Brain Representations of Natural Sound Statistics

Yousef Mohammadi[a,b], Alexander J. Billig[c], Joel I. Berger[d], Timothy D. Griffiths[a,b]

[a] Department of Imaging Neuroscience, University College London
12 Queen Square, London, WC1N 3AR, United Kingdom

[b] Newcastle University Medical School
Framlington Place, Newcastle upon Tyne, NE2 4HH, United Kingdom

[c] UCL Ear Institute, University College London
332 Grays Inn Road, London, WC1X 8EE, United Kingdom

[d] Department of Neurosurgery, The University of Iowa
Iowa City, IA, 52242, United States of America

[*] Correspondence concerning this article should be addressed to Yousef Mohammadi, 12 Queen Square, London, WC1N 3AR, United Kingdom. E-mail: y.mohammadi@ucl.ac.uk

Number of **pages**: 14
Number of **figures**: 4
Number of **tables**: 1
Number of words for **Abstract**: 220
Number of words for **Introduction**: 654
Number of words for **Discussion**: 488

Conflict of interest statement: The authors declare no competing financial interests.

Acknowledgments: This work was supported by the Medical Research Council, United Kingdom (MR/T032553/1). We thank Martina Callaghan, Fred Dick and Oliver Josephs, for designing the sound delivery system and MRI sequences; Karl Friston and Peter Zeidman for their advice on experimental design; Richard McWalter for providing code and commenting on the manuscript; and the Imaging Support team at the UCL Department of Imaging Neuroscience.

Abstract

Natural sound textures, such as rain or crackling fire, are defined by time-averaged summary statistics that shape their perceptual identity. Variations in these statistics provide a controlled means to examine how the human brain processes complex auditory structure. In this study, we investigated how such statistics are represented along the ascending auditory pathway, within auditory cortex, and in high-level regions involved in natural pattern analysis. Using fMRI, we measured brain responses to synthetic sound textures in which higher-order statistical structure was systematically degraded while preserving the overall texture category. Participants listened to sounds with varying levels of naturalness, defined by their statistical fidelity to real-world textures, and performed a perceptual comparison task, judging whether two sequentially presented sounds matched in naturalness. We observed that increasing naturalness elicited stronger BOLD responses across bilateral auditory cortex for both reference and test sounds. Activity in medial temporal lobe regions, including the entorhinal cortex and hippocampus, was modulated by naturalness in a position-dependent manner. The entorhinal cortex activity was modulated only during the test sound, suggesting a role in perceptual-mnemonic comparison. The hippocampal connectivity with auditory cortex increased when reference textures were more degraded or less natural, indicating top-down inference under uncertainty. Together, these findings highlight the interplay between bottom-up encoding and memory-based mechanisms in supporting judgments of auditory realism based on summary statistics.

Introduction

Everyday listening environments are filled with rich acoustic backgrounds, such as rustling leaves, falling rain, or crackling fire. These sounds are known as sound textures as they are produced by a superposition of many similar acoustic events. Unlike speech or music, textures are characterized by stationarity over time, meaning their perceptual identity is maintained through time-averaged statistics rather than transient or sequential structure (McDermott and Simoncelli, 2011; McDermott et al., 2013). These time-averaged summary statistics, extracted from spectral and modulation-based representations, provide a compact and robust description of the perceptual identity of textures despite local temporal variability. Crucially, the perceptual naturalness of synthetic sound textures depends on how well these statistics preserve the higher-order structure of real-world sounds (McDermott and Simoncelli, 2011; Norman-Haignere and McDermott, 2018).

Recent neuroimaging and electrophysiological studies (Zhai et al., 2020; Peng et al., 2024) suggest that these time-averaged summary statistics may be computed and represented in midbrain ensembles, particularly the inferior colliculus, along the ascending auditory pathway and subsequently utilized in auditory cortical areas for sound object detection and categorization (Lewis et al., 2009; Leaver and Rauschecker, 2010; Norman-Haignere and McDermott, 2018; Landemard et al., 2021). Within the auditory cortex, a functional hierarchy has been observed: primary regions are tuned to low-level acoustic features such as spectral energy and

temporal modulations, while nonprimary regions respond to higher-order statistical regularities, including cross-correlations in cochlear and modulation channels (Leaver and Rauschecker, 2010; Norman-Haignere and McDermott, 2018). Notably, Norman-Haignere and McDermott (2018) found that nonprimary auditory cortex exhibits stronger responses to natural textures than to synthetic counterparts that preserve low-level features but lack higher-order structure. Similar results were reported in a visual study by Freeman et al. (2013), which found differential neuronal responses in areas V1 and V2 of anesthetized macaque monkeys and humans to naturalistic texture images that preserved higher-order dependencies, compared to spectrally matched noise. While V1 neurons responded similarly to both types of stimuli, V2 neurons showed significantly stronger responses to naturalistic texture images. These findings suggest that cortical sensitivity to statistics both in auditory and visual textures follow a hierarchical organization. However, in natural sound textures, low- and high-order statistics typically co-vary, making it difficult to isolate their individual contributions to perceptual processing. In this study, we aimed to probe cortical sensitivity to texture naturalness, defined as the degree to which a synthetic sound captures the full statistical structure of a real-world texture. Using a sound texture synthesis model (McWalter & McDermott, 2019), we parametrically varied the statistics while preserving the categorical identity, generating sounds that ranged from spectrally matched noise to highly naturalistic exemplars. This design allowed us to assess how graded changes in statistical complexity influence perception and neural responses measured by functional magnetic resonance imaging (fMRI).

