Hierarchical whole-brain dynamics of auditory predictive coding at the fine-grained milliseconds level

Bonetti, L.[1,2,3,4*], Fernández Rubio, G.[1], Carlomagno, F.[1,5], Pantazis, D.[6], Vuust, P.[1], Kringelbach, M. L.[1,2,3]

[1]*Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music, Aarhus, Aalborg, Denmark*

[2]*Centre for Eudaimonia and Human Flourishing, Linacre College, University of Oxford, Oxford, United Kingdom*

[3]*Department of Psychiatry, University of Oxford, Oxford, United Kingdom*

[4]*Department of Psychology, University of Bologna*

[5]*Department of Education, Psychology, Communication, University of Bari Aldo Moro, Italy*

[6]*McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), MA 02139, Cambridge, USA*

**Corresponding author: leonardo.bonetti@psych.ox.ac.uk*

Abstract

Humans extract information from the key features of the physical word, namely space and time. However, while much is known about the neural processing of visuospatial patterns, there is little information on the hierarchical brain mechanisms underlying conscious recognition of auditory sequences extended over time and their associated prediction errors. Using magnetoencephalography (MEG), we studied the brain activity of 83 participants during recognition of original and varied auditory sequences.

Results showed a network comprising auditory, ventro-medial prefrontal and inferior temporal cortices, hippocampus, and medial cingulate gyrus underlying sequence processing. The hierarchical pathway from auditory cortex to hippocampus and ventro-medial prefrontal cortex was associated with both originals and variations, but its temporal dynamics crucially differed. Recognition of original versus varied sequences elicited stronger gamma power.

Our results complement the knowledge on neural processing of visuospatial patterns by providing the hierarchical brain mechanisms underlying auditory, temporal sequence recognition.

Keywords

Recognition memory, Temporal sequences, Predictive coding (PC), Magnetoencephalography (MEG), Global neuronal workspace (GNW)

Introduction

One of the keys for understanding the brain is unravelling how its different areas are hierarchically organized in space and time. These hierarchical structural connections and their sequential activity and interaction are essential for comprehending human perception and cognition [1].

Along this line, pivotal works on the neural basis of perception highlighted the brain hierarchies in the visual system. Here, visual inputs are captured by the retina and streamed first to the lateral geniculate nuclei and then to the primary visual cortex (V1). This is essential to create a bottom-up saliency map of the visual field which is necessary to direct attention to relevant locations in the visual field [2, 3]. Afterwards, the visual information flows through a hierarchical pathway in the secondary visual areas (i.e., V2, V3, V4 and V5), which process distinct visual features. For instance, V1 and V2 selectively respond to bars of different orientations, allowing the detection of edges and corners, and to diverse colors and motion information [4]. Hierarchical processing is also needed for elementary perceptual grouping of visual items. For instance, Grossberg and colleagues [5] showed that the brain hierarchy from the lateral geniculate nuclei to the interblob cortical stream via V1 and V2 was required for creating a context-sensitive perceptual groupings of visual inputs.

Auditory perception has also a hierarchical structure from the cochlea to the primary auditory cortex. In fact, from the first neuronal processing happening in the cochlear nucleus, sound information travels along the brainstem and pons, following a hierarchical pathway comprising the trapezoid body, superior olivary complex, lateral lemniscus, and inferior colliculi in a progressive manner [6-8]. Finally, the information reaches the medial geniculate nucleus of the thalamus and is then streamed to the primary auditory cortex [9-12].

This research suggested the importance of brain hierarchy for perception, a concept that has been further expanded in the context of high-order human cognitive functions such as memory encoding and recognition. In a key review, Mesulam [1] highlighted the hierarchical organization of the brain, proposing that sensorial information becomes cognitively meaningful through an extensive associative elaboration process relying on a hierarchy of core neural layers (levels) and brain regions. These include the primary sensory, upstream unimodal, downstream unimodal, heteromodal, paralimbic, and limbic zones of the cerebral cortex. In his view, the connections from one layer to another are reciprocal and allow higher layers to provide top-down feedbacks to the earlier levels of processing. Mesulam [1] argued that the first neural levels are responsible for unimodal associations and encode basic features

of sensory information such as motion, colour, form, and pitch. Then, more complex concepts (e.g., faces, objects, word-forms, spatial locations) originate from the integration of the information coming from the lower levels with the highest brain layers (or transmodal areas), which are built upon heteromodal, paralimbic and limbic cortices. For instance, transmodal areas in the medial lobe such as the hippocampal-entorhinal complex and Wernicke's area are crucial for transforming perception of sensory stimuli into conscious, cognitive recognition [13].

