Decoding reveals the neural representation of perceived and imagined musical sounds

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

(a) Helen Wills Neuroscience Institute & Department of Psychology and Neuroscience, University of California Berkeley, Berkeley, CA
(b) Psychology Department, University of Copenhagen, Copenhagen, Denmark
(c) Center for Music in the Brain, Department of Clinical Medicine, Aarhus University and The Royal Academy of Music, Aarhus, Denmark
(d) Center for Eudaimonia and Human Flourishing, Linacre College, University of Oxford, Oxford UK
(e) Department of Psychiatry, University of Oxford, Oxford UK
(f) Institute of Computer Science, University of Bern, Bern, Switzerland
(g) 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

tel: +1-5103656529

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

Classification: Biological Sciences

Keywords: Imagery, music, decoding, MEG

Abstract

Vividly imagining a song or a melody is a skill that many people accomplish with relatively little effort. However, we are only beginning to understand how the brain represents, holds, and manipulates these musical "thoughts". Here, we decoded perceived and imagined melodies from magnetoencephalography (MEG) brain data (N = 71) to characterize their neural representation. We found that, during perception, auditory regions represent the sensory properties of individual sounds. In contrast, a widespread network including fronto-parietal cortex, hippocampus, basal nuclei, and sensorimotor regions hold the melody as an abstract unit during both perception and imagination. Furthermore, the mental manipulation of a melody systematically changes its neural representation, reflecting volitional control of auditory images. Our work sheds light on the nature and dynamics of auditory representations, informing future research 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 (1). 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 (2).

Here, we consider two kinds of auditory 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 (3–17), while manipulation further involves the frontal and parietal lobes (18,19). With the exception of the visual cortex, these brain areas are largely consistent with those engaged in visual imagery (20–22). 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) (23), 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, functional magnetic resonance imaging (fMRI) studies showed sound-specific representations in primary and secondary auditory cortex (24–27) and frontoparietal association areas (28,29) during maintenance in working memory and imagination. Other studies demonstrated imagined sounds decoding from scalp electroencephalography (EEG) (30–32). However, it remains unclear **1)** how sound sequences are represented in auditory and association areas, **2)** how these representations evolve in time, and **3)** how they change when mentally manipulated.

Here, we used MVPA of brain activity recorded with magnetoencephalography (MEG, **Fig. 1a**) to investigate how perceived, imagined and mentally manipulated short auditory sequences are represented in the brain. For each trial in the task, participants listened and then were instructed to vividly imagine 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.

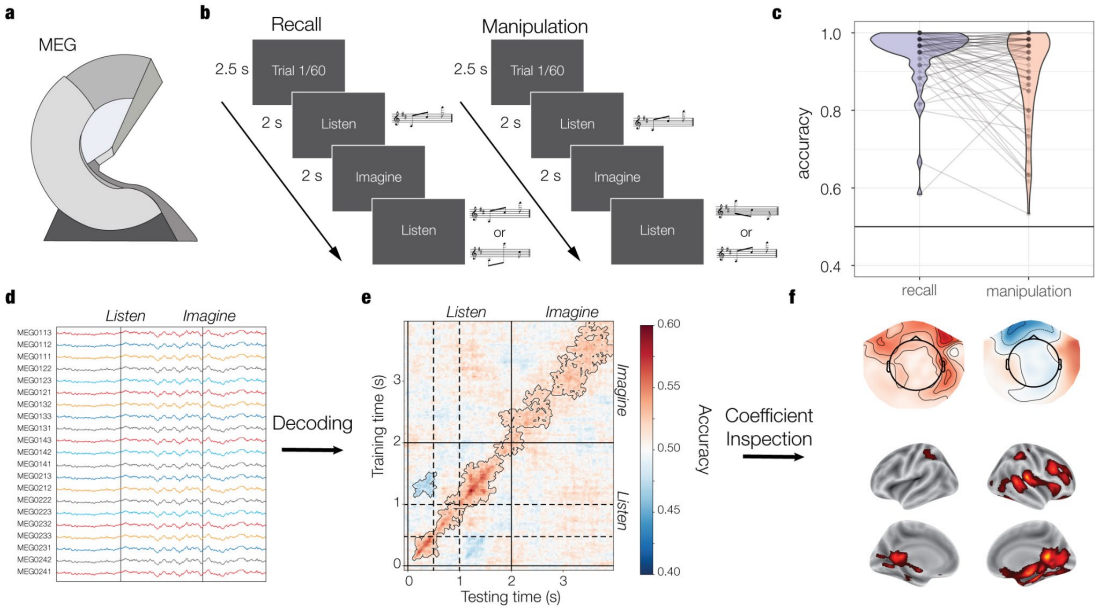

Figure 1. Materials and methods. We used MEG (a) to record the brain activity of 71 participants while they performed an imagery task (b). On each trial, participants heard and then imagined a short three-note melody. In the recall block, they imagined the melody as presented, while in the manipulation 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.

