Decoding reveals the neural representation of held and manipulated musical thoughts

David R. Quiroga Martinez[a,b*], Gemma Fernandez Rubio[b],
Leonardo Bonetti[b,c,d], Kriti G. Achyutuni[a], Athina Tzovara[a,e,f], Robert T. Knight[a], Peter Vuust[b]

(a) Helen Wills Neuroscience Institute & Department of Psychology, University of California Berkeley, Berkeley, CA
(b) Center for Music in the Brain, Department of Clinical Medicine, Aarhus University and The Royal Academy of Music, Aarhus, Denmark
(c) Center for Eudaimonia and Human Flourishing, Linacre College, University of Oxford, Oxford UK
(d) Department of Psychiatry, University of Oxford, Oxford UK
(e) Institute of Computer Science, University of Bern, Bern, Switzerland
(f) Center for Experimental Neurology, Sleep Wake Epilepsy Center, NeuroTec, Department of Neurology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
*Corresponding author

Corresponding Author: David Ricardo Quiroga Martinez **e-mail:** dquiroga@berkeley.edu

Address: 210 Barker Hall, UC Berkeley, Berkeley CA, 94703

Classification: Biological Sciences

Keywords: Imagery, music, decoding, MEG

Significance statement

Imagining vividly a sequence of sounds is a skill that most humans exert with relatively little effort. However, it is unknown how the brain achieves such an outstanding feat. Here, we used decoding techniques and non-invasive electrophysiology to investigate how sequences of sounds are represented in the brain. We report that auditory regions represent the sensory properties of individual sounds while association areas represent melodies as abstract entities. Moreover, we show that mentally manipulating a melody changes its neural representation across the brain. Understanding auditory representations and their volitional control opens the path for future work on decoding of imagined auditory objects and possible applications in cognitive brain computer interfaces.

Abstract

Imagine a song you know by heart. With little effort you could sing it or play it vividly in your mind. However, we are only beginning to understand how the brain represents, holds, and manipulates these musical "thoughts". Here, we decoded listened and imagined melodies from MEG brain data (N = 71) to show that auditory regions represent the sensory properties of individual sounds, whereas cognitive control (prefrontal cortex, basal nuclei, thalamus) and episodic memory areas (inferior and medial temporal lobe, posterior cingulate, precuneus) hold and manipulate the melody as an abstract unit. Furthermore, the mental manipulation of a melody systematically changes its neural representation, reflecting the volitional control of auditory images. Our work sheds light on the nature and dynamics of auditory representations and paves the way for future work on neural decoding of auditory imagination.

Introduction

Imagine your friends throwing a birthday party for you. At the climax, you begin to hear the first sounds of a well-known tune. "Happy birthday to you…" they cheerfully sing while you blow the candles and slice a delicious cake. If you are like most people, you can vividly recall the tune that your friends sing for you. You may even recall the voice of a cherished friend or the crowd singing painfully out of tune. Yet, we are only beginning to understand how the brain represents, holds, and manipulates these musical thoughts.

Here, we consider two kinds of imagination: Recall and manipulation. During ***recall***, we accurately imagine previously known sounds. During ***manipulation***, we imagine a modified version of the original sounds. In the brain, recall engages a widespread network including superior temporal gyrus, motor cortex, supplementary motor area, thalamus, parietal lobe, and frontal lobe (1–11), while manipulation further involves the frontal and parietal lobes (12, 13). However, it is unclear how these regions represent imagined sounds. By ***representation*** we mean the neural activity patterns that distinguish an auditory object from others. Understanding neural representations is crucial for elucidating how the brain recreates and transforms auditory images in the mind's ear.

A powerful technique to study auditory representations is ***multivariate pattern analysis*** (MVPA) (14), where patterns of neural activity are used to decode features of mentally held objects. If neural signals carry object-specific information, decoding accuracy is different from chance. By inspecting decoding model coefficients, it is possible to identify the features of neural activity that underlie mental representations. Using similar techniques, fMRI studies showed sound-specific representations in primary and secondary auditory cortex (15–18) and frontoparietal association areas (19, 20) during maintenance in working memory and imagination. Other studies demonstrated imagined sounds decoding from scalp EEG (21, 22). However, no study has elucidated how imagined auditory representations are held and manipulated in the brain.

Here, we used MVPA of brain activity recorded with magnetoencephalography (MEG, **Fig. 1a**) to investigate how mentally held and manipulated short auditory sequences are represented in the brain. For each trial in the task, participants listened and then vividly imagined a short three-note melody (**Fig. 1b**). In the recall block, participants imagined the melody as presented, whereas in the manipulation block they imagined it backwards (e.g., A-C#-E becomes E-C#-A). After a delay, they heard a second melody, which was the same as the first one, its backward version or a totally different one. Participants

answered whether the second melody was the same as the first one or not (recall block) or the inverted version of the first one or not (manipulation block). Importantly, there were only two melodies to imagine in the task, which were backward versions of each other.