While most prior work has focused on the ascending auditory system, emerging evidence suggests that regions beyond auditory cortex, particularly within the medial temporal lobe (MTL), may also contribute to auditory pattern analysis (Billig et al., 2022). This raises the possibility that non-primary auditory cortex interacts with memory-related structures, such as the hippocampus, during the analysis of higher-order acoustic structure. Here, we tested this hypothesis by examining functional connectivity between auditory cortex and MTL during the perceptual evaluation of sound texture naturalness.

During scanning, participants performed an active comparison task, listening to pairs of synthetic textures ("reference" and then "test") that either matched or differed in their level of naturalness. Our primary aim was to investigate how cortical areas, particularly those beyond auditory cortex, contribute to perceptual realism derived from summary statistics. We found that both primary and non-primary auditory regions showed graded sensitivity to texture naturalness. Moreover, MTL structures, especially the hippocampus and entorhinal cortex, were modulated by naturalness levels, suggesting a role in top-down processes that support the recognition and evaluation of texture realism. These results support a distributed cortical mechanism, spanning auditory and memory-related regions, that tracks statistical structure to support the perception of natural textures.

Methods

Participants

Thirty individuals (aged 18-40 years, median 26 years, 10 males, 20 females) with self-reported normal hearing took part in the experiment. Informed consent was obtained from all participants, who were compensated for their time. The study was approved by the University College London Research Ethics Committee.

Stimuli

Sound texture stimuli were synthesized based on an auditory texture model (McDermott and Simoncelli, 2011). Texture statistics ($\xi_{texture}$) were first measured in 7-s original recordings of real-world sound textures. The statistics include marginal moments of each cochlear envelope (the mean, coefficient of variation, skew and kurtosis), the pairwise correlation coefficient between cochlear envelopes, the power in each modulation band, and pairwise correlations across modulation bands. With the same procedure, statistics (ξ_{noise}) of 7-s Gaussian noise were also calculated. The cochlear envelope means in the noise statistics, ξ_{noise}, were replaced with those of the texture statistics, $\xi_{texture}$, in order to avoid spectral discontinuities in the stimuli. Texture morphs with different degrees of naturalness were generated by synthesizing a texture from a point along a line in the space of statistics between the original texture and spectrally matched noise ($\xi_{synthesis} = \Delta\xi_{texture} + (1-\Delta)\xi_{noise}$), using 3-s samples of Gaussian noise as a seed (McWalter and McDermott, 2018). Four logarithmic-scaled levels of naturalness Δ were selected (0.25, 0.4, 0.63, 1.00). For each level, several different exemplars were synthesized using different random seeds. All synthesis parameters were set as in (McWalter and McDermott, 2018). The 2-s segments (with random beginning point between 0-1s) were selected from synthesized texture sounds. The root mean square (RMS) of each sound was normalized to a desired level of 0.1, and then the long-term average loudness of each sound was adjusted to 22 sones using the Glasberg and Moore method (Moore et al., 2016). A 10-ms cosine ramp was then applied to the beginning and end of each sound. Sounds were delivered using a bespoke high-fidelity delivery system based on a manifold of parallel piezoelectric transducers. An equalizer filter was applied (prior to RMS normalization) to compensate for sound delivery system attenuation.

Procedure

The experiment used a match-to-sample task (Figure 1A). Each trial began with a 1-second visual alert of "Sounds to start" displayed on the screen, followed by the presentation of the reference and test sounds with an interstimulus interval of 2-4 s while fixation cross displayed on the screen. After the sounds, a message "Same or different?" appeared on screen. Participants were instructed to respond as quickly as possible without making mistakes within a 3-second time window, using one of two available buttons on a keypad. They had to decide whether the reference and test sounds were the same or different. Each trial ended with a resting period of variable length, jittered between 2 and 4 s during which a fixation cross was displayed on the screen.

Trials were presented in 5 runs of about 10 min, each containing 52 trials, totalling 260 trials. Between runs, participants were permitted a short rest but were required to remain still. Four texture sounds ("Birds," "Rain," "Fire," and "Insects") were used. In each run, trials were divided into 13 groups, each containing all four texture types. Within each group, the order of the textures was randomized.

Different naturalness levels (Δ) were used for the reference and test sounds in 75% of the trials ("different" trials), while identical levels were used in the remaining 25% ("same" trials). The order of different and same trials was randomized. An equal number of each morph level was utilized in both different and same trials. In different trials, the number of trials with a greater Δ for the reference and test sounds was counterbalanced, and the order of these trials was randomized.

Prior to the main experiment, participants completed a practice session outside the scanner room consisting of 16 trials with different sound exemplars than those used in the main experiment, intended to familiarize them with the task. The experiment was controlled using Psychtoolbox and custom MATLAB code.