We first used a time generalized decoding technique (23) to characterize the neural dynamics of auditory representations. Then, we assessed whether mentally manipulating the melodies changed their neural representation. In the manipulation block, participants suppressed the forward pattern and mentally 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. Incorrect trials were excluded from MEG analyses. After the experiment, participants rated task-related imagery vividness on a 7-point Likert scale (from -3 to 3), with 72% of them rating 0 or above, which ranges from mild to strong vividness (Table 1). The good task performance and the vividness ratings suggest the presence of melody-specific information during imagination. Behavioral accuracy was associated with general working memory skills (33) (Wechsler Adult Intelligence Scale – WAIS; recall: r(69) = 0.3, p = .012; manipulation: r(69) = 0.29, p = .016; **Fig. S1**). No significant relationship with music training was found (34) (Goldsmiths Musical Sophistication Index – GMSI; recall: r(69) = 0.13, p = .28; manipulation: r(69) = 0.23, p = .063; **Fig. S2**; see also **Fig. S6** and **S7** for further exploratory analyses on possible associations of neural decoding with behavioral accuracy, vividness ratings, and music training).

Table 1. Vividness ratings at the end of the task across participants.

Rating	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
Number of participants	1	4	8	10	25	12	4
Percentage	1%	6%	11%	14%	35%	17%	6%

2.1. Above-chance decoding of perceived 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 S2** for full statistical report). This further confirms that mental representations were present during the imagination 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

We used between-condition testing to decode the identity of the perceived melody at all timepoints in the trial and detect manipulation-related changes in neural representations. Thus, if during manipulation participants inhibited the representation of the perceived melody and reinstated the representation of its backward version, between-condition tests should systematically predict the opposite of the perceived melody, resulting in below-chance accuracy in the imagination period. Indeed, we found below-chance accuracy 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 S1**). In both cases, accuracies were lower for between-condition than within-condition testing ($p \leq .035$; **Fig. 2c**). This indicates a flip in neural representations such that models trained in one condition consistently predicted the opposite when tested on the other condition.