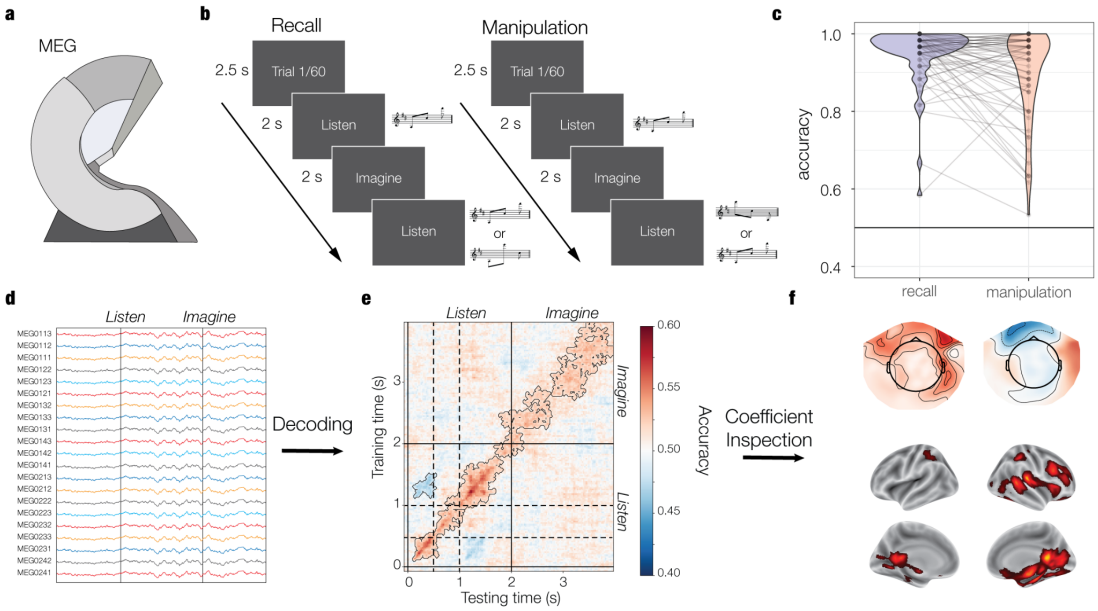

Figure 1. Materials and methods. We used MEG (a) to record participants' brain activity while they performed an imagery task (b). On each trial, participants heard and then imagined a short three-note melody. In the recall block, they imagined the melody as presented, while in the manipulation block, they imagined it backwards. Afterwards, they answered whether a test melody was the same as the first one (recall) or its backward version (manipulation). Participants performed with high accuracy (c) in both blocks. MEG signals (d) were used to decode melody identity. We used a time-generalization approach (e) in which models were trained at each time point of the training trials and tested at each time point of the test trials, resulting in time-generalized accuracy matrices. We transformed model coefficients into patterns of activation (f) and localized their brain generators. Dashed lines mark the onset of the second (0.5s) and third (1s) sounds of the melodies.

To understand neural auditory representations, we first used a time generalized decoding technique (14) to characterize their dynamics. Then, we assessed whether mentally manipulating the melodies changed their neural representation. In the manipulation block, participants suppressed the forward pattern and reinstated its backward version. Therefore, we predicted below-chance performance when training on manipulation and testing on recall and vice versa. Finally, we inspected model coefficients to identify the brain regions and neural activity features that discriminate between melodies and assessed how they changed between listening and imagination.

2. Results

2.1. Behavior

Participants (N = 71, 44 female, age = 28.77 +/- 8.43 SD) performed with high accuracy (**Fig. 1c**) and were better (*OR* = 1.85, CI = [1.25 - 2.72], *p* = .002) in the recall (96.7%, CI = [95.6 - 97.6]) than the manipulation (94.1%, CI = [91.9 - 95.8]) block. This confirms the presence of melody specific information during the imagination period. Incorrect trials were excluded from MEG analyses. Task performance was significantly associated with general working memory skills (Wechsler Adult Intelligence Scale - (WAIS; 23) (recall: r(69) = 0.3, p = .012; manipulation: r(69) = 0.29, p = .016; **Fig. S1**). No significant relationship with music training was found (Goldsmiths Musical Sophistication Index - GMSI; 24) (recall: r(69) = 0.13, p = .28; manipulation: r(69) = 0.23, p = .063; **Fig. S2**).

2.1. Above-chance decoding of listened and imagined melodies.

To investigate the neural dynamics of musical representations, we trained logistic regression models on MEG sensor data (**Fig. 1d**) to classify melody identity (melody 1: ***A-C#-E*** vs melody 2: ***E-C#-A***) at each time point of the trials. To assess whether representations recurred over time, we evaluated the models at each time point of the test data, resulting in time-generalized accuracy matrices (**Fig. 1e**). We used two types of testing: ***Within-condition*** (training and testing on recall trials or training and testing on manipulation trials) and ***between-condition*** (training on manipulation trials and testing on recall trials or training on recall trials and testing on manipulation trials). The latter aimed to reveal whether mentally manipulating the melodies changed their neural representations.

We observed above-chance within-condition decoding during listening and imagination for both recall and manipulation ($p < .001$; **Fig 2a**; see **Table S1** for full statistical report). This further confirms that mental representations were present during the delay period. Furthermore, we observed below chance performance when training around 0.3s and testing around 1.3s and vice versa reflecting the fact that the first sound (starting at 0s) in one melody was the third sound (starting at 1s) in the other melody ($p \leq .016$). This indicates that sound specific representations discriminated between melodies at these time points. Note that the second sound (C#) was always the same.

2.2. Volitional control over imagined melodies

Using between-condition testing, we found below-chance accuracy during the imagination period both when training on recall and testing on manipulation ($p \leq .048$) and when training on manipulation and testing on recall ($p \leq .018$; **Fig. 2b; Table** S**1**). In both cases, accuracies were lower for between-condition than within-condition testing ($p \leq .035$; **Fig. 2c**). This is indicative of a flip in neural representations such that models trained in one condition were consistently "wrong" when tested on the other condition.