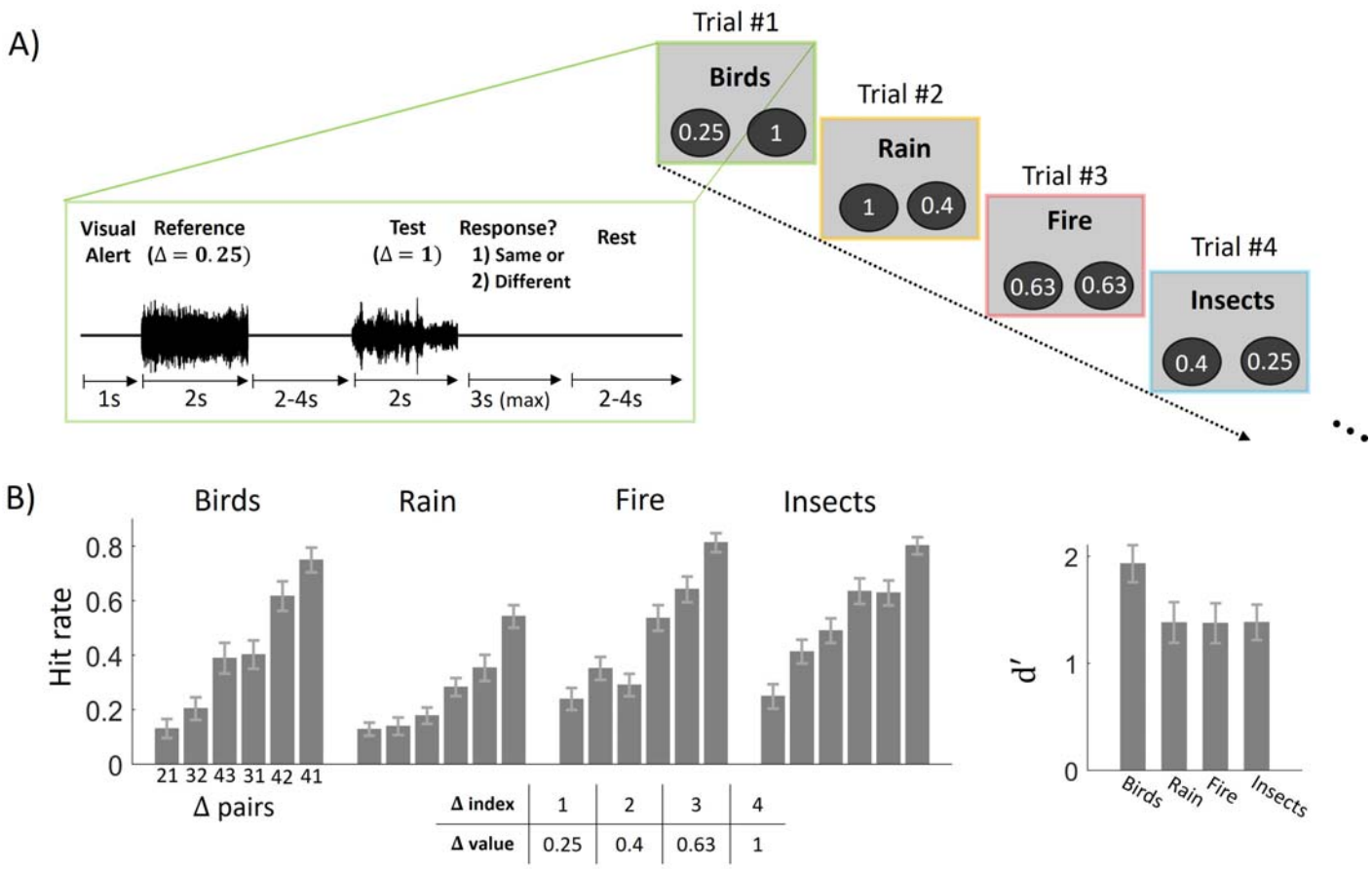

Figure 1. Texture perception paradigm and discrimination performance across participants. A) Each trial consisted of two sounds—a "reference" and a "test"—that either shared the same or differed in their level of naturalness, quantified by Δ: [0.25, 0.4, 0.63, 1]. These sounds were separated by a short interstimulus interval. Four types of natural texture sounds were used: Birds, Rain, Fire, and Insects. B) Discrimination performance (measured as hit rate) improved as the difference between Δ values increased. The x-axis indexes correspond to specific Δ pairs arranged in order of increasing change in statistics; for example, index "21" indicates the

reference sound with $\Delta = 0.4$ and the test sound with $\Delta = 0.25$. The right panel shows the d′ score for each texture sound. Error bars represent the standard error of the mean.

MRI data acquisition

Data were acquired using a 3-Tesla Siemens MAGNETOM Prisma MR scanner (Siemens Healthcare) with a 32-channel receiver head coil at the Department of Imaging Neuroscience (London, UK). Whole-brain T2*-weighted functional images were acquired using 3D echo-planar imaging (EPI) with field of view 192 x 192 x 144 mm, voxel size 3 mm isotropic, 48 partitions, 3D acceleration factor 2, repetition time (TR) 44ms, volume acquisition time 1.056 s, anterior-to-posterior EPI phase encoding, and bandwidth 2298 Hz/Px. Field maps (short echo time (TE) = 10.00 ms, long TE = 12.46 ms) and a whole-brain T1-weighted anatomical image (MPRAGE, field of view 256 x 256 x 176 mm, voxel size 1 mm isotropic, GRAPPA acceleration factor 2, TR = 2530 ms, TE = 3.34 ms) were acquired during the same scanning session.

During scanning sessions peripheral measurements of subject pulse rate and respiratory signals were made together with scanner slice synchronisation pulses using the Spike2 data acquisition system (Cambridge Electronic Design Limited, Cambridge UK). The cardiac pulse signal was measured using an MRI compatible pulse oximeter (Model 8600 F0, Nonin Medical, Inc. Plymouth, MN) attached to the subject's left index finger. The respiratory signal, thoracic movement, was monitored using a pneumatic belt positioned around the abdomen close to the diaphragm.

A physiological noise model was constructed to account for artifacts related to cardiac and respiratory phase and changes in respiratory volume using an in-house developed Matlab toolbox (Hutton et al., 2011). This resulted in a total of 14 regressors which were sampled at a reference slice in each image volume to give a set of values for each time point. The resulting regressors were included as confounds in the first level analysis for each subject.