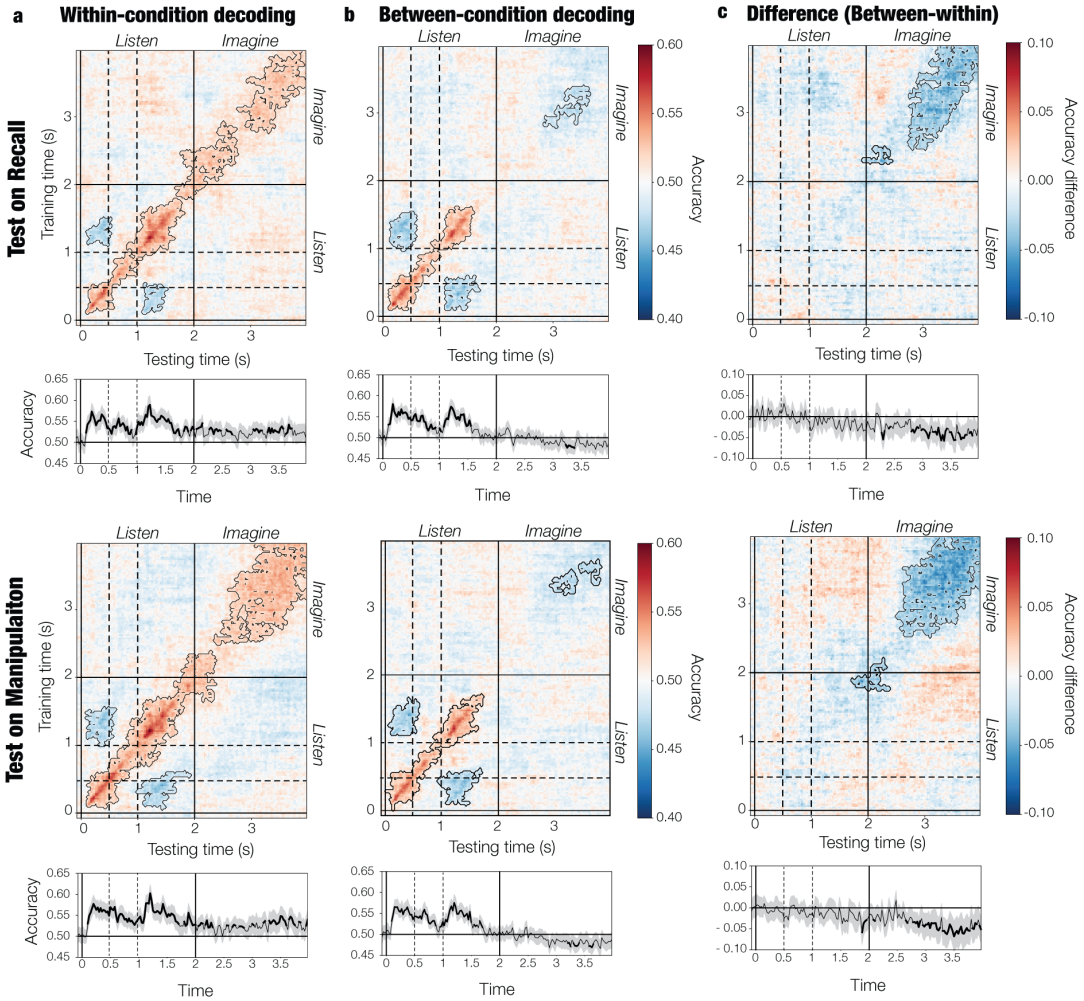

Figure 2. Accuracy for (a) within-condition decoding (train and test in the same condition), (b) between-condition decoding (train in one condition and test on the other), and the difference between the two (c). A time-generalization technique was used in which models were trained at each time point 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. 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, thus leading to lower generalization across listening and imagination. Overall, these findings indicate that, in the manipulation block, participants inhibited the perceived melody and

reinstated its backward version, resulting in a flip of neural representations. This provides evidence of ***volitional control*** over mental auditory representations.

2.2. Musical sound sequences are represented in auditory, association, sensorimotor, and subcortical areas

To elucidate the brain regions and neural features that distinguish between the melodies, we transformed the model coefficients into interpretable patterns of activation as described in (35), and localized their brain generators (**Fig. 1f**). The resulting patterns can be interpreted as the differences in neural activity that discriminate between melodies and 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 several brain areas. For the first sound (**Fig. 3a**), we found clusters of regions in both conditions ($p \leq .006$, see **Table S3** for full statistical reports) with peak activity patterns in ***auditory areas*** such as superior temporal gyrus and Heschl's gyrus, but also, in ***somatosensory*** (postcentral gyrus) and ***association areas*** (fusiform, hippocampus, retrosplenial, posterior cingulate, angular gyrus, inferior parietal cortex; see **Tables S4** and **S5** for a full report of anatomical regions). In addition, activity patterns in another cluster 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, ***motor cortex*** (precentral gyrus), and ***subcortical structures*** including the basal nuclei (putamen, caudate, accumbens, pallidum) and the thalamus. Interestingly, after the second sound information about melody identity was present in association, sensorimotor, and subcortical structures ($p < .005$), 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 across the brain.

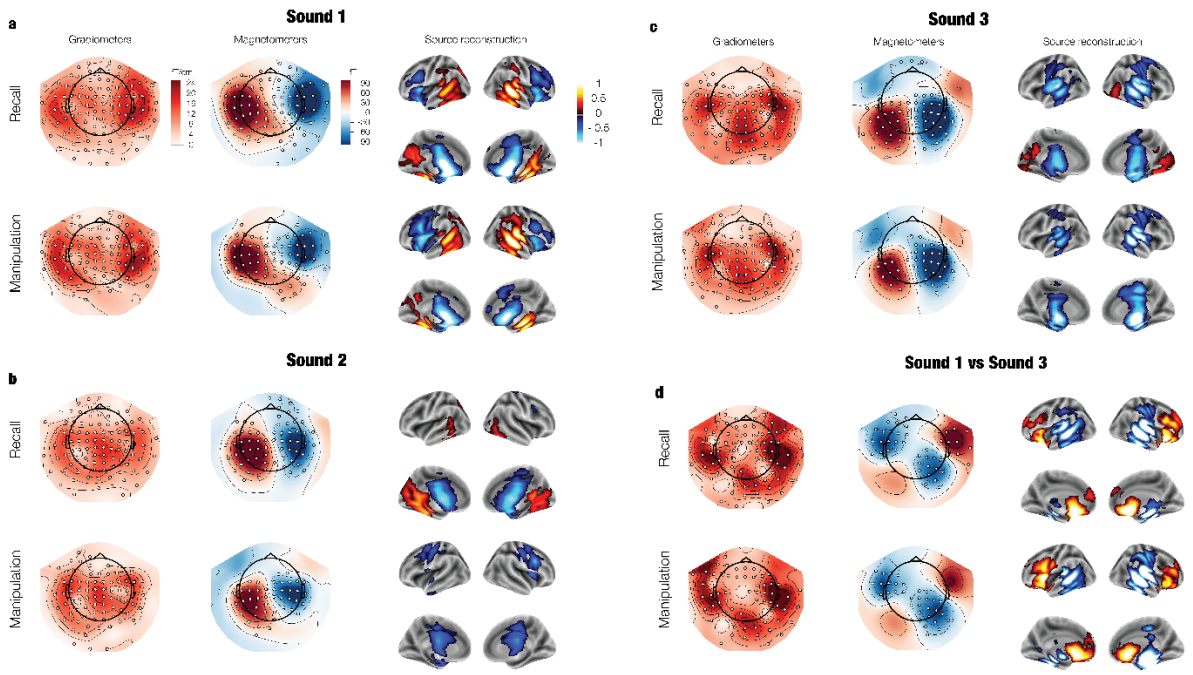

Figure 3. Patterns of neural activity that discriminate between melodies during listening, as derived from decoding coefficients and averaged over time. Patterns are shown for three time windows: (a) Sound 1 (0.2s - 0.5s), (b) Sound 2 (0.7s - 1s), and (c) Sound 3 (1.2s - 1.5s). The difference between sounds 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 areas outlined above 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 prefrontal cortex 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 (36,37). 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 association, sensorimotor, and subcortical structures.

