effect emerged only after centroid alignment, extending to temporal ventral regions in addition to auditory and superior temporal areas. This indicates that attentional engagement predicts fine-grained, within-category shifts in the neural manifold — subtle changes in how individual stimuli are represented relative to their category centroid. The emergence of these effects only after alignment suggests that behaviorally meaningful information is encoded at this finer level, rather than in overall category-level distances. These findings demonstrate that the "geometric attraction" of population responses closely reflects participants' subjective estimates of their level of attention, supporting a direct link between representational geometry and behavior.

Limitations and future directions. The geometry of brain spaces can be altered through multiple mechanisms. In this study, the predominant effect we observed was an "attraction" of neural population responses toward the object-alone configuration induced by selective attention. We also examined potential differences in the rotation of brain spaces, quantified using Principal Angles, but only sparse significant effects were found (see Fig. 5). Notably, our analyses focused exclusively on between-space differences—that is, the geometric relationship between attended, distractor, and object-alone manifolds—without investigating within-space geometry, such as the Euclidean distances between objects of the same category. Previous work suggests that selective attention may reshape the internal structure of neural manifolds by collapsing or expanding representations of individual objects within a category (Sereno & Lehky, 2018). For example, the internal structure of the attended manifold may become more similar to that of the object-alone configuration than to distractor manifolds. Future work could examine these within-space effects to provide a more comprehensive account of how attention reshapes representational geometry.

A second limitation is related to the temporal resolution: fMRI averages activity over the stimulus window, preventing observation of how geometry evolves over time. According to the hierarchical decomposition hypothesis (Cusack et al., 2004), auditory representations are dynamically segregated along the cortical hierarchy, with attention gradually gating higher-level semantic and linguistic processing (Puschmann et al., 2024). Using short stimulus segments, attention may not have fully modulated higher-order representations. Future work with high-temporal-resolution recordings could reveal both within-space effects and the temporal unfolding of attentional modulation across hierarchical stages.

Conclusions. Our study demonstrates that selective attention reshapes neural population codes in a context- and category-dependent manner, effectively "attracting" representations of a soundscape towards the template of the attended sound object. This effect is modulated by the surrounding soundscape, with stronger attentional modulation when distractors belong to different categories (3OA), and varies across stimulus types, engaging left frontal regions for speech but remaining largely confined to auditory areas for non-speech sounds. Importantly, these geometric changes are

behaviorally meaningful: participants' subjective attention ratings predict the proximity of attended representations to their object-alone templates. Together, these findings highlight the functional relevance of representational geometry in auditory perception and provide a population-level framework for understanding how attention prioritizes behaviorally relevant information in complex environments. Future work could explore how these mechanisms interact with other cognitive processes such as learning and memory, and whether similar geometric principles govern attentional selection in other sensory modalities or real-world listening contexts.

References

Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*, *61*(3), 467–480. https://doi.org/10.1016/j.neuron.2008.12.027

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews. Neuroscience*, *22*(6), 359–371. https://doi.org/10.1038/s41583-021-00448-6

Barbosa, J., Nejatbakhsh, A., Duong, L., Harvey, S. E., Brincat, S. L., Siegel, M., Miller, E. K., & Williams, A. H. (2025). Quantifying differences in neural population activity with shape metrics. In *bioRxiv*. https://doi.org/10.1101/2025.01.10.632411

Boersma, P., & Weenink, D. (2001). *Praat, doing phonetics by computer* [Computer program]. Retrieved from https://www.praat.org

Caruso, V. C., Mohl, J. T., Glynn, C., Lee, J., Willett, S. M., Zaman, A., Ebihara, A. F., Estrada, R., Freiwald, W. A., Tokdar, S. T., & Groh, J. M. (2018). Single neurons may encode simultaneous stimuli by switching between activity patterns. *Nature Communications*, *9*(1), 2715. https://doi.org/10.1038/s41467-018-05121-8

Cheng, Y.-A., Sanayei, M., Chen, X., Jia, K., Li, S., Fang, F., Watanabe, T., Thiele, A., & Zhang, R.-Y. (2025). A neural geometry approach comprehensively explains apparently conflicting models of visual perceptual learning. *Nature Human Behaviour*, *9*(5), 1023–1040. https://doi.org/10.1038/s41562-025-02149-x

Chou, C.-N., Kim, R., Arend, L. A., Yang, Y.-Y., Mensh, B. D., Shim, W. M., Perich, M. G., & Chung, S. (2025). Geometry Linked to Untangling Efficiency reveals structure and computation in neural populations. In *bioRxivorg*. https://doi.org/10.1101/2024.02.26.582157