Further research has proposed new theories that are in line with Mesulam's [1] thesis. For instance, the concept of neural hierarchical organization is at the basis of the two-stream hypothesis of the brain [14, 15]. This perspective proposes two main pathways for high-order elaboration of visual and auditory information. On the one hand, the ventral stream shows a hierarchical pathway from sensorial areas (e.g., visual and auditory cortices) to the medial temporal lobe and is responsible for processing features associated to object recognition [15, 16]. On the other hand, the dorsal stream shows a different hierarchical pathway, leading from sensory cortices to the parietal lobe, and is necessary for elaborating spatial information [17].

In addition to perception and cognition, brain hierarchies have also been studied with resting state, showing the coupling between functional (temporal) and structural (spatial) hierarchies [18-20]. For instance, Mahjoory and colleagues [18] investigated the association between cortical thickness and peak frequency of neural oscillations using magnetoencephalography (MEG). They discovered that the dominant peak frequency in a brain area decreased along the posterior-anterior axis and followed the brain global hierarchy from early sensory to higher order areas. Moreover, this spatial gradient of peak frequency was anticorrelated with the one of cortical thickness. In another study, Deco and Kringelbach [19] proposed a new framework for capturing the key principles of interaction between local and global neural activity and information flow in the brain. They showed that the propagation of neural activity between different regions could be described as intrinsic perturbations which followed hierarchical pathways. Similarly, Deco and colleagues [20], using both resting state and task data in over 1000 participants, showed evidence that the complex functional architecture of the human brain is built upon a hierarchical organization.

Taken together, this research has clearly shown that successful brain functioning requires a perfectly balanced hierarchical interplay between brain areas and networks. This is especially true when dealing with complex stimuli such as patterns where the meaning emerges from the harmonious combination of different items posed in relation to each other [21].

In this context, once again neuroscience has primarily focused on the visual system, discovering some of the brain hierarchies that allow humans to perceive and recognize complex images, visual scenes, and spatial patterns. For instance, several studies investigated the neural signature of face recognition [22-24] and detection of spatial patterns and configurations [25-27]. These works converged in showing the robust involvement of the fusiform gyrus in face encoding and recognition [28, 29]. As a more specific example, Dobs and colleagues [23] recently unfolded some of the hierarchical brain dynamics underlying the recognition of independent features of faces. Using representational similarity analysis, the authors showed that facial age and gender information were decoded before personal identity. With regards to other categories of visual patterns, previous studies revealed the crucial role of hippocampal regions and several areas of visual cortex such as calcarine and occipital gyri for object recognition and processing of spatial patterns provided through visual channels [26, 30, 31].

Previous studies have also focused on induced responses, revealing oscillatory bursts of gamma band associated to recognition of familiar objects. For instance, in an electroencephalography (EEG) study, Gruber and colleagues [32] showed augmented induced gamma responses following the presentation of meaningful, familiar versus meaningless, unfamiliar stimuli. Moreover, they revealed a dense pattern of phase-locking values between distant electrodes during recognition of familiar images. Similarly, in another EEG study, Slobounov and colleagues [33] reported increased gamma burst during the recognition of target stimuli.

Although vision is the primary sense of humans, to pursue a successful and flourishing existence auditory information is essential. Moreover, while the visual system relies on the recognition of patterns arranged in space, the auditory system extracts information from patterns and sequences extended over time [34]. Thus, investigating complex cognitive processes in both visual and auditory systems is critical to understand how the brain comes to term with the two key physical dimensions of our world: space and time.