2.2.2. Auditory representations during imagination

In the imagination period, melodies were mainly represented in non-auditory areas including basal nuclei, thalamus, mid-posterior cingulate, motor, and parietal cortex ($p < .001$, **Fig. 4a**). Additional recruitment of 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 temporal areas becoming negative, and negative fields in anterior association, sensorimotor, and subcortical areas becoming positive after 2s (**Fig. 4b; Fig. S4**). A similar switch was reported in studies that decoded imagined sounds from scalp EEG (30,31) and auditory working memory content with fMRI (25).

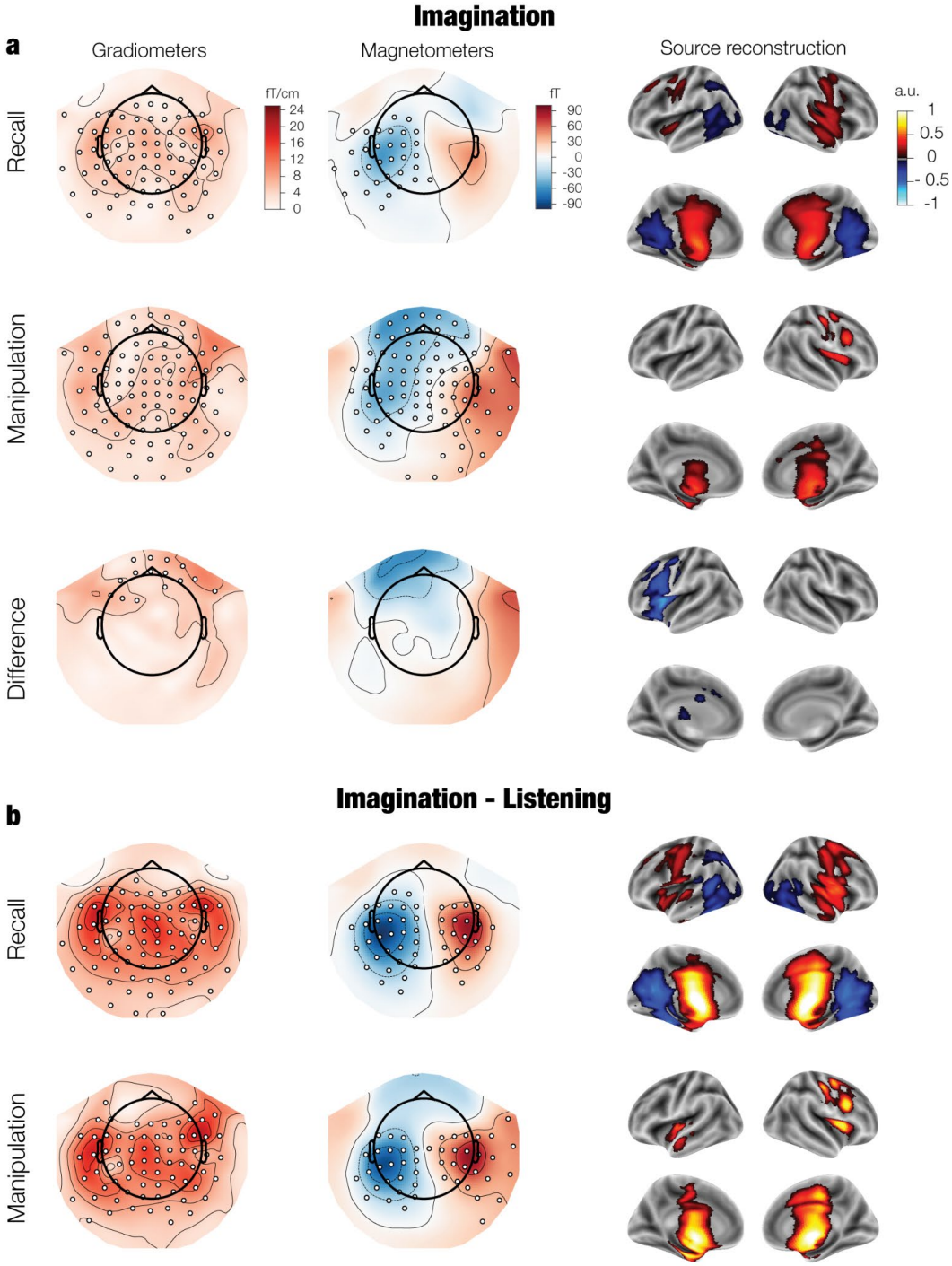

Figure 4. (a) Patterns of neural activity that discriminate between melodies during imagination in both conditions (2s - 4s), as derived from decoding coefficients and averaged over time. The difference between listening and Imagination is also presented (b). Patterns are depicted for two types of MEG 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 perceived and imagined melodies from neural activity to demonstrate that musical sound sequences are represented in auditory, association, sensorimotor, and subcortical areas, and that these representations systematically change when mentally manipulated. While previous studies have decoded imagined sound information from brain data (24,26,29–32), 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 (23). During listening, this could be due to the constantly changing sensory input. During imagination, this might reflect temporal variability between participants. The lack of generalization further suggests that representations were different between listening and imagination. This contrasts with research suggesting that perceived and imagined sounds share neural substrates and representations. For example, both imagined and actual sounds activate secondary auditory areas (13) and fMRI studies decoded imagined auditory representations from primary and secondary auditory cortex (24–27). Moreover, some studies found that representations during the omission of predictable sounds are similar to those of the actual sounds (38,39).

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 (e.g., 40). Second, experimental paradigms have either minimized the temporal variability of the imagined representations (e.g., omission studies)(38,39) or did not take time into account (fMRI studies)(24–27). 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 (25,30,31), representations in association, sensorimotor, and subcortical areas 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 (41,42). 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 different brain areas. 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, association and subcortical areas (inferior and medial temporal lobe, ventromedial prefrontal cortex, thalamus, basal nuclei) remained stable, while representations in dorsal association areas (lateral prefrontal cortex) were involved only at sequence onset. Moreover, during imagination, association and subcortical areas were the main carriers of representations, with auditory and temporal 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 temporal areas, and another one concerned with the encoding, retrieval, and manipulation of auditory sequences in association and subcortical 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 (43).

The regions that carried auditory representations in our study overlap with those identified in previous neuroimaging activation studies as important for imagery in audition and other modalities (4,6–22). 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 (44). Nevertheless, we found representations in the motor and somatosensory cortex, which is consistent with previous reports (7,45,46) 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 auditory imagery research. From these nuclei, the putamen has been related to motor imagery (47). Moreover, the basal nuclei are typically studied with the hemodynamic response in fMRI, which correlates best with high gamma (> 60 Hz) power (48) in EEG. In this study, we used instead MEG broadband signals to decode auditory objects, which might be why basal nuclei representations were found here but not in fMRI. Future research examining high gamma activity and other frequency bands will be needed to elucidate their relationship with the hemodynamic response.