Chung, S., Lee, D. D., & Sompolinsky, H. (2018). Classification and geometry of general perceptual manifolds. *Physical Review. X*, *8*(3). https://doi.org/10.1103/physrevx.8.031003

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(4), 643–656. https://doi.org/10.1037/0096-1523.30.4.643

Da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., & Saenz, M. (2013). Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(5), 1858–1863. https://doi.org/10.1523/JNEUROSCI.4405-12.2013

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(29), 11854–11859. https://doi.org/10.1073/pnas.1205381109

Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine in cognitive neuroscience. In *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/2104.00145

Engel, T. A., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., & Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science (New York, N.Y.)*, *354*(6316), 1140–1144. https://doi.org/10.1126/science.aag1420

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J.

(2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fascianelli, V., Battista, A., Stefanini, F., Tsujimoto, S., Genovesio, A., & Fusi, S. (2024). Neural representational geometries reflect behavioral differences in monkeys and recurrent neural networks. *Nature Communications*, *15*(1), 6479. https://doi.org/10.1038/s41467-024-50503-w

Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews. Neuroscience*, *25*(5), 289–312. https://doi.org/10.1038/s41583-024-00802-4

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254. https://doi.org/10.1016/j.conb.2012.10.002

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention--focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. https://doi.org/10.1016/j.conb.2007.07.011

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, *6*(11), 1216–1223. https://doi.org/10.1038/nn1141

Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., & Miller, L. E. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, *9*(1), 4233. https://doi.org/10.1038/s41467-018-06560-z

Gao, S., Mishne, G., & Scheinost, D. (2021). Nonlinear manifold learning in functional magnetic resonance imaging uncovers a low-dimensional space of brain dynamics. *Human Brain Mapping*, *42*(14), 4510–4524. https://doi.org/10.1002/hbm.25561

Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(9), 2025–2037. https://doi.org/10.1093/cercor/bhs162

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*(1), 33–51. https://doi.org/10.1007/bf02291478

Guntupalli, J. S., Feilong, M., & Haxby, J. V. (2018). A computational model of shared fine-scale structure in the human connectome. *PLoS Computational Biology*, *14*(4), e1006120. https://doi.org/10.1371/journal.pcbi.1006120

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *26*(6), 2919–2934. https://doi.org/10.1093/cercor/bhw068

Hausfeld, L., Hamers, I. M. H., & Formisano, E. (2024). FMRI speech tracking in primary and non-primary auditory cortex while listening to noisy scenes. *Communications Biology*, *7*(1), 1217. https://doi.org/10.1038/s42003-024-06913-z

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*, *9*. https://doi.org/10.7554/eLife.56601

Ibos, G., & Freedman, D. J. (2016). Interaction between spatial and feature attention in posterior parietal cortex. *Neuron*, *91*(4), 931–943. https://doi.org/10.1016/j.neuron.2016.07.025

Ilomäki, M., Lindblom, J., Flykt, M., Vänskä, M., Punamäki, R.-L., & Wikman, P. (2025). Similarity in early life stress exposure is associated with similarity in neural representations in early adulthood. *Human Brain Mapping*, *46*(14), e70373. https://doi.org/10.1002/hbm.70373

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(22), 8210–8219. https://doi.org/10.1523/JNEUROSCI.6153-09.2011

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6

Kauramäki, J., Jääskeläinen, I. P., & Sams, M. (2007). Selective attention increases both gain and feature