To ensure that sounds were presented at a comfortable listening level, we conducted a calibration test inside the MRI scanner with an EPI sequence running, prior to the main experiment. Participants listened to a continuous sound composed of a mixture of synthesis textures, presented initially at a baseline sound pressure level (SPL) of 68 dB. They were instructed to press a button to gradually increase the volume until they reached a level that they found comfortable, while ensuring that all texture sounds remained clearly audible. The average SPL across participants was 82 dB.

MRI preprocessing

MRI data were processed in SPM12 (Department of Imaging Neuroscience, UCL). Each participant's functional images were unwarped using their field map and realigned to the first volume. Functional and anatomical images were coregistered to the mean functional image. For the univariate analyses, the coregistered images were normalised to the SPM12 template (avg305T1) in MNI space and the functional images were spatially smoothed using with a Gaussian kernel with 8 mm full-width at half-maximum (FWHM).

After the data were preprocessed using SPM, the artifact detection (ART) toolbox, implemented in CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012), was used to detect outliers from smoothed images. Time points in the series were marked as outliers if the global signal exceeded 3 standard deviations and the movement exceeded 1 mm. Additionally, five covariates generated using the aCmpCorr method (Behzadi et al., 2007), which uses principal component analysis on the measurements made in the white matter (WM) and cerebrospinal fluid (CSF) of each individual subject's segmented WM and CSF masks, were included as confounds in the first level analysis for each subject.

Univariate fMRI Analysis

To examine how BOLD activity varied with the naturalness of sound textures, we conducted a whole brain voxel-wise univariate analysis using a general linear model in SPM12. At the first level, each participant's time series was modelled with separate regressors for each texture category (Birds, Rain, Fire, and Insects) at both reference and test sound positions. This allowed us to control for potential category-specific differences while examining naturalness effects across all textures. For each regressor, we also included a parametric modulator corresponding to the level of naturalness (expressed as Δ, the proportion of the texture statistics: 0.25, 0.4, 0.63, and 1.0). All regressors were convolved with the canonical hemodynamic response function. Data were high-pass filtered (cut-off: 1/128 Hz) to remove low-frequency drifts. First-level contrasts of the main effect of the parametric modulation of naturalness were estimated for each participant. These contrast images were then entered into second-level group analyses using a random-effects model. To identify regions showing a positive linear relationship with naturalness, we conducted a one-sample t-test on the parametric modulation contrast across participants. Group-level statistical maps were thresholded using a cluster-level family-wise error (FWE) correction applied at $p < 0.05$ to control for multiple comparisons. Anatomical labelling of auditory subregions was based on the Julich-Brain Atlas (Amunts et al., 2020) (Figure 2).

Psychophysiological Interaction (PPI) Analysis

We used the psychophysiological interaction (PPI) method, as implemented in the CONN toolbox, to test whether connectivity between auditory subregions and medial temporal lobe (MTL) areas was modulated by texture naturalness. We defined seed regions in auditory regions of interest (ROIs; Te1.0, Te1.1, Te1.2, Te2.1, Te2.2, Te3, STS1, STS2) according to the Julich-Brain Atlas (Amunts et al., 2020), and for the MTL, we included the entorhinal cortex, hippocampus, parahippocampal gyrus, and amygdala for each hemisphere. PPI analysis was performed for each ROI pair (ROI-to-ROI analysis) for every subject. The effect of movement and physiological parameters on the BOLD signal was reduced by regressing out motion and physiological artifacts, along with their first-order temporal derivative, and covariates generated using the aCmpCorr method from WM and CSF by running whole-brain voxel wise regression.

The analysis was carried out separately for the reference and test sounds, based on the hypothesis that connectivity patterns might vary depending on the phase of sound presentation (i.e., early vs. late perceptual processing) such as MTL regions being involved in mnemonic processing of naturalistic textures (e.g. comparison of sounds to reference during retrieval). At the group level, one-sample t-tests (two-tailed) were

conducted with the contrast averaged across all texture categories to identify brain regions showing significant modulations ($p < 0.05$, connection-level and ROI-level FDR-corrected) in connectivity between auditory regions and MTL regions as a function of increasing or decreasing texture naturalness.

Code and data availability

All fMRI and behavioural data are available *here*.

Results

Behavioural performance

In this experiment, participants were presented with two sound texture morphs separated by a short interval. Each pair consisted of sounds with either the same or differing degrees of naturalness, that is constructed based on the same or differing proportions of natural sound statistics (Figure 1A). Participants were asked to judge whether the two sounds were the same or different on this dimension. Discrimination performance was poor when the naturalness levels were close (e.g., when the reference and test sounds had Δ values of 0.25 and 0.4, or vice versa), and improved as the difference in naturalness between the two sounds increased (Figure 1B). This pattern of performance was consistent across all texture categories.

Neural responses to texture naturalness

Univariate results revealed significant positive parametric modulation of texture naturalness in bilateral primary and non-primary auditory cortex, including subregions Te1.0, Te1.1, Te1.2, Te2.1, Te2.2, and Te3 during reference and test sound presentation (Figure 2). No significant negative parametric effects (i.e., stronger activity for less natural textures) were observed in auditory cortex. In addition to auditory cortex, we observed bilateral clusters in medial temporal lobe regions, including the entorhinal cortex, which showed increasing activity with naturalness during test sound presentation at an uncorrected threshold ($p < 0.001$) (Figure 3).