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 and mentally manipulate the melodies, whereas in working memory experiments maintenance strategies usually remain unspecified. Thus, while it is possible that our participants used unconscious maintenance strategies without imagery, the explicit task instruction, the good task performance, the vividness ratings, and the between-condition decoding results suggest that they engaged in active mental recall and manipulation. Future experiments where imagery is not required are needed to further elucidate the nature of maintenance strategies and the relationship of imagery with working memory.

This caveat aside, task performance was associated with general working memory scores and the brain regions 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 (19,28,49,50). Moreover, auditory representations in working memory have been decoded from the auditory, frontal, and parietal cortices (24,28,43) and from the functional interaction of these regions (51,52). Most of these decoding studies, however, addressed working memory for individual sounds and none 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.

Two methodological caveats need to be considered. First, we localized auditory representations to both cortical and deep brain areas (basal nuclei, thalamus, hippocampus), raising concerns given the bias towards the head center of beamforming algorithms (53) 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, with implementation of appropriate controls, the use of beamformers has made the detection of deep sources increasingly common, including the basal nuclei, the medial temporal lobe, and even the cerebellum (54–57). 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.

In conclusion, our results provide evidence regarding the nature and dynamics of perceived and imagined sound representations in the brain and contribute to a growing body of work investigating musical imagery and its relationship with other modalities (2,58–63). Our findings also 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 employ different recording modalities (e.g., optical MEG, intracranial EEG), bigger datasets (e.g., by increasing the number of trials), and models that are larger and account for the temporal variability in imagination (e.g., deep learning)(64) 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 behavioral performance and three due to noisy or corrupted neural data, resulting in a final group of 71 participants (44 female, age = 28.77 +/- 8.43 SD). Three of these participants were excluded from source level analyses due to absence of anatomical images. Participants 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) (34) and of 96 (IQR = [93-105]) in the Wechsler Adult Intelligence Scale (WAIS) (33). Musical expertise was not a factor in recruitment for this experiment. 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 using piano sounds (musical pitch: A3, C#5, E6; F0: 220Hz, 554Hz, 1318Hz) synthesized with MuseScore (v3.6.2; see materials' GitHub repository for the actual sounds used). 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 individual sounds was 500ms. The sounds were normalized to peak amplitude.

3.3. Task

The experiment was implemented in Psychopy v3.1.2 (65) (see materials' GitHub repository for details). 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, encompassing the two different blocks in the experiment. 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 written 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. The task lasted approximately 20 minutes. Other experimental paradigms testing recognition memory were recorded together with this task. Results are reported elsewhere (57). The order of the paradigms was counterbalanced across participants. After the experiment, participants were asked to rate the vividness of imagery during the task on a 7-point Likert scale ranging from -3 to 3.