selectivity of the human auditory cortex. *PloS One*, *2*(9), e909. https://doi.org/10.1371/journal.pone.0000909
Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology: CB*, *25*(5), 595–600. https://doi.org/10.1016/j.cub.2014.12.050
Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience*, *42*(1), 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906
Lage-Castellanos, A., De Martino, F., Ghose, G. M., Gulban, O. F., & Moerel, M. (2022). Selective attention sharpens population receptive fields in human auditory cortex. *Cerebral Cortex* . https://doi.org/10.1093/cercor/bhac427
Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750–761. https://doi.org/10.1016/j.neuron.2012.11.034
Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(22), 7604–7612. https://doi.org/10.1523/JNEUROSCI.0296-10.2010
Li, H.-H., & Curtis, C. E. (2023). Neural population dynamics of human working memory. *Current Biology: CB*, *33*(17), 3775–3784.e4. https://doi.org/10.1016/j.cub.2023.07.067
Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236. https://doi.org/10.1038/nature11020
Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2013). Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(29), 11888–11898. https://doi.org/10.1523/JNEUROSCI.5306-12.2013
Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(8), 4277–4291. https://doi.org/10.1093/cercor/bhx138
Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035
Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4
Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, *18*(6), 903–911. https://doi.org/10.1038/nn.4021
Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601–605. https://doi.org/10.1038/s41586-021-03390-w
Perich, M. G., Narain, D., & Gallego, J. A. (2025). A neural manifold view of the brain. *Nature Neuroscience*, *28*(8), 1582–1597. https://doi.org/10.1038/s41593-025-02031-z
Puschmann, S., Regev, M., Fakhar, K., Zatorre, R. J., & Thiel, C. M. (2024). Attention-driven modulation of auditory cortex activity during selective listening in a multispeaker setting. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *44*(15), e1157232023. https://doi.org/10.1523/JNEUROSCI.1157-23.2023
Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *98*, 262–268. https://doi.org/10.1016/j.cortex.2017.10.020
Santo-Angles, A., Yang, J., Zhou, Y., Chu, W. K. H., Lindsay, G. W., & Sreenivasan, K. K. (2025). Neural subspaces encode sequential working memory, but neural sequences do not. In *bioRxiv*. https://doi.org/10.1101/2025.09.05.674385
Schäfer, P. J., Corona-Strauss, F. I., Hannemann, R., Hillyard, S. A., & Strauss, D. J. (2018). Testing the limits of the stimulus reconstruction approach: Auditory attention decoding in a four-speaker free field environment. *Trends in Hearing*, *22*, 233121651881660. https://doi.org/10.1177/2331216518816600
Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*(1), 1–10. https://doi.org/10.1007/bf02289451
Sereno, A. B., & Lehky, S. R. (2018). Attention effects on neural population representations for shape and location are stronger in the ventral than dorsal stream. *eNeuro*, *5*(2). https://doi.org/10.1523/ENEURO.0371-17.2018

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123. https://doi.org/10.1016/j.tins.2010.11.002

Shamma, S., Fritz, J., David, S., Elhilali, M., Winkowski, D., & Yin, P. (2010). Correlates of auditory attention and task performance in primary auditory and prefrontal cortex. In *The Neurophysiological Bases of Auditory Perception* (pp. 555–570). Springer New York. https://doi.org/10.1007/978-1-4419-5686-6_51

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216. https://doi.org/10.1146/annurev.neuro.24.1.1193

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Sorscher, B., Ganguli, S., & Sompolinsky, H. (2022). Neural representational geometry underlies few-shot concept learning. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(43), e2200800119. https://doi.org/10.1073/pnas.2200800119

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. https://doi.org/10.1038/nn.3574

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365. https://doi.org/10.1038/s41586-019-1346-5

St-Yves, G., Kay, K., & Naselaris, T. (2025). Variation in the geometry of concept manifolds across human visual cortex. *PLoS Computational Biology*, *21*(9), e1013416. https://doi.org/10.1371/journal.pcbi.1013416

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews. Neuroscience*, *15*(6), 355–366. https://doi.org/10.1038/nrn3731

Wikman, P., Muukkonen, I., Kauramaki, J., Laaksonen, V., Waris, O., Petkov, C., & Rauschecker, J. (2025). Selective attention network in naturalistic auditory scenes is object and scene specific. In *bioRxiv*. https://doi.org/10.1101/2025.01.03.631190. In press in Journal of Neuroscience.

Wikman, P., Sahari, E., Salmela, V., Leminen, A., Leminen, M., Laine, M., & Alho, K. (2021). Breaking down the cocktail party: Attentional modulation of cerebral audiovisual speech processing. *NeuroImage*, *224*(117365), 117365. https://doi.org/10.1016/j.neuroimage.2020.117365

Wikman, P., Salmela, V., Sjöblom, E., Leminen, M., Laine, M., & Alho, K. (2024). Attention to audiovisual speech shapes neural processing through feedback-feedforward loops between different nodes of the speech network. *PLoS Biology*, *22*(3), e3002534. https://doi.org/10.1371/journal.pbio.3002534

Wikman, P., Ylinen, A., Leminen, M., & Alho, K. (2022). Brain activity during shadowing of audiovisual cocktail party speech, contributions of auditory-motor integration and selective attention. *Scientific Reports*, *12*(1), 18789. https://doi.org/10.1038/s41598-022-22041-2

Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X.-J., Yang, T., Dehaene, S., Tang, S., Min, B., & Wang, L. (2022). Geometry of sequence working memory in macaque prefrontal cortex. *Science (New York, N.Y.)*, *375*(6581), 632–639. https://doi.org/10.1126/science.abm0204

Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage*, *10*(5), 544–554. https://doi.org/10.1006/nimg.1999.0491