Previous research on the neural basis of the auditory system has provided a wide array of findings, shedding light, for example, on the brain mechanisms underlying the automatic processing of simple sound sequences and on the brain correlates of auditory memory tasks. For instance, previous investigations focusing on sound processing revealed the primary role of the auditory cortex [34, 35]. These studies uncovered the early, well-known components of the event-related potentials/fields (ERP/F) occurring in response to sounds, such as the N100, mismatch negativity, and P3a [35, 36]. Research on the brain correlates of auditory processing

has also involved memory tasks. For instance, Gaab and colleagues [37] observed the brain activity of participants who were requested to compare different simple melodies. When they successfully performed the task, their brain showed activity mainly in the superior temporal, superior parietal, posterior dorsolateral frontal, and dorsolateral cerebellar regions, supramarginal and left inferior frontal gyri. A more recent study by Kumar and colleagues [38] showed the key role of the activity and connectivity between primary auditory cortex, inferior frontal gyrus and hippocampus for performing an auditory working memory (WM) task consisting of maintaining a series of single sounds. Further, in a recent MEG study, Albouy and colleagues [39] explored the brain activity underlying memory retention, showing that theta oscillations in the dorsal stream of the participants' brain anticipated their abilities to perform an auditory WM task which consisted of manipulating and maintaining sound information. Finally, as reported for the visual domain, EEG research has also examined induced responses during recognition in the auditory system. For instance, previous studies reported increased gamma power and phase locking values between auditory areas when participants recognized familiar stimuli such as environmental sounds [40] and words [41].

Altogether, research on auditory neuroscience has covered several topics, but little is known on the hierarchical brain mechanisms underlying conscious encoding and recognition of auditory sequences. Recently, Bonetti et al. [42-44] revealed that encoding and recognition of sound information recruited a large network of brain areas, spanning from auditory cortex to medial cingulate, inferior temporal cortex and hippocampus. While these studies provided a first glimpse of the neural basis of auditory encoding and recognition, the brain hierarchies underlying conscious recognition of temporal sequences and their associated prediction error remain obscure. Thus, in this study we revealed the temporal dynamics of these hierarchies, providing new evidence in terms of both evoked and induced responses associated to complex auditory memory processes.

Results

Experimental design and behavioral results

Eighty-three participants performed an old/new auditory recognition task [42-46] during magnetoencephalography (MEG) recordings. After learning a short musical piece, participants were presented with 135 five-tone musical excerpts lasting 1750 ms each and were requested to state whether each excerpt belonged to the original music ('memorized' sequence (M), old) or it was a varied musical sequence ('novel' sequence (N), new) (**Figure 1a**).

Twenty-seven excerpts were taken from the original musical piece and 108 were variations of the original melodies. We organized these variations in four categories and created them by changing every musical tone of the sequence after the first (NT1), second (NT2), third (NT3) or fourth (NT4) tone (**Figure 1b**). Thus, all the original sequences and variations had the first same tone. **Figure S1** shows a depiction in musical notation of all the sequences used in the study.