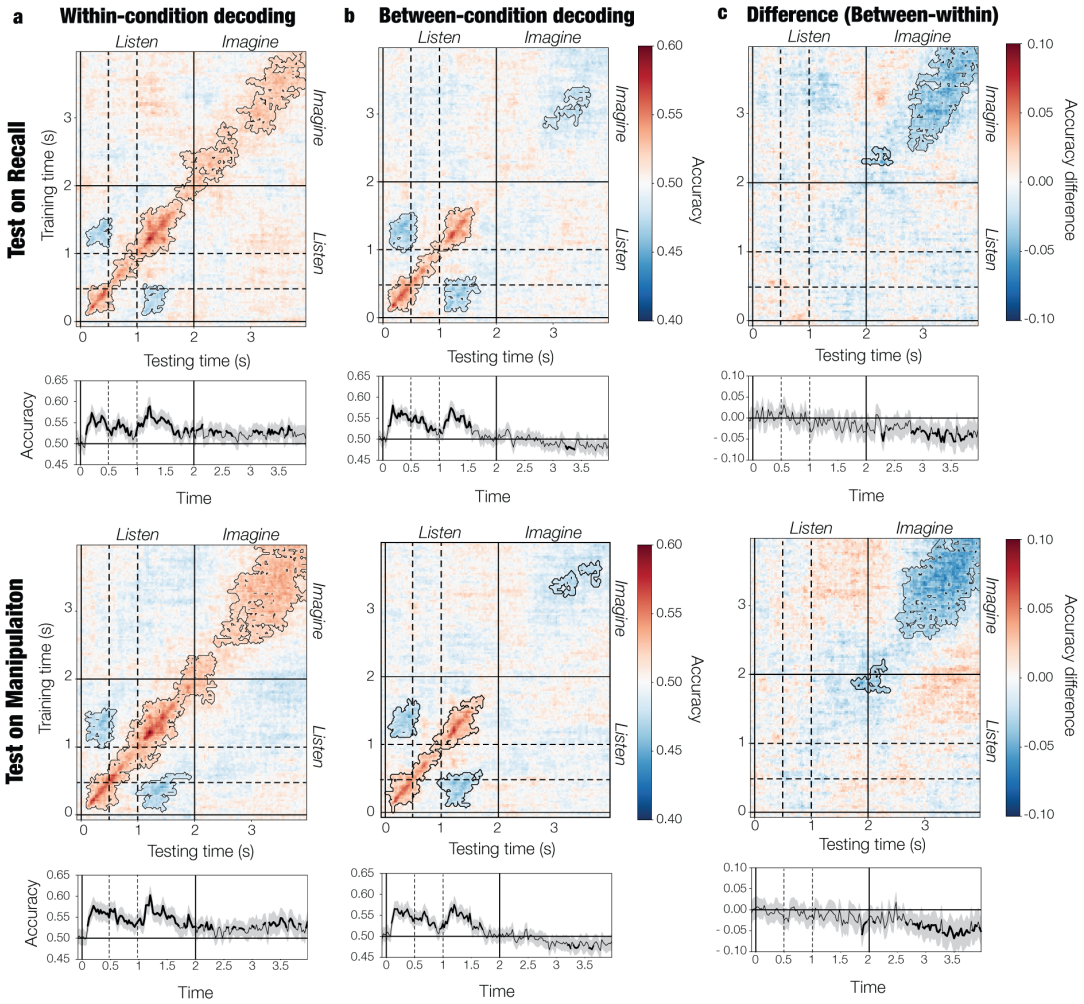

Figure 2. Accuracy for (a) within-condition decoding (train and test in the same condition), (b) between-condition decoding (train in one condition and test on the other) and the difference between the two (c). A time-generalization technique was used in which models were trained at each time point of the training data and tested at each time point of the test data. Accuracy across the diagonal is shown at the bottom of each plot. Contours and bold segments highlight significant clusters of above-chance or below-chance accuracy. Note how between-condition testing yields below-chance accuracy during imagination, suggesting a flip in neural representations. Dashed lines mark the onset of the second (0.5s) and third (1s) sounds of the melodies.

We also considered the possibility that representational dynamics were different between recall and manipulation. This is interesting because, during manipulation, the representation of the listened melody is meant to be the opposite of the imagined melody, potentially yielding different temporal generalization patterns. Indeed, when models were trained in the imagination period (~3.5 s) and tested on the listening period (~1.2 s) or vice versa, within-condition accuracy was lower ($p \leq .033$) for manipulation than recall (**Fig. S3**). This may reflect the fact that, for the manipulation block, the representation of the first melody was inhibited during the silent delay, thus leading to lower

generalization across listening and imagination. Overall, these findings indicate that, in the manipulation block, participants inhibited the listened melody and reinstated its backward version, resulting in a flip of neural representations. This is evidence of ***volitional control*** over mental auditory representations.

2.2. Musical sound sequences are represented in auditory, episodic memory and cognitive control networks.

To elucidate the brain regions and neural features that distinguish between the melodies, we transformed the model coefficients into interpretable patterns of activation (25) and localized their brain generators (**Fig. 1f**). The resulting activity patterns can be interpreted as the differences between melodies that underlie successful decoding. We focused on average brain activity at four different periods: Three during listening (0.2s - 0.5s, 0.7s - 1s and 1.2s - 1.5s) and one during imagination (2s - 4s). For listening, we chose 200 to 500 ms after onset of each sound, starting at accuracy peaks (0.2s and 1.2s) and including sustained activity until sound offset. For the imagination period, we included the whole time interval due to the lower signal to noise ratio, the lack of prominent peaks and the inherent temporal variability of mental images.

2.2.1. Auditory representations during listening

Patterns of neural activity distinguished between melodies in three main networks. For the first sound (**Fig. 3a**), we found clusters of regions in both conditions ($p \leq .006$, see **Table S2** for full statistical reports) with peak activity patterns in auditory areas such as superior temporal gyrus and Heschl's gyrus; but also, in object recognition and episodic memory areas in the inferior temporal, medial temporal and medial parietal cortex, including the fusiform, the hippocampus, the retrosplenial and the cingulate (see **Tables S3** and **S4** for a full report of anatomical regions). These ***auditory*** and ***episodic memory*** networks are likely involved in the encoding of sensory information and auditory objects. In addition, activity patterns in another cluster of regions in both blocks ($p < .001$) peaked at anteromedial (orbitofrontal, anterior cingulate), posteromedial (mid-posterior cingulate) and lateral (inferior, middle and superior frontal gyri) prefrontal cortex, as well as insula, basal nuclei (putamen, caudate, accumbens, pallidum) and the thalamus. These regions are typically involved in ***cognitive control*** (26, 27). Note that we include in this network the motor cortex (precentral gyrus) and some parietal regions detected together with the cluster of memory and auditory areas (angular gyrus, inferior parietal, postcentral gyrus), which also carried representations during the manipulation block. Moreover, we classify other structures detected together with the cognitive control cluster (entorhinal, amygdala, temporal pole, anterior hippocampus)

as part of the memory network described above. Interestingly, after the second sound melodies were discriminated in the episodic memory and cognitive control networks ($p < .005$), but not in auditory areas (**Fig. 3b**). This reflects the fact that the second sound is the same in both melodies, inducing similar sensory representations in superior temporal cortex while maintaining distinct melody-wise representations in episodic memory and cognitive control structures.