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) (66). 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 the joint gradiometer covariance across listening and imagination periods. For similar results obtained with the separate covariance of the listening and imagination periods see **Fig. S8.** Importantly, forward models and inverse solutions were normalized to eliminate the bias towards the center of the head inherent to beamformers (53).

3.6. Decoding analysis

We used a time-generalized decoding approach (Fig. 1d) (23) 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 non-parametric cluster-based permutations (67) 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 $p = 0.05$ based on one-sample t-tests ($p = 0.025$ 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 (**W**) into interpretable patterns of activation (**A**) for each participant using the method detailed in (35) and defined by the equation:

$$\mathbf{A} = \Sigma_x \mathbf{W} \Sigma_{\hat{y}}^{-1}$$

Where $\Sigma_{\hat{y}}$ is the covariance of model predictions and Σ_x is the covariance of neural signals.

We localized the neural generators of these patterns using the inverse solutions described in section 3.5. (**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 (68), 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 groups of regions of interest (ROI) including (1) right auditory, (2) right posteroventral association, (3) right dorsal association, (4) left dorsal association, and (5) right anteroventral/subcortical 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.

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) , Mutua Madrileña Foundation (GFR), 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 Fernández-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 materials for the experiment can be found here:

https://github.com/drqm/memory_music_iEEG.git

The code used to perform the analyses can be found here:

https://github.com/drqm/working_memory.git

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

Competing interests

The authors declare no competing interests.

References

1. Cotter KN, Christensen AP, Silvia PJ. Understanding inner music: A dimensional approach to musical imagery. Psychol Aesthet Creat Arts. 2019 Nov;13(4):489–503.
2. Küssner MB, Taruffi L, Floridou GA, editors. Music and Mental Imagery. London: Routledge; 2022. 318 p.
3. Halpern AR. Dynamic aspects of musical imagery. Ann N Y Acad Sci. 2012 Apr;1252:200–5.
4. Bastepe-Gray SE, Acer N, Gumus KZ, Gray JF, Degirmencioglu L. Not all imagery is created equal: A functional Magnetic resonance imaging study of internally driven and symbol driven musical performance imagery. J Chem Neuroanat. 2020 Mar 1;104:101748.
5. Bunzeck N, Wuestenberg T, Lutz K, Heinze HJ, Jancke L. Scanning silence: Mental imagery of complex sounds. NeuroImage. 2005 Jul 15;26(4):1119–27.
6. Foster NE, Halpern AR, Zatorre RJ. Common parietal activation in musical mental transformations across pitch and time. Neuroimage. 2013;75:27–35.
7. Halpern AR. When That Tune Runs Through Your Head: A PET Investigation of Auditory Imagery for Familiar Melodies. Cereb Cortex. 1999 Oct 1;9(7):697–704.
8. Halpern AR, Zatorre RJ, Bouffard M, Johnson JA. Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia. 2004 Jan;42(9):1281–92.
9. Herholz SC, Halpern AR, Zatorre RJ. Neuronal correlates of perception, imagery, and memory for familiar tunes. J Cogn Neurosci. 2012 Jun;24(6):1382–97.
10. Huijbers W, Pennartz CMA, Rubin DC, Daselaar SM. Imagery and retrieval of auditory and visual information: Neural correlates of successful and unsuccessful performance. Neuropsychologia. 2011 Jun 1;49(7):1730–40.
11. Pando-Naude V, Patyczek A, Bonetti L, Vuust P. An ALE meta-analytic review of top-down and bottom-up processing of music in the brain. Sci Rep. 2021 Oct 21;11(1):20813.
12. Yoo SS, Lee CU, Choi BG. Human brain mapping of auditory imagery: event-related functional MRI study. NeuroReport. 2001 Oct 8;12(14):3045–9.
13. Zatorre RJ, Halpern AR, Perry DW, Meyer E, Evans AC. Hearing in the Mind's Ear: A PET Investigation of Musical Imagery and Perception. J Cogn Neurosci. 1996 Jan 1;8(1):29–46.
14. Zvyagintsev M, Clemens B, Chechko N, Mathiak KA, Sack AT, Mathiak K. Brain networks underlying mental imagery of auditory and visual information. Eur J Neurosci. 2013;37(9):1421–34.
15. Herff SA, Herff C, Milne AJ, Johnson GD, Shih JJ, Krusienski DJ. Prefrontal High Gamma in ECoG Tags Periodicity of Musical Rhythms in Perception and Imagination. eNeuro [Internet]. 2020 Jul 1 [cited 2022 Apr 26];7(4). Available from: https://www.eneuro.org/content/7/4/ENEURO.0413-19.2020
16. Cheng THZ, Creel SC, Iversen JR. How Do You Feel the Rhythm: Dynamic Motor-Auditory Interactions Are Involved in the Imagination of Hierarchical Timing. J Neurosci Off J Soc Neurosci. 2022 Jan 19;42(3):500–12.
17. Schaefer RS. Images of time: temporal aspects of auditory and movement imagination. Front Psychol. 2014 Aug 12;5:877.
18. Zatorre RJ. Beyond auditory cortex: working with musical thoughts. Ann N Y Acad Sci. 2012;1252(1):222–8.
19. Albouy P, Weiss A, Baillet S, Zatorre RJ. Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. Neuron. 2017 Apr;94(1):193-206.e5.
20. Dijkstra N, Kok P, Fleming SM. Perceptual reality monitoring: Neural mechanisms dissociating imagination from reality. Neurosci Biobehav Rev. 2022;135:104557.
21. Dijkstra N, Bosch SE, van Gerven MAJ. Shared Neural Mechanisms of Visual Perception and Imagery. Trends Cogn Sci. 2019 May 1;23(5):423–34.