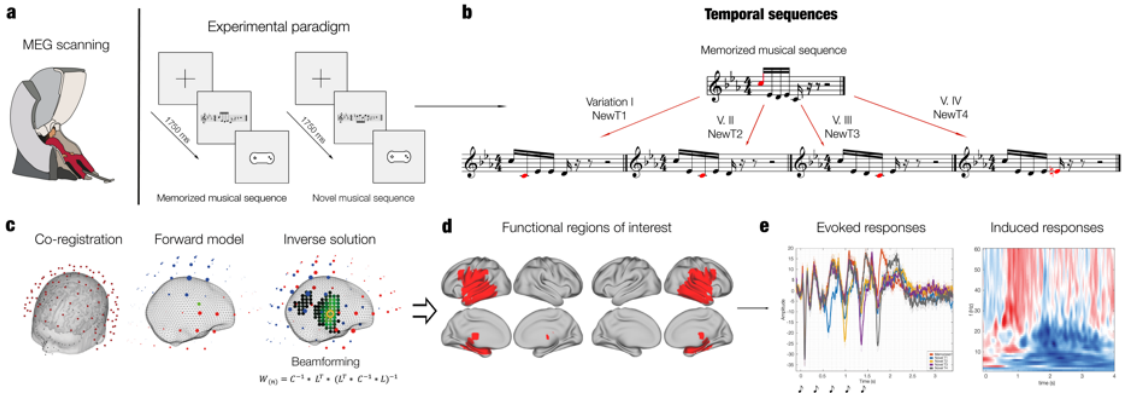


Figure 1. Experimental design, stimuli, and analysis pipeline

a *- Participants' brain activity was collected using magnetoencephalography (MEG) while they performed an old/new auditory recognition task. One at a time, five-tone temporal sequences (i.e., melodies) were presented. These could either belong to the musical piece participants had previously listened to (memorized musical sequences, 'old') or could be novel musical sequences ('new'). In this figure, below the title 'Experimental paradigm', we depicted an example of a memorized musical sequence ('old', left, 2nd square) with the relative response pad that participants used to state whether they recognized the excerpt as 'old' or 'new' (left, 3th square). Then, we depicted an example of a novel musical sequence ('new', right, 2nd square). The total number of trials was 135 (27 memorized and 108 novel musical sequences) and their order was randomized.* ***b*** *– Five types of temporal sequences (M, NT1, NT2, NT3, NT4) were used in the study. There were three M sequences*

that comprised the first five tones of the first three measures of the memorized musical piece. The figure shows one of them, as an example (top row). The N sequences were created through systematic variations of each of the three M sequences. This procedure consisted of changing every musical tone of the sequence after the first (NT1), second (NT2), third (NT3) or fourth (NT4) tone, as illustrated in the example reported in the bottom row. ***C*** *– The MEG data was pre-processed, co-registered with the individual anatomical MRI data, and source reconstructed using a beamforming algorithm. This procedure returned one time series for each of the 3559 reconstructed brain sources.* ***d*** *– Six main functional brain parcels were derived from the activity of the previously described 3559 brain sources (the figure illustrates only two ROIs as an example, see* ***Figure S2*** *for all parcels).* ***e*** *- We studied the evoked (left plot) and induced (right plot) responses for each ROI and experimental condition.*

Before focusing on the recorded brain data, we performed statistical analyses on the MEG task behavioral data (see **Table 1** for descriptive statistics). We computed two independent Kruskal-Wallis H test (non-parametric one-way analysis of variance) to assess whether the five categories of temporal sequences (M, NT1, NT2, NT3, and NT4) differed in terms of response accuracy and reaction times (**Figure 2a**).

The Kruskal-Wallis H test for response accuracy was significant ($H(4) = 36.38$, $p < .001$), indicating a difference between categories in the number of correct responses. The Tukey-Kramer correction for multiple comparisons highlighted that NT4 trials were correctly identified with a lower frequency than M ($p = .001$), NT1 ($p = .001$), NT2 ($p = .0003$), NT3 ($p < .0001$).

The Kruskal-Wallis H test for the reaction times was also significant ($H(4) = 22.53$, $p = .0002$).The Tukey-Kramer correction for multiple comparisons highlighted that NT4 trials were correctly identified with a greater reaction time than M ($p = .0016$), NT1 ($p = .0013$), NT2 ($p = .0054$), NT3 ($p = .0008$).

Behavioral variables	***M***	***NT1***	***NT2***	***NT3***	***NT4***
Correct recognition	22.33 ± 5.30	22.36 ± 4.27	21.58 ± 5.31	21.66 ± 5.34	17.04 ± 7.12
Reaction times (ms)	2426 ± 226	2407 ± 284	2431 ± 282	2415 ± 272	2578 ± 259

Table 1. MEG task behavioral results

Mean and standard deviations across participants of number of correctly recognized trials and reaction times (ms) for the five experimental conditions (previously memorized [‘M’], novel [‘NT1’], novel [‘NT2’], novel [‘NT3’], novel [‘NT4’]).