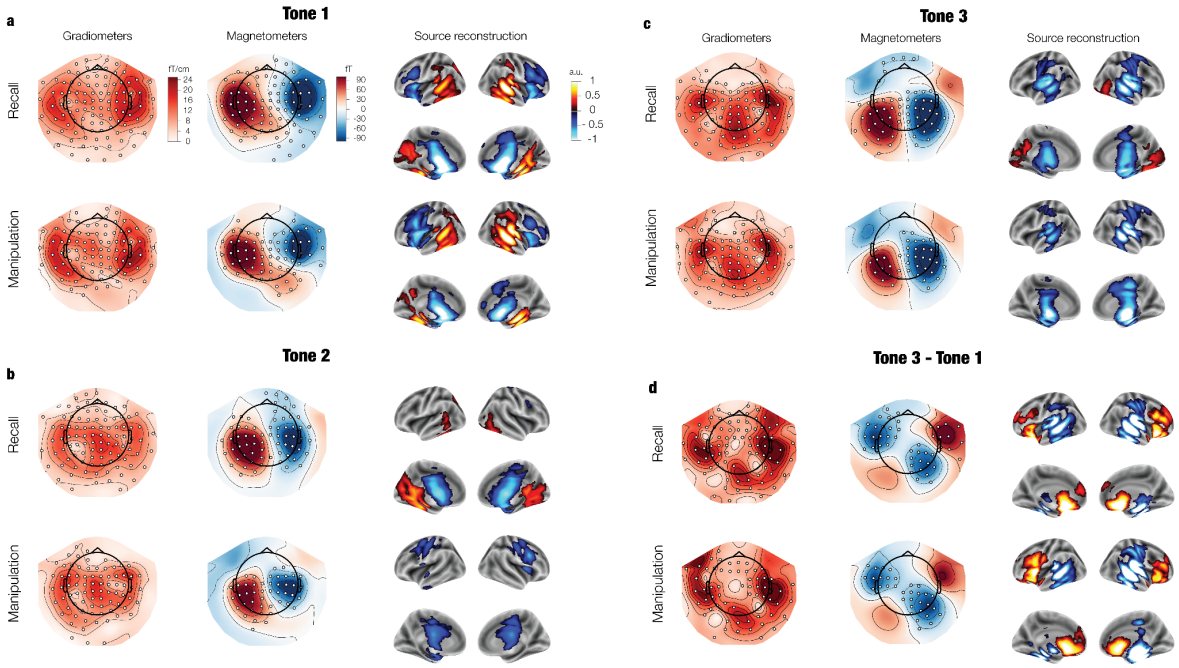

Figure 3. Patterns of neural activity that discriminate between melodies during listening, as derived from decoding coefficients and averaged over time. Patterns are shown for three time windows: (a) Tone 1 (0.2s - 0.5s), (b) Tone 2 (0.7s - 1s) and (c) Tone 3 (1.2s - 1.5s). The difference between tones 1 and 3 is also shown (d). Patterns are depicted for two types of MEG sensors (planar gradiometers and magnetometers) and after source reconstruction. For visualization, pairs of planar gradiometers were combined by taking their root mean square. Significant channels are highlighted with white dots. Source-level activation is shown for significant clusters. fT = "femto Tesla".

The same three networks represented the melodies after the third sound ($p \leq .021$; **Fig. 3c**). Crucially, representations flipped sign in auditory areas and anterior medial temporal areas ($p \leq .003$; **Fig. 3d**) such that, during sound one, melody 2 elicited more positive local field potentials than melody 1, whereas for sound three melody 1 elicited more positive local field potentials than melody 2 (**Fig. S4**). This representational flip underlies below-chance decoding after sounds 1 and 3 (**Fig. 4a**) and reflects the fact that the two melodies are backward versions of each other. In addition, representations in the

cognitive control network were more prominent after sound 1 ($p \leq .017$, **Fig. 3d**) than sound 3, possibly indicating a more automatic evaluation at the end than at the beginning of the sequence (27, 28). Overall, these pieces of evidence suggest two types of processing: One concerned with ***individual sound*** encoding in auditory and anterior memory regions and another one concerned with holding the melody ***as a sequence*** in posterior episodic memory and cognitive control areas.

2.2.1. Auditory representations during imagination

In the imagination period, melodies were mainly represented in structures of the cognitive control network, including basal nuclei, thalamus, mid-posterior cingulate, motor and parietal cortex ($p < .001$, **Fig. 4a**). Additional recruitment of the episodic memory network (inferior temporal cortex, posterior cingulate, precuneus) and auditory areas was observed in the recall block, and of the lateral prefrontal cortex in the manipulation block. Furthermore, representations changed in the left lateral prefrontal cortex during manipulation compared to recall ($p = .033$; **Fig. 4a**) with possible further changes in the right prefrontal cortex and retrosplenial (**Fig. S5**). These changes likely underlie the manipulation-driven representational flip identified through between-condition testing (**Fig. 2b**).

2.3. Opposite neural activity during listening compared to imagination.

Interestingly, patterns of activity switched sign ($p < .001$) between listening and imagination, with positive local fields in episodic memory areas becoming negative, and negative fields in cognitive control areas becoming positive during the silent delay (**Fig. 4b; Fig. S4**). A similar switch was reported in studies that decoded imagined sounds from scalp EEG (21, 22) and auditory working memory content with fMRI (16).