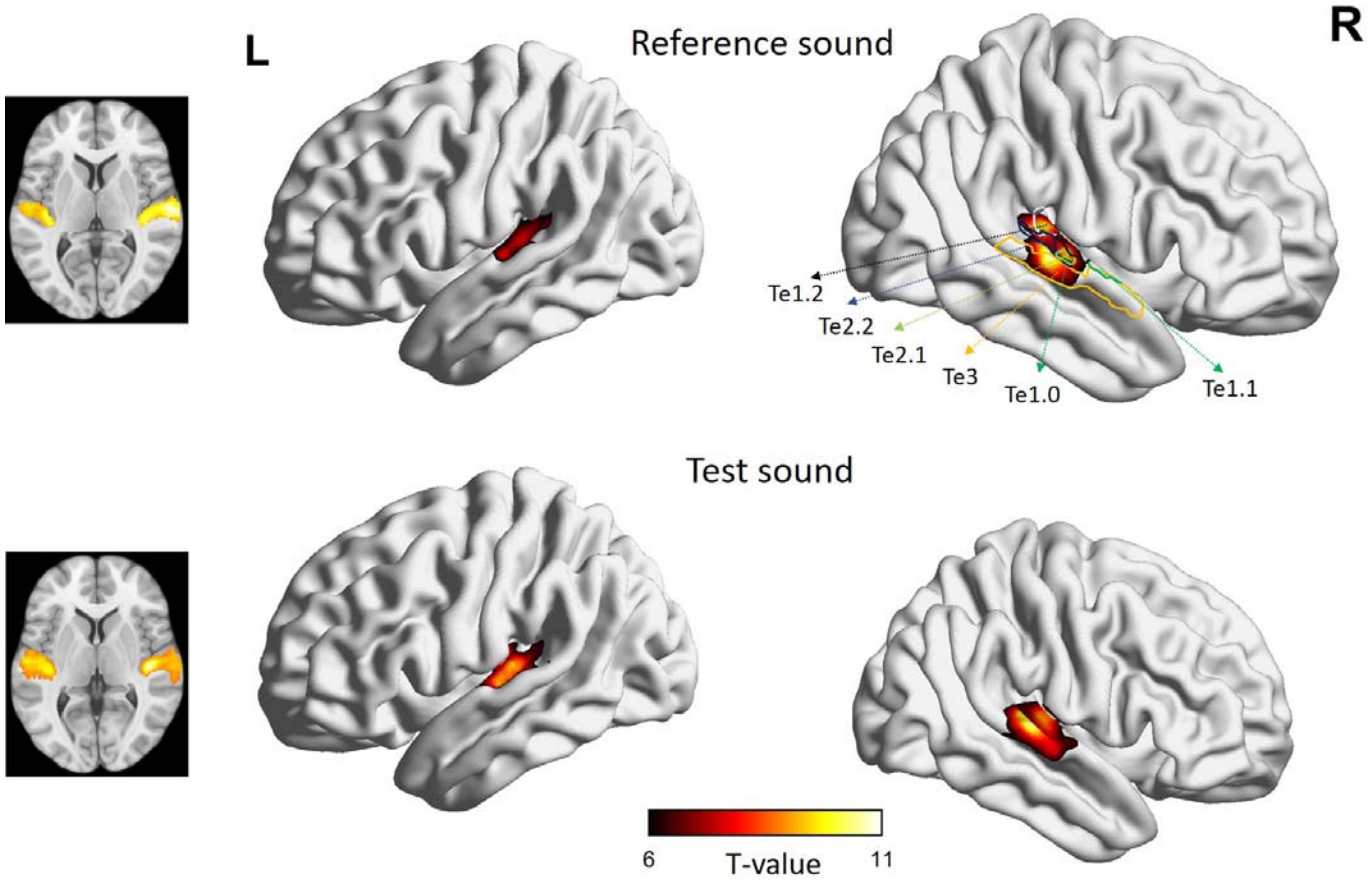

Figure 2. Auditory cortical responses to texture naturalness during reference and test sounds (at $p<0.05$, FWE-corrected). BOLD activity was measured while participants listened to pairs of synthetic textures varying in naturalness. During both the reference and test intervals, bilateral auditory cortex showed increased activation with higher texture naturalness.