Multivariate pattern analysis

Using a support vector machine (SVM) classifier (see details in Methods), we performed multivariate pattern analyses to decode different neural activity associated with the recognition of M versus N. Specifically, we computed four independent analyses, decoding M versus each of the four categories of novel sequences (i.e., M versus NT1, M versus NT2, M versus NT3, and M versus NT4).

As shown in **Figure 2b** and **Figure S3**, each of these analyses returned a decoding time series showing how the neural activity differentiated the pair of experimental conditions. The decoding time series were significantly different from chance level in several time windows. Decoding M versus NT1 returned the following main significant time windows: 0.53-0.73 sec; 0.91 – 0.95 sec; 1.27 – 1.30 sec; 1.62 – 1.88 sec ($q < .012$, false-discovery rate [FDR]-corrected). Decoding M versus NT2 gave rise to the following main significant time windows: 0.89 – 1.18 sec; 1.26 – 1.42 sec; 1.54 – 1.89 sec ($q < .012$, FDR-corrected). Decoding M versus NT3 returned one main significant time window: 1.25-2.07 sec ($q < .012$, FDR-corrected). Finally, decoding M versus NT4 showed the following main significant time window: 1.64-2.07 ($q < .012$, FDR-corrected). Detailed statistical results are reported in **Table S1** and illustrated in **Figure S3**.

To evaluate the persistence of discriminable neural information over time, we used a temporal generalization approach by training the SVM classifier at a given timepoint *t* and testing it across all timepoints. This was calculated for the four pairs of experimental conditions described above. The signed-rank test against chance level and cluster-based Monte-Carlo simulations [42-44, 46, 47] (MCS; $\alpha = .01$, MCS *p-value* = .001) showed that the performance of the classifier was significantly above chance even a few hundreds of milliseconds beyond the diagonal, for all pairs of conditions. Notably, the neural difference between M and N was comparable across diverse musical tones, as shown by the recurrent patterns depicted in **Figure 2c**. Detailed statistical results are reported in **Table S2**.