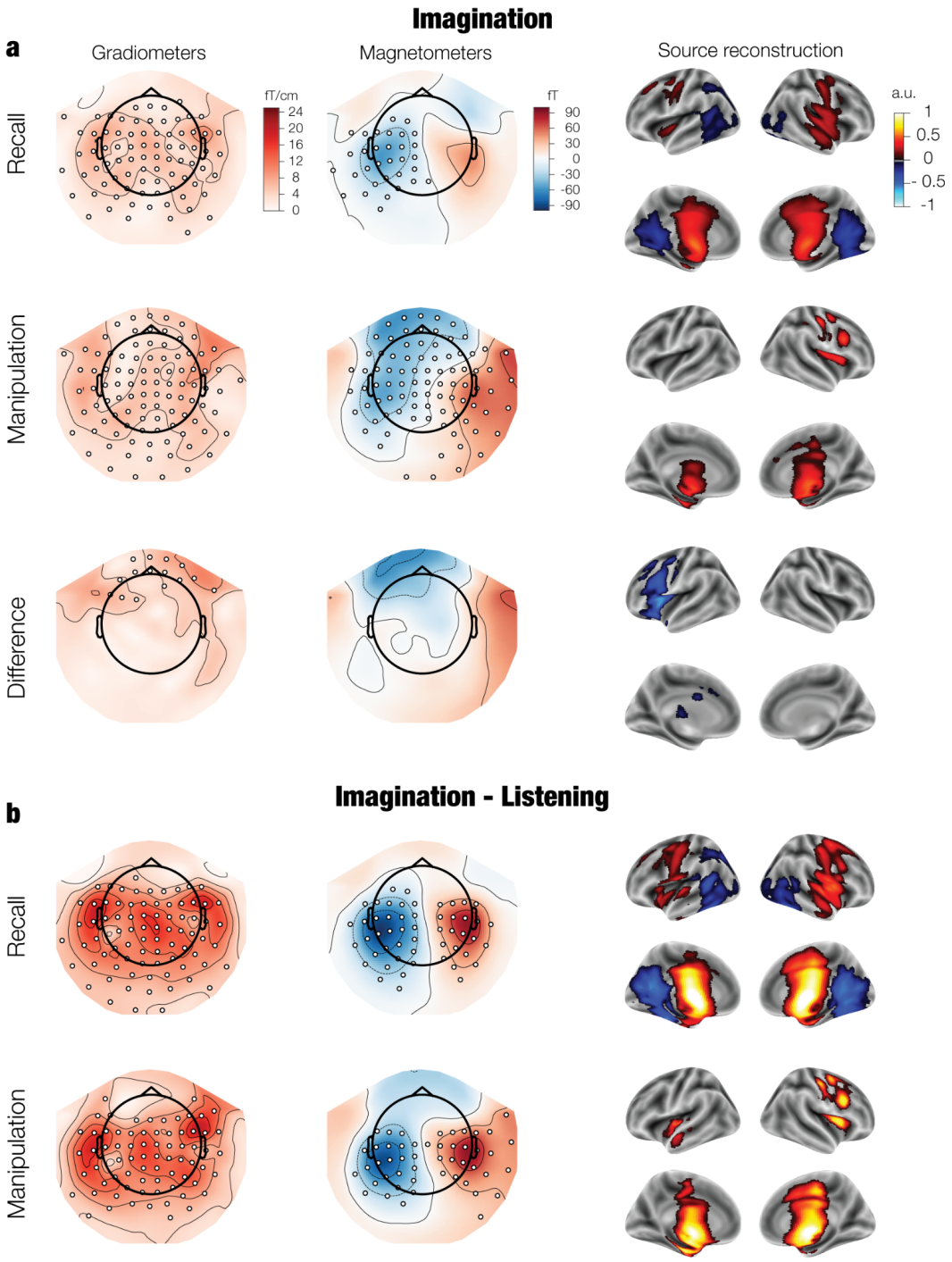

Figure 4. (a) Patterns of neural activity that discriminate between melodies during imagination (2s - 4s), as derived from decoding coefficients and averaged over time. The difference between listening and Imagination is also presented (b). Patterns are depicted for two types of MEG sensors (planar gradiometers and magnetometers) and after source reconstruction. For visualization, pairs of planar gradiometers were combined by taking their root mean square. Significant channels are highlighted with white dots. Source-level activation is shown for significant clusters. fT = “femto Tesla”.

Discussion

In this study, we decoded listened and imagined melodies from neural activity to demonstrate that musical sound sequences are represented in auditory, episodic memory and cognitive control networks, and that these representations systematically change when mentally manipulated. While previous studies have decoded imagined sound information from brain data (15, 17, 20–22), here we define the nature and dynamics of the underlying auditory representations and show how they change during manipulation.

Above-chance decoding peaked after the onset of the first and third sounds, and was sustained during imagination. The highest decoding performance was detected around the diagonal of the matrices, which indicates that representations were dynamic and had marginal generalization over time. During listening, this is probably due to the constantly changing sensory input. During imagination, this might reflect temporal variability between participants.

This lack of generalization further suggests that representations were different between listening and imagination. This contrasts with research suggesting that heard and imagined sounds share neural substrates and representations. For example, both imagined and actual sounds activate secondary auditory areas (10) and fMRI studies decoded imagined auditory representations from primary and secondary auditory cortex (15–18). Moreover, some studies found that representations during the omission of predictable sounds are similar to those of the actual sounds (29, 30).

The lack of generalization in our results might arise from three factors. First, the melodies differed in the temporal order of their constituent sounds, which were otherwise the same. Temporal order is an abstract feature that might generalize less across listening and imagination than sensory features such as pitch, which are typically the target of decoding. Second, experimental paradigms have either minimized the temporal variability of the imagined representations (e.g., omission studies) or didn't take time into account (fMRI studies). In contrast, our design allows temporal flexibility within a relatively long imagination period (2s), which might introduce between-trial and between-subject variations that blur sound-specific representations. Finally, consistent with previous EEG and fMRI findings (17, 21, 22), representations in the cognitive control and episodic memory networks were opposite between listening and imagination (**Fig. 3b**). This flip could reflect a change in the direction of information flow across the brain, from bottom-up (listening) to top-down (imagination). Some models propose specific roles for different layers of the cortical sheet, with superficial pyramidal cells conveying bottom up sensory input,

and deep pyramidal cells conveying top-down expectations (31, 32). It is possible that these layer-specific efferent and afferent activity patterns result in detectable changes in the direction of local field potentials. Layer-sensitive recordings are needed to test this hypothesis.

The fact that the melodies were backward versions of each other allowed us to dissociate the role of the three identified networks. During listening, auditory regions encoded the sensory information of individual sounds such that representations were opposite after the first and third tones, and equal during the second sound. In contrast, episodic memory areas (inferior and medial temporal lobe) and the ventral portion of the cognitive control network (ventromedial prefrontal cortex, thalamus, basal nuclei) remained stable, while representations in dorsal cognitive control areas (lateral prefrontal cortex) were involved only at sequence onset. Moreover, during imagination the cognitive control network was the main carrier of representations, with auditory and episodic memory areas further involved during recall; and lateral prefrontal cortex further engaged in manipulation. Overall, these dynamics suggest two types of processing: One concerned with the encoding of sound-specific sensory information in superior temporal cortex and anterior memory areas, and another one concerned with the encoding, retrieval, and manipulation of temporal auditory sequences in posterior episodic memory and cognitive control areas. This dissociation between the sensory and abstract properties of sound sequences is consistent with a previous scalp EEG study that disentangled pitch and temporal order representations during sound maintenance in auditory working memory (33).

The regions that carried auditory representations in our study overlap with those identified in previous neuroimaging activation studies as important for auditory imagery (1, 2, 4–11). However, a discrepancy of our study is the lack of substantial melody-specific information in the supplementary motor area, identified as a key region for auditory imagination (34). Nevertheless, we found representations in the motor and somatosensory cortex, which is consistent with previous reports (4, 35, 36) and might reflect the generation of auditory expectations through motor simulation. Furthermore, we observed melody-specific representations in the basal nuclei, a set of areas involved in both cognitive and motor control that have not been identified in previous fMRI research. From these nuclei, the putamen has been related to motor imagery (37). Moreover, we used MEG broadband signals to decode auditory representations, while BOLD signals correlate best with high gamma power (38), which might be why basal nuclei representations were found here but not in fMRI. Future research should examine high gamma activity and other frequency bands to elucidate their relationship with the BOLD signal and previous results.

The task used in this study is similar to the classical delayed match-to-sample paradigm employed in working memory research. An important difference, however, is that we asked participants to vividly imagine the melodies, whereas in working memory experiments maintenance strategies usually remain unspecified. Thus, while it is possible that some participants use imagery to hold information in working memory, this is not required in a classic delayed match-to-sample task. This caveat aside, task performance was associated with general working memory scores and the three networks identified overlap with those exhibiting delay-period activity in auditory working memory, including the auditory cortex, the prefrontal cortex, the parietal cortex, and the medial temporal lobe (12, 19, 39, 40). Moreover, auditory representations in working memory have been decoded from the auditory, frontal and parietal cortices (15, 19, 33) and from the functional interaction of these regions (41, 42). Most of these decoding studies, however, addressed working memory for individual sounds and none of them investigated sound manipulation. In addition, there is a tradeoff, with fMRI studies having good spatial but low temporal resolution, and EEG studies having good temporal but low spatial resolution. The use of MEG allowed a good localization of auditory representations both in space and time.