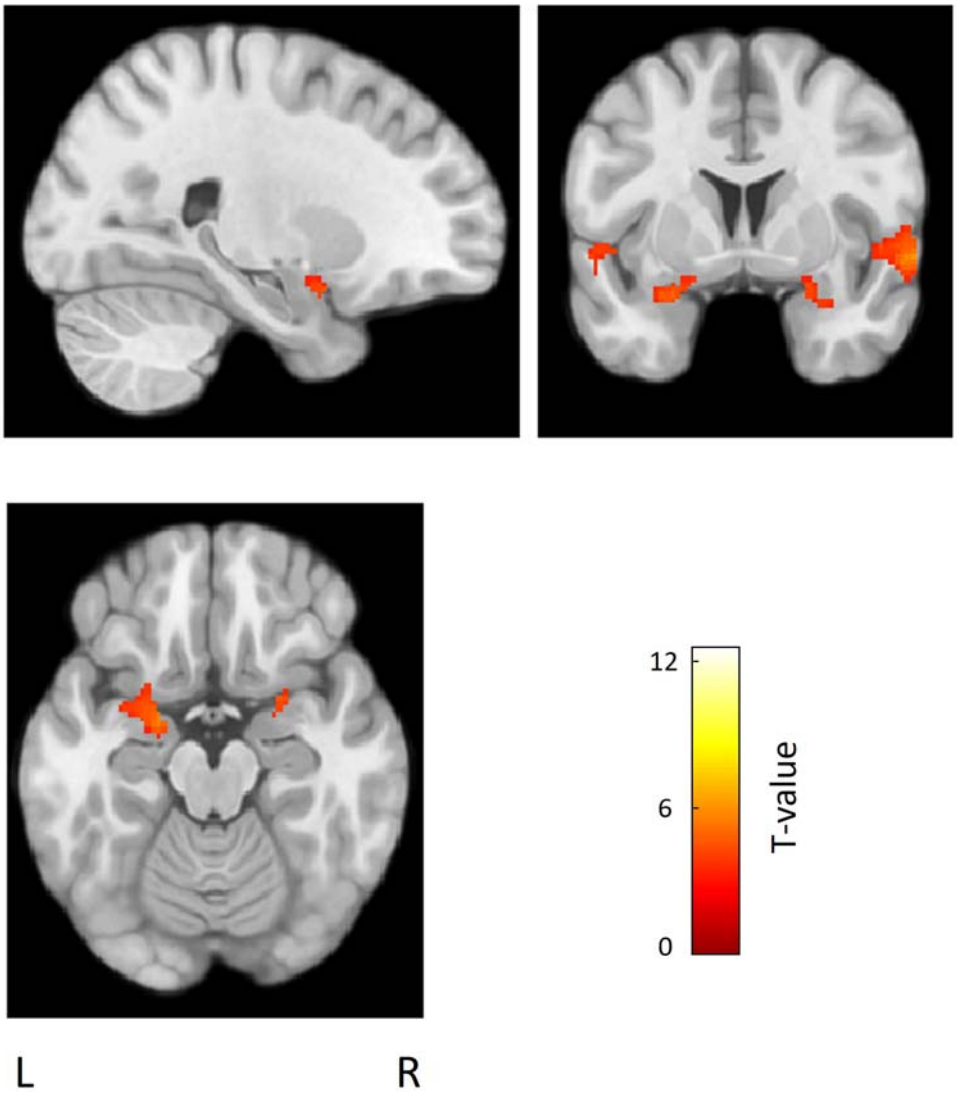

Figure 3. Medial temporal lobe responses to texture naturalness ($p < 0.001$, uncorrected). BOLD responses were measured while participants compared pairs of synthetic textures (a reference and a test) that varied in statistical naturalness. Activity in the entorhinal cortex was selectively modulated by the naturalness of the test sound.

PPI Analysis

The PPI results revealed a significant negative modulatory effect of naturalness exclusively in the right hemisphere only during the reference sound presentation (Figure 4, Table 1). Specifically, as the sound texture became unnatural (i.e., lower levels of statistical fidelity), the functional connectivity between the right hippocampus and right higher-order auditory areas (Te1.2, Te2.1, Te2.2, and Te3) increased ($p < 0.05$, FDR-corrected). Notably, this effect was evident in both seed configurations: significant connectivity was observed when the right hippocampus was used as the seed and when right auditory subregions (Te1.2 and Te3) were used as seeds. This bidirectional pattern suggests that the functional connectivity between hippocampus and auditory regions is robust across different seed choices. To determine whether sensory-hippocampus

connectivity was specific to the auditory cortex, a separate PPI analysis was conducted using visual regions (inferior, middle, and posterior occipital gyri) as seeds and MTL regions as targets, and vice versa. No significant modulatory effect of naturalness on connectivity between visual and MTL areas was observed.