22. Pearson J. The human imagination: the cognitive neuroscience of visual mental imagery. Nat Rev Neurosci. 2019 Oct;20(10):624–34.
23. King JR, Dehaene S. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci. 2014 Apr;18(4):203–10.
24. Deutsch P, Czoschke S, Fischer C, Kaiser J, Bledowski C. Decoding of Working Memory Contents in Auditory Cortex Is Not Distractor-Resistant. J Neurosci. 2023 May 3;43(18):3284–93.
25. Linke AC, Vicente-Grabovetsky A, Cusack R. Stimulus-specific suppression preserves information in auditory short-term memory. Proc Natl Acad Sci. 2011 Aug 2;108(31):12961–6.
26. Linke AC, Cusack R. Flexible Information Coding in Human Auditory Cortex during Perception, Imagery, and STM of Complex Sounds. J Cogn Neurosci. 2015 Jul;27(7):1322–33.
27. Regev M, Halpern AR, Owen AM, Patel AD, Zatorre RJ. Mapping Specific Mental Content during Musical Imagery. Cereb Cortex N Y N 1991. 2021 Jul 5;31(8):3622–40.
28. Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD. A Brain System for Auditory Working Memory. J Neurosci. 2016 Apr 20;36(16):4492–505.
29. Uluç I, Schmidt TT, Wu Y hao, Blankenburg F. Content-specific codes of parametric auditory working memory in humans. NeuroImage. 2018 Dec;183:254–62.
30. Marion G, Di Liberto GM, Shamma SA. The Music of Silence. Part I: Responses to Musical Imagery Encode Melodic Expectations and Acoustics. J Neurosci. 2021 Aug 2;JN-RM-0183-21.
31. Di Liberto GM, Marion G, Shamma SA. The Music of Silence: Part II: Music Listening Induces Imagery Responses. J Neurosci. 2021 Sep 1;41(35):7449–60.
32. Di Liberto GM, Marion G, Shamma SA. Accurate Decoding of Imagined and Heard Melodies. Front Neurosci. 2021;15:673401.
33. Wechsler D. Wechsler Adult Intelligence Scale--Fourth Edition [Internet]. 2012 [cited 2023 Aug 5]. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/t15169-000
34. Müllensiefen D, Gingras B, Musil J, Stewart L. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. Snyder J, editor. PLoS ONE. 2014 Feb 26;9(2):e89642.
35. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014 Feb;87:96–110.
36. Vishne G, Gerber EM, Knight RT, Deouell LY. Distinct ventral stream and prefrontal cortex representational dynamics during sustained conscious visual perception. Cell Rep. 2023 Jul;42(7):112752.
37. Voytek B, Knight RT. Prefrontal cortex and basal ganglia contributions to visual working memory. Proc Natl Acad Sci. 2010 Oct 19;107(42):18167–72.
38. Chouiter L, Tzovara A, Dieguez S, Annoni JM, Magezi D, De Lucia M, et al. Experience-based Auditory Predictions Modulate Brain Activity to Silence as do Real Sounds. J Cogn Neurosci. 2015 Oct;27(10):1968–80.
39. Demarchi G, Sanchez G, Weisz N. Automatic and feature-specific prediction-related neural activity in the human auditory system. Nat Commun. 2019 Dec;10(1):3440.
40. Higgins C, Es MWJ van, Quinn A, Vidaurre D, Woolrich M. The relationship between frequency content and representational dynamics in the decoding of neurophysiological data [Internet]. bioRxiv; 2022 [cited 2022 May 9]. p. 2022.02.07.479399. Available from: https://www.biorxiv.org/content/10.1101/2022.02.07.479399v1
41. Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical Microcircuits for Predictive Coding. Neuron. 2012 Nov;76(4):695–711.
42. Bastos AM, Lundqvist M, Waite AS, Kopell N, Miller EK. Layer and rhythm specificity for predictive routing. Proc Natl Acad Sci. 2020 Dec 8;117(49):31459–69.