Despite the advantages of MEG, two methodological caveats need to be considered. First, we localized auditory representations to deep brain areas (basal nuclei, thalamus, hippocampus), which might raise concerns given the bias towards the head center of beamforming algorithms (43) and the fact that activity in such areas is typically hard to detect with MEG. However, we eliminated the depth bias by normalizing the forward and inverse solutions and verified that the localized activations are consistent with sensor topographies, especially at the midline (e.g., **Fig. 3c**). In addition, differences were still found in deep structures when two conditions were contrasted (e.g., imagination vs listening), arguing against a depth bias which should cancel out in condition contrasts. Furthermore, the detection of deep sources has become increasingly common, including the basal nuclei, the medial temporal lobe and even the cerebellum (44–49). Therefore, it is unlikely that these deep activity patterns are localization artifacts. The other caveat is the possibility that successful decoding is partly due to extracerebral, motion-related activity. However, this is also unlikely because we thoroughly cleaned the data from the main sources of contamination (eye movements and heartbeats), the sensor topographies suggest brain generators, and beamforming algorithms are particularly good at filtering out extracerebral sources.

Finally, our results demonstrate the feasibility of decoding mental auditory representations at a fine temporal resolution with non-invasive methods. This opens the path to clinical applications where decoding of imagined objects is relevant (e.g., communication impairments). Future work might use

larger models (e.g., deep learning) and different modalities (e.g., optical MEG, intracranial EEG) to maximize the decoding of auditory images.

3. Methods

3.1. Participants

We recorded MEG (**Fig. 1a**) data from 80 participants. From these, six were excluded due to chance performance in the task and three due to noisy or corrupted neural data. The final group of 71 participants (44 female, age = 28.77 +/- 8.43 SD) had mixed musical backgrounds with most of them (50) never having played a musical instrument (including voice). The other 21 participants had a median of 11 (IQR = [7-16]) years of musical training. In addition, participants had a median score of 17 (IQR = [13-26], maximum possible score = 49) in the training subscale of the Goldsmiths Musical Sophistication Index (GMSI; 24) and of 96 (IQR = [93-105]) in the Wechsler Adult Intelligence Scale (WAIS; 23). Participants gave their explicit informed consent and received a small monetary compensation. The study was approved by the Institutional Review Board (IRB) of Aarhus University (case number: DNC-IRB-2020-006) and conducted in accordance with the Helsinki declaration.

3.2. Stimuli

We employed short three-note melodies forming a major chord arpeggio in a piano timbre (musical pitch: A3, C#5, E6; F0: 220Hz, 554Hz, 1318Hz). The sounds were arranged in ascending order in melody 1 (A-C#-E) and descending order in melody 2 (E-C#-A). Two foil test melodies were also included: A-E-C# and E-A-C#. The inter-onset interval between tones was 500ms. The sounds were normalized to peak amplitude.

3.3. Task

On each trial (**Fig. 1b**), participants heard melody 1 or melody 2, together with the word “Listen” appearing on the screen. After 2s, participants saw the word “Imagine”, which indicated that they had to vividly reproduce the melody in their minds. There were two conditions. In the recall block, they imagined the melody as presented whereas in the manipulation block, they imagined it backwards. Four seconds after trial onset, participants heard a test melody, which could be the same as the first one, its inverted version or a different melody. Participants answered whether the second melody was the same as the first one or not (recall block) or its inverted version or not (manipulation block). A response time limit of 3.5 seconds was set. There were 60 trials per block (30 same/inverted, 30 different/other). The

trial number was displayed on the screen for 2.5s before trial onset. A quick pause was allowed after the 30th trial. Two practice trials were presented at the beginning of each block. Conditions were counterbalanced across participants.

3.4. Procedure

At the beginning of the session, we explained in detail the procedure to the participants and instructed them to vividly imagine the melodies without humming them or moving any part of the body. We made sure the participant fully understood the nature of the task and was able to perform practice trials correctly before the MEG recording. After giving informed consent, the participants changed into medical clothes, and we attached electrocardiogram (ECG) and electrooculogram (EOG) electrodes to their skin for heartbeat and eye movement monitoring. Head shape was digitized with a Polhemus system and head position was continuously tracked during the recording with the help of three coils. During the task, the participant sat in the MEG chair inside a magnetically shielded room and looked at the screen where instructions and trial information were displayed. The subjects responded to each trial by making a button press in a response pad with their right hand. Sound stimulation was delivered through magnetically isolated ear tubes. Other experimental paradigms testing recognition memory were recorded together with this task. Results are reported elsewhere (50). The order of the paradigms was counterbalanced across participants.

3.5. MEG recording and preprocessing

MEG data were collected with a 306-channel (102 magnetometers, 204 planar gradiometers) Elekta Neuromag system and Maxwell-filtered with proprietary software. This step also involved correcting the data for continuous head movements. Data analyses were conducted in MNE Python (v0.24) (51). Vertical and horizontal eye movements as well as heartbeat artifacts were corrected with ICA in a semi-automatic routine. Visual inspection was used to ensure data quality. After high-pass filtering (0.05 Hz cutoff), epochs were extracted from -0.1s to 4s around trial onset. For source reconstruction, T1 brain anatomical images were collected with a 3T MRI scanner and segmented and aligned with MEG sensors using Freesurfer. Source reconstruction was done for the 68 participants with an available MRI. Using the boundary element method and a single shell mesh (5 mm resolution), volumetric forward models were created and subsequently inverted with linearly constrained minimum variance (LCMV) beamforming employing gradiometer covariance. Importantly, forward models and inverse solutions were normalized to eliminate the bias towards the center of the head inherent to beamformers (43).

3.6. Decoding analysis

We used a time-generalized decoding approach (Fig. 1d) (King & Dehaene, 2014) based on L1 regularized logistic regression to classify melody identity (melody 1 vs melody 2) at each time point of the trials, for each participant separately. To assess the representational dynamics, we evaluated the models at each time point of the test data. We did two types of testing. In within-condition testing, we trained and tested the models with trials of the same condition. In between-condition testing, we trained the models with trials of one condition (e.g., manipulation) and tested on trials of the other (e.g., recall). Five-fold cross validation was used for within-condition testing. To avoid biases in model fitting due to class imbalances related to the exclusion of incorrect trials, we used a balanced scoring strategy in which the average accuracy was computed separately for each class and then combined across classes.