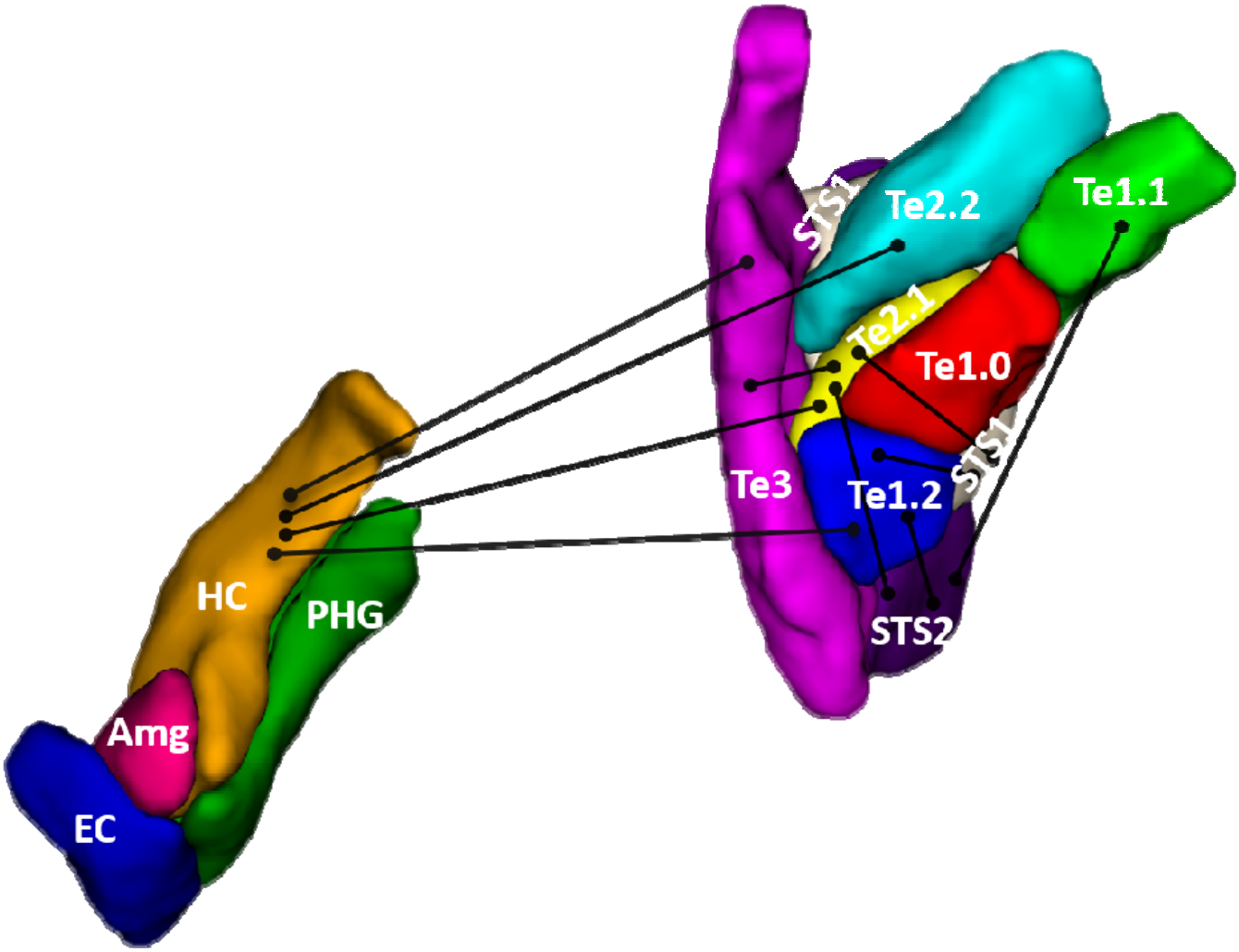

Figure 4. Group-level psychophysiological interaction (PPI) results showing connectivity modulated by texture naturalness ($p < 0.05$, FDR-corrected). Results are displayed on a 3D reconstructed surface of the right hemisphere, highlighting anatomical regions of interest, visualised using ITK-SNAP (*www.itksnap.org*, Yushkevich et al., 2006). (HC = hippocampus, Amg = amygdala, EC = entorhinal cortex, PHG = parahippocampal gyrus, STS = superior temporal sulcus).

Table 1. Significant connections between auditory regions and hippocampus modulated by naturalness

Seed Region	Target Region	T-value	P-value (FDR-corrected)
HC	Te1.2	-3.00	0.009
Te1.2	HC	-3.22	0.009
HC	Te2.1	-2.67	0.012
HC	Te2.2	-2.84	0.01
HC	Te3	-2.99	0.009
Te3	HC	-3.47	0.006
STS1	Te2.1	-3.82	0.004
STS1	Te1.2	-2.96	0.009
STS2	Te2.1	-3.91	0.004
STS2	Te1.2	-2.78	0.01
STS2	Te1.1	-3.09	0.009
Te3	Te2.1	-2.85	0.01

HC = hippocampus, STS = superior temporal sulcus

Discussion

In this study, we investigated neural responses to sound textures with different degrees of naturalness. Our results demonstrate that texture naturalness modulates both local activity in auditory cortex and large-scale connectivity between auditory and medial temporal lobe regions. As texture naturalness increased, we observed stronger responses in bilateral auditory cortex as well as in medial temporal regions, particularly the entorhinal cortex. Hippocampal connectivity with auditory cortex in the right hemisphere was enhanced when textures were less natural.

In this study, sound texture naturalness was defined based on the principle that perceptual realism emerges from the preservation of higher-order statistical regularities characteristic of natural sounds (McDermott and Simoncelli, 2011; McDermott et al., 2013; Norman-Haignere and McDermott, 2018; McWalter and McDermott, 2019) and natural images (Portilla and Simoncelli, 2000; Freeman et al., 2013). By parametrically varying these summary statistics along a continuum from spectrally matched noise to naturalistic textures, we found that activity in both primary and non-primary auditory regions increased proportionally with texture naturalness. This graded response suggests that auditory cortex is not merely detecting the presence or absence of structure but is instead sensitive to the degree of statistical regularity, tracking how closely a synthetic texture approximates the properties of real-world sounds.

Interestingly, while auditory cortical activity tracked naturalness across both reference and test sounds, activity in the medial temporal lobe, particularly entorhinal cortex (EC), was modulated only by the

naturalness of the test sound. That this modulation was parametric, tracking perceptual variation in naturalness, suggests that EC may contribute to the comparison process itself, rather than reflecting retrieval alone. This pattern aligns with accounts proposing that the EC supports the integration of perceptual input with stored representations, particularly under conditions requiring fine-grained discriminations (Ferko et al., 2022). In our task, the EC may be specifically recruited during the test sound when participants must compare it to a previously heard reference, supporting a perceptual-mnemonic interface sensitive to the degree of naturalness (Baxter, 2009; Suzuki and Baxter, 2009).

Additionally, we found that naturalness negatively modulated functional connectivity between the right hippocampus and higher-order auditory areas during the reference sound. This suggests stronger hippocampal-auditory coupling when textures are more degraded or become unnatural, possibly reflecting top-down perceptual inference under uncertainty (Billig et al., 2022). Although the task includes a delay period following the reference sound, and the observed hippocampal-auditory connectivity may involve working memory processes (Kumar et al., 2016), the fact that this connectivity varies parametrically with naturalness suggests it reflects more than simple sensory encoding. Our findings thus support a model in which naturalness judgments emerge from interactions between bottom-up encoding of statistics in auditory cortex and top-down modulation from memory-related regions. Future research could explore how these dynamics change with learning or in perceptual disorders.