43. Fan Y, Han Q, Guo S, Luo H. Distinct Neural Representations of Content and Ordinal Structure in Auditory Sequence Memory. J Neurosci. 2021 Jul 21;41(29):6290–303.
44. Lima CF, Krishnan S, Scott SK. Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. Trends Neurosci. 2016 Aug;39(8):527–42.
45. Ding Y, Zhang Y, Zhou W, Ling Z, Huang J, Hong B, et al. Neural Correlates of Music Listening and Recall in the Human Brain. J Neurosci. 2019 Oct 9;39(41):8112–23.
46. Tian X, Zarate JM, Poeppel D. Mental imagery of speech implicates two mechanisms of perceptual reactivation. Cortex. 2016 Apr 1;77:1–12.
47. Li C shan R. Impairment of motor imagery in putamen lesions in humans. Neurosci Lett. 2000 Jun 16;287(1):13–6.
48. Niessing J, Ebisch B, Schmidt KE, Niessing M, Singer W, Galuske RAW. Hemodynamic Signals Correlate Tightly with Synchronized Gamma Oscillations. Science. 2005 Aug 5;309(5736):948–51.
49. Grimault S, Lefebvre C, Vachon F, Peretz I, Zatorre R, Robitaille N, et al. Load-dependent Brain Activity Related to Acoustic Short-term Memory for Pitch. Ann N Y Acad Sci. 2009;1169(1):273–7.
50. Sreenivasan KK, D'Esposito M. The what, where and how of delay activity. Nat Rev Neurosci. 2019 Aug;20(8):466–81.
51. Ahveninen J, Uluç I, Raij T, Nummenmaa A, Mamashli F. Spectrotemporal content of human auditory working memory represented in functional connectivity patterns. Commun Biol. 2023 Mar 20;6(1):1–11.
52. Mamashli F, Khan S, Hämäläinen M, Jas M, Raij T, Stufflebeam SM, et al. Synchronization patterns reveal neuronal coding of working memory content. Cell Rep. 2021 Aug 24;36(8):109566.
53. Hillebrand A, Barnes GR. Beamformer Analysis of MEG Data. In: International Review of Neurobiology [Internet]. Academic Press; 2005 [cited 2023 Jun 1]. p. 149–71. (Magnetoencephalography; vol. 68). Available from: https://www.sciencedirect.com/science/article/pii/S0074774205680063
54. Andersen LM, Jerbi K, Dalal SS. Can EEG and MEG detect signals from the human cerebellum? NeuroImage. 2020 Jul 15;215:116817.
55. Andersen LM, Dalal SS. The cerebellar clock: Predicting and timing somatosensory touch. NeuroImage. 2021 Sep 1;238:118202.
56. Bonetti L, Brattico E, Bruzzone SEP, Donati G, Deco G, Pantazis D, et al. Brain recognition of previously learned versus novel temporal sequences: a differential simultaneous processing. Cereb Cortex. 2023 May 1;33(9):5524–37.
57. Bonetti L, Brattico E, Carlomagno F, Donati G, Cabral J, Haumann NT, et al. Rapid encoding of musical tones discovered in whole-brain connectivity. NeuroImage. 2021 Dec 15;245:118735.
58. Taruffi L, Ayyildiz C, Herff SA. Thematic Contents of Mental Imagery are Shaped by Concurrent Task-Irrelevant Music. Imagin Cogn Personal. 2023 Dec;43(2):169–92.
59. Herff SA, Cecchetti G, Taruffi L, Déguernel K. Music influences vividness and content of imagined journeys in a directed visual imagery task. Sci Rep. 2021 Aug 6;11(1):15990.
60. Küssner MB, Taruffi L. Modalities and causal routes in music-induced mental imagery. Trends Cogn Sci. 2023 Feb;27(2):114–5.
61. Herff SA, McConnell S, Ji JL, Prince JB. Eye Closure Interacts with Music to Influence Vividness and Content of Directed Imagery. Music Sci. 2022 Jan 1;5:20592043221142711.
62. Margulis EH, Jakubowski K. Music, Memory, and Imagination. Curr Dir Psychol Sci. 2024 Apr 1;33(2):108–13.
63. Margulis EH, McAuley JD. Mechanisms and individual differences in music-evoked imaginings. Trends Cogn Sci. 2023 Feb;27(2):116–7.
64. Metzger SL, Littlejohn KT, Silva AB, Moses DA, Seaton MP, Wang R, et al. A high-performance neuroprosthesis for speech decoding and avatar control. Nature. 2023 Aug;620(7976):1037–46.

65. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behav Res Methods. 2019 Feb 1;51(1):195–203.
66. Gramfort A. MEG and EEG data analysis with MNE-Python. Front Neurosci [Internet]. 2013 [cited 2019 Mar 1];7. Available from: http://journal.frontiersin.org/article/10.3389/fnins.2013.00267/abstract
67. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007 Aug;164(1):177–90.
68. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 2006 Jul 1;31(3):968–80.