At the group level, we used cluster-based, permutation-corrected, two-sided one-sample t-tests, to evaluate whether accuracies in the time-generalization matrices were significantly above or below chance. Here, chance level corresponds to 0.5 accuracy, as we classified binary melody identity from brain data. We used a two-sided cluster defining threshold of $p = 0.05$ ($p = 0.025$ one-sided, $t > 1.99$) and max sum as the cluster statistic. The cluster-level significance threshold was set at $p = .05$. The number of permutations was 5000. The same statistical approach was used to evaluate whether within-condition accuracy was different from between-condition accuracy.

3.6. Coefficient inspection

We transformed decoding coefficients into patterns of activation for each participant (25) and localized their neural generators using the inverse solutions described above (**Fig. 1f**). For each voxel, the magnitude and sign of the orientation with maximum power were retained. For sensor activity patterns, we used cluster based permutations (see above) in the whole epoch (0-4) to test whether group-level activity patterns were different from zero. After projecting individual source-level time-courses into MNI standard space, we also tested against zero the localized patterns averaged across time in the three listening (0.2s – 0.5s, 0.7s – 1s, 1.2s – 1.5s) and one imagination (2s – 4s) time windows. For all these periods, differences between recall and manipulation were also tested. Furthermore, we compared patterns of average activity between the listening (0s – 2s) and imagination (2s – 4s) periods and between sounds 1 (0.2s – 0.5s) and 3 (1.2s – 1.5s). Using the Desikan-Killiany parcellation (52), we obtained the significant peak activation for each region that overlapped with significant clusters. We report the regions with the most prominent peaks.

Finally, in a supplemental analysis, we inspected the time courses of activity patterns in five regions of interest (ROI) including (1) right auditory, (2) right episodic memory, (3) dorsal cognitive control, (4) left dorsal cognitive control and (5) right ventral cognitive control areas (**Fig. S4**). We used cluster based permutations as described above to evaluate significant differences from zero. We display these patterns together with the evoked response calculated between –0.1s and 4s around trial onset, for each of the two melodies and the two conditions (**Fig. S4**). These evoked responses were source localized with the same inverse operator as the activity patterns derived from decoding and were subject to the same statistical tests.

3.7. Relationship with behavior: Exploratory analyses

For completeness, we investigated possible relationships between neural decoding accuracy and behavior. We used exploratory single-trial logistic mixed-effects models (R, lme4; , 53, 54) to evaluate whether participants' target melody identification predicted correct melody discrimination from neural activity at each time point of the trial. We did this analysis separately for the recall and manipulation blocks, and for the listening and imagination periods. We included participant as a random effect for both intercept and slopes. We obtained *p*-values for the relevant coefficients using the Satterthwaite approximation of degrees of freedom (55). The same analysis was done to predict neural accuracy from response times, as measured from the beginning of the test melody. No significant relationships were detected in these analyses, likely due to the ceiling effect in behavioral performance. See **Fig. S6** and **7** for the results.

Acknowledgements

We thank Francesco Carlomagno for his help with data collection and Ludovic Bellier for feedback. This work was supported by NINDS R37NS21135 (RTK), Brain Initiative (U19NS107609-03 and U01NS108916) (RTK), CONTE Center PO MH109429 (RTK), the Independent Research fund, Denmark (DQM), the Carlsberg Foundation (CF20-0239) (LB), Lundbeck Foundation (Talent Prize 2022) (LB), Linacre College of the University of Oxford (Lucy Halsall fund) (LB), Nordic Mensa Fund (LB), the Danish National Research Foundation (DNRF 117) (GFR, LB, PV) and the Interfaculty Research Cooperation "Decoding Sleep: From Neurons to Health & Mind" of the University of Bern and the Swiss National Science Foundation (#320030_188737) (AT).

CReDIT authorship contribution statement

David R. Quiroga-Martinez: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization, Investigation. Gemma Fernandez-Rubio: Data curation, Visualization, Formal analysis. Leonardo Bonetti: Conceptualization, Methodology, Data curation. Kriti G. Achyutuni: Data curation, Visualization. Athina Tzovara: Conceptualization. Robert T. Knight: Conceptualization, Supervision. Peter Vuust: Conceptualization, Supervision, Funding acquisition.

Code and data sharing

The code used to perform the analyses can be found in this location: https://github.com/drqm/working_memory.git

The data for this paper will be publicly shared upon acceptance in a peer-reviewed journal.

References

1. S. E. Bastepe-Gray, N. Acer, K. Z. Gumus, J. F. Gray, L. Degirmencioglu, Not all imagery is created equal: A functional Magnetic resonance imaging study of internally driven and symbol driven musical performance imagery. *Journal of Chemical Neuroanatomy* **104**, 101748 (2020).
2. N. Bunzeck, T. Wuestenberg, K. Lutz, H.-J. Heinze, L. Jancke, Scanning silence: Mental imagery of complex sounds. *NeuroImage* **26**, 1119–1127 (2005).
3. N. E. Foster, A. R. Halpern, R. J. Zatorre, Common parietal activation in musical mental transformations across pitch and time. *Neuroimage* **75**, 27–35 (2013).
4. A. R. Halpern, When That Tune Runs Through Your Head: A PET Investigation of Auditory Imagery for Familiar Melodies. *Cerebral Cortex* **9**, 697–704 (1999).
5. A. R. Halpern, R. J. Zatorre, M. Bouffard, J. A. Johnson, Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* **42**, 1281–1292 (2004).
6. S. C. Herholz, A. R. Halpern, R. J. Zatorre, Neuronal correlates of perception, imagery, and memory for familiar tunes. *J Cogn Neurosci* **24**, 1382–1397 (2012).
7. W. Huijbers, C. M. A. Pennartz, D. C. Rubin, S. M. Daselaar, Imagery and retrieval of auditory and visual information: Neural correlates of successful and unsuccessful performance. *Neuropsychologia* **49**, 1730–1740 (2011).
8. V. Pando-Naude, A. Patyczek, L. Bonetti, P. Vuust, An ALE meta-analytic review of top-down and bottom-up processing of music in the brain. *Sci Rep* **11**, 20813 (2021).
9. S.-S. Yoo, C. U. Lee, B. G. Choi, Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport* **12**, 3045–3049 (2001).
10. R. J. Zatorre, A. R. Halpern, D. W. Perry, E. Meyer, A. C. Evans, Hearing in the Mind's Ear: A PET Investigation of Musical Imagery and Perception. *Journal of Cognitive Neuroscience* **8**, 29–46 (1996).
11. M. Zvyagintsev, *et al.*, Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience* **37**, 1421–1434 (2013).
12. P. Albouy, A. Weiss, S. Baillet, R. J. Zatorre, Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. *Neuron* **94**, 193-206.e5 (2017).
13. R. J. Zatorre, Beyond auditory cortex: working with musical thoughts. *Annals of the New York Academy of Sciences* **1252**, 222–228 (2012).
14. J.-R. King, S. Dehaene, Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences* **18**, 203–210 (2014).
15. P. Deutsch, S. Czoschke, C. Fischer, J. Kaiser, C. Bledowski, Decoding of Working Memory Contents in Auditory Cortex Is Not Distractor-Resistant. *J. Neurosci.* **43**, 3284–3293 (2023).
16. A. C. Linke, A. Vicente-Grabovetsky, R. Cusack, Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences* **108**, 12961–12966 (2011).
17. A. C. Linke, R. Cusack, Flexible Information Coding in Human Auditory Cortex during Perception, Imagery, and STM of Complex Sounds. *Journal of Cognitive Neuroscience* **27**, 1322–1333 (2015).
18. M. Regev, A. R. Halpern, A. M. Owen, A. D. Patel, R. J. Zatorre, Mapping Specific Mental Content during Musical Imagery. *Cereb Cortex* **31**, 3622–3640 (2021).
19. S. Kumar, *et al.*, A Brain System for Auditory Working Memory. *Journal of Neuroscience* **36**, 4492–4505 (2016).
20. I. Uluç, T. T. Schmidt, Y. Wu, F. Blankenburg, Content-specific codes of parametric auditory working memory in humans. *NeuroImage* **183**, 254–262 (2018).
21. G. M. Di Liberto, G. Marion, S. A. Shamma, The Music of Silence: Part II: Music Listening Induces Imagery Responses. *J. Neurosci.* **41**, 7449–7460 (2021).