Overall, our findings highlight texture naturalness as a meaningful perceptual dimension grounded in summary statistical representations, and they identify distributed neural mechanisms, spanning auditory cortex and medial temporal lobe, that support its evaluation.

Conclusion

In conclusion, our findings show that the human brain tracks sound texture naturalness through both localized responses in auditory cortex and distributed interactions with memory-related regions. Auditory areas appear sensitive to time-averaged summary statistics that define the statistical structure of textures, while the medial temporal lobe, particularly the hippocampus, may support top-down processes that aid in recognizing and evaluating texture realism. These results point to a hierarchical, distributed system in which perceptual judgments of naturalness emerge from the integration of bottom-up sensory encoding and memory-based modulation.

References

Amunts K, Mohlberg H, Bludau S, Zilles K (2020) Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. Science (1979) 369:988–992 Available at: https://www.science.org/doi/10.1126/science.abb4588 [Accessed April 14, 2025].

Baxter MG (2009) Involvement of Medial Temporal Lobe Structures in Memory and Perception. Neuron 61:667–677.

Behzadi Y, Restom K, Liau J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37:90–101 Available at: https://pubmed.ncbi.nlm.nih.gov/17560126/ [Accessed June 12, 2025].

Billig AJ, Lad M, Sedley W, Griffiths TD (2022) The hearing hippocampus. Prog Neurobiol 218:102326 Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10510040/ [Accessed April 20, 2025].

Ferko KM, Blumenthal A, Martin CB, Proklova D, Minos A, Saksida LM, Bussey TJ, Khan AR, Köhler S (2022) Activity in perirhinal and entorhinal cortex predicts perceived visual similarities among category exemplars with highest precision. Elife 11:e66884 Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9020819/ [Accessed April 19, 2025].

Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. Nature Neuroscience 2013 16:7 16:974–981 Available at: https://www.nature.com/articles/nn.3402 [Accessed April 17, 2025].

Hutton C, Josephs O, Stadler J, Featherstone E, Reid A, Speck O, Bernarding J, Weiskopf N (2011) The impact of physiological noise correction on fMRI at 7 T. Neuroimage 57:101–112 Available at: https://www.sciencedirect.com/science/article/pii/S105381191100396X?via%3Dihub [Accessed May 23, 2025].

Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD (2016) A Brain System for Auditory Working Memory. Journal of Neuroscience 36:4492–4505 Available at: https://www.jneurosci.org/content/36/16/4492 [Accessed June 13, 2025].

Landemard A, Bimbard C, Demené C, Shamma S, Norman-Haignere S, Boubenec Y (2021) Distinct higher-order representations of natural sounds in human and ferret auditory cortex. Elife 10.

Leaver AM, Rauschecker JP (2010) Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. The Journal of Neuroscience 30:7604 Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC2930617/ [Accessed April 19, 2025].

Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA (2009) Human Cortical Organization for Processing Vocalizations Indicates Representation of Harmonic Structure as a Signal Attribute. The Journal of Neuroscience 29:2283 Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC2774090/ [Accessed April 19, 2025].

McDermott JH, Schemitsch M, Simoncelli EP (2013) Summary statistics in auditory perception. Nat Neurosci 16:493–498.

McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. Neuron 71:926–940.

McWalter R, McDermott JH (2018) Adaptive and Selective Time Averaging of Auditory Scenes. Current Biology 28:1405-1418.e10.

McWalter R, McDermott JH (2019) Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. Nature Communications 2019 10:1 10:1–18 Available at: https://www.nature.com/articles/s41467-019-12893-0 [Accessed April 19, 2025].

Moore BCJ, Glasberg BR, Varathanathan A, Schlittenlacher J (2016) A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. Trends Hear 20:2331216516682698 Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC5318944/ [Accessed April 24, 2025].

Norman-Haignere S V., McDermott JH (2018) Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. PLoS Biol 16:e2005127 Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC6292651/ [Accessed April 19, 2025].

Peng F, Harper NS, Mishra AP, Auksztulewicz R, Schnupp JWH (2024) Dissociable Roles of the Auditory Midbrain and Cortex in Processing the Statistical Features of Natural Sound Textures. Journal of Neuroscience 44 Available at: https://www.jneurosci.org/content/44/10/e1115232023 [Accessed April 19, 2025].

Portilla J, Simoncelli EP (2000) Parametric texture model based on joint statistics of complex wavelet coefficients. Int J Comput Vis 40:49–71 Available at: https://link.springer.com/article/10.1023/A:1026553619983 [Accessed April 19, 2025].

Suzuki WA, Baxter MG (2009) Memory, Perception, and the Medial Temporal Lobe: A Synthesis of Opinions. Neuron 61:678–679.

Whitfield-Gabrieli S, Nieto-Castanon A (2012) Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. Brain Connect 2:125–141 Available at: /doi/pdf/10.1089/brain.2012.0073?download=true [Accessed April 24, 2025].

Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31:1116–1128 Available at: https://pubmed.ncbi.nlm.nih.gov/16545965/ [Accessed April 17, 2025].

Zhai X, Khatami F, Sadeghi M, He F, Read HL, Stevenson IH, Escabí MA (2020) Distinct neural ensemble response statistics are associated with recognition and discrimination of natural sound textures. Proc Natl

Acad Sci U S A 117:31482–31493 Available at: https://pubmed.ncbi.nlm.nih.gov/33219122/ [Accessed April 19, 2025].