22. G. Marion, G. M. Di Liberto, S. A. Shamma, The Music of Silence. Part I: Responses to Musical Imagery Encode Melodic Expectations and Acoustics. *J. Neurosci.*, JN-RM-0183-21 (2021).
23. D. Wechsler, Wechsler Adult Intelligence Scale--Fourth Edition (2012) https:/doi.org/10.1037/t15169-000 (August 5, 2023).
24. D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE* **9**, e89642 (2014).
25. S. Haufe, *et al.*, On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **87**, 96–110 (2014).
26. F. McNab, T. Klingberg, Prefrontal cortex and basal ganglia control access to working memory. *Nat Neurosci* **11**, 103–107 (2008).
27. B. Voytek, R. T. Knight, Prefrontal cortex and basal ganglia contributions to visual working memory. *PNAS* **107**, 18167–18172 (2010).
28. G. Vishne, E. M. Gerber, R. T. Knight, L. Y. Deouell, Distinct ventral stream and prefrontal cortex representational dynamics during sustained conscious visual perception. *Cell Reports* **42**, 112752 (2023).
29. L. Chouiter, *et al.*, Experience-based Auditory Predictions Modulate Brain Activity to Silence as do Real Sounds. *J Cogn Neurosci* **27**, 1968–1980 (2015).
30. G. Demarchi, G. Sanchez, N. Weisz, Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nat Commun* **10**, 3440 (2019).
31. A. M. Bastos, *et al.*, Canonical Microcircuits for Predictive Coding. *Neuron* **76**, 695–711 (2012).
32. A. M. Bastos, M. Lundqvist, A. S. Waite, N. Kopell, E. K. Miller, Layer and rhythm specificity for predictive routing. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31459–31469 (2020).
33. Y. Fan, Q. Han, S. Guo, H. Luo, Distinct Neural Representations of Content and Ordinal Structure in Auditory Sequence Memory. *J. Neurosci.* **41**, 6290–6303 (2021).
34. C. F. Lima, S. Krishnan, S. K. Scott, Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends Neurosci* **39**, 527–542 (2016).
35. Y. Ding, *et al.*, Neural Correlates of Music Listening and Recall in the Human Brain. *J. Neurosci.* **39**, 8112–8123 (2019).
36. X. Tian, J. M. Zarate, D. Poeppel, Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex* **77**, 1–12 (2016).
37. C. R. Li, Impairment of motor imagery in putamen lesions in humans. *Neuroscience Letters* **287**, 13–16 (2000).
38. J. Niessing, *et al.*, Hemodynamic Signals Correlate Tightly with Synchronized Gamma Oscillations. *Science* **309**, 948–951 (2005).
39. S. Grimault, *et al.*, Load-dependent Brain Activity Related to Acoustic Short-term Memory for Pitch. *Annals of the New York Academy of Sciences* **1169**, 273–277 (2009).
40. K. K. Sreenivasan, M. D'Esposito, The what, where and how of delay activity. *Nat Rev Neurosci* **20**, 466–481 (2019).
41. J. Ahveninen, I. Uluç, T. Raij, A. Nummenmaa, F. Mamashli, Spectrotemporal content of human auditory working memory represented in functional connectivity patterns. *Commun Biol* **6**, 1–11 (2023).
42. F. Mamashli, *et al.*, Synchronization patterns reveal neuronal coding of working memory content. *Cell Reports* **36**, 109566 (2021).
43. A. Hillebrand, G. R. Barnes, "Beamformer Analysis of MEG Data" in *International Review of Neurobiology*, Magnetoencephalography., (Academic Press, 2005), pp. 149–171.
44. R. A. Adams, *et al.*, Impaired theta phase coupling underlies frontotemporal dysconnectivity in schizophrenia. *Brain* **143**, 1261–1277 (2020).
45. L. M. Andersen, K. Jerbi, S. S. Dalal, Can EEG and MEG detect signals from the human cerebellum?

NeuroImage **215**, 116817 (2020).
46. L. M. Andersen, S. S. Dalal, The cerebellar clock: Predicting and timing somatosensory touch. *NeuroImage* **238**, 118202 (2021).
47. L. Bonetti, *et al.*, Brain recognition of previously learned versus novel temporal sequences: a differential simultaneous processing. *Cerebral Cortex* **33**, 5524–5537 (2023).
48. L. Bonetti, *et al.*, Rapid encoding of musical tones discovered in whole-brain connectivity. *NeuroImage* **245**, 118735 (2021).
49. E. Ruzich, M. Crespo-García, S. S. Dalal, J. F. Schneiderman, Characterizing hippocampal dynamics with MEG: A systematic review and evidence-based guidelines. *Hum Brain Mapp* **40**, 1353–1375 (2019).
50. L. Bonetti, *et al.*, Revealing the spacetime hierarchical whole-brain dynamics of auditory predictive coding. 2022.11.19.517195 (2023).
51. A. Gramfort, MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7** (2013).
52. R. S. Desikan, *et al.*, An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
53. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67** (2015).
54. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
55. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Soft.* **82** (2017).