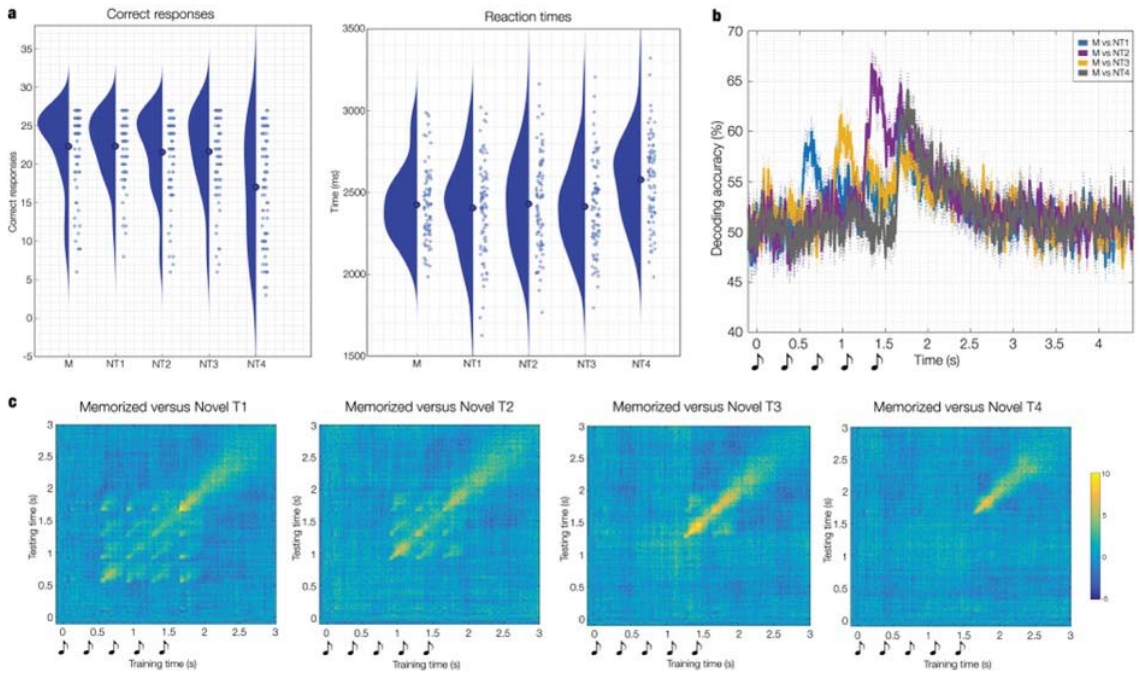


Figure 2. Behavioural results, and decoding of neural data

a *– Scatter and violin plots for the correct responses (left) and reaction times (right) for the experimental task performed by the participants in the MEG. The plots separately illustrate each experimental condition. Each dot represents a participant.* ***b*** *- Multivariate pattern analysis decoding the different neural activity associated with memorized versus novel musical sequences. The plot shows the decoding time series for four rounds of pairwise decoding (M vs NT1, M vs NT2, M vs NT3, M vs NT4). The sketch of the musical tones represents the onset of the sounds forming the temporal sequences.* ***c*** *- Temporal generalization decoding accuracy showing prediction of training time on testing time for each time-point. The sketch of the musical tones represents the onset of the sounds forming the temporal sequences.*

Neural sources of the differential brain activity between M and N

We employed a local-spheres forward model and a beamforming approach as inverse solution in an 8-mm grid (**Figure 1c**). To detect the brain sources underlying the differential signal observed for M and N, we considered both the results from the decoding analyses and the inspection of the MEG data after preprocessing. This was necessary since decoding can capture differences between conditions, but it does not discriminate which condition is associated with a larger neural signal (see Methods for details).

This procedure returned the following significant time windows: 0.50 – 0.60 sec; 0.70 – 0.80 sec; 0.98 – 1.02 sec; 1.05 – 1.15 sec; 1.33 – 1.39 sec; 1.45 – 1.55 sec; 1.70 – 1.75 sec; 1.75 – 1.85 sec. For each time window and condition, we averaged the time series of all brain sources over time and computed t-tests contrasting each combination of M versus Ns. Finally,

we corrected for multiple comparisons using a 3D cluster-based MCS (MCS, α = .003, MCS *p-value* = .001).

As reported in **Table 2** and illustrated in **Figure 3**, this analysis returned several significant clusters of differential brain activity between M and N, primarily located in the bilateral medial cingulate gyrus (MC), left (HITL) and right hippocampal area and inferior temporal cortex (HITR), left (ACL) and right auditory cortex (ACR), and bilateral ventro-medial prefrontal cortex (VMPFC).

Detailed statistical results are extensively reported in **Table S3**.

Time-windows	**Contrast**	**Cluster size**	**Main ROI(s)**	**Peak *t-value***	***MCS p-value***
0.50 - 0.60	M vs NT1	10	MC	-4.18	< .001
0.70 - 0.80	M vs NT1	132	MC	5.59	< .001
1.05 - 1.15	M vs NT1	162	MC	5.38	< .001
1.45 – 1.55	M vs NT1	270	ACR MC	4.91	< .001
1.75 – 1.85	M vs NT1	175	HITR	5.94	< .001
0.98 – 1.02	M vs NT2	94	VMPFC	-4.33	< .001
1.05 – 1.15	M vs NT2	57	MC	4.47	< .001
1.45 – 1.55	M vs NT2	9	MC	3.7	< .001
1.75 – 1.85	M vs NT2	28	HITR	4.16	< .001
1.33 – 1.39	M vs NT3	300	HITR VMPFC	-5.34	< .001
1.45 – 1.55	M vs NT3	58	ACL	-4.57	< .001
1.75 – 1.85	M vs NT3	79	HITR	4.42	< .001
1.70 – 1.75	M vs NT4	584	VMPFC	-5.57	< .001
1.75 – 1.85	M vs NT4	197	ACL	-5.39	< .001

Table 2. Brain sources of decoding significant time windows

Significant clusters of brain sources of the significant time windows returned by the decoding analysis. The table shows the time windows, contrasts, cluster size (number of voxels forming the cluster), main ROI(s) involved in the cluster, peak t-value, and MCS p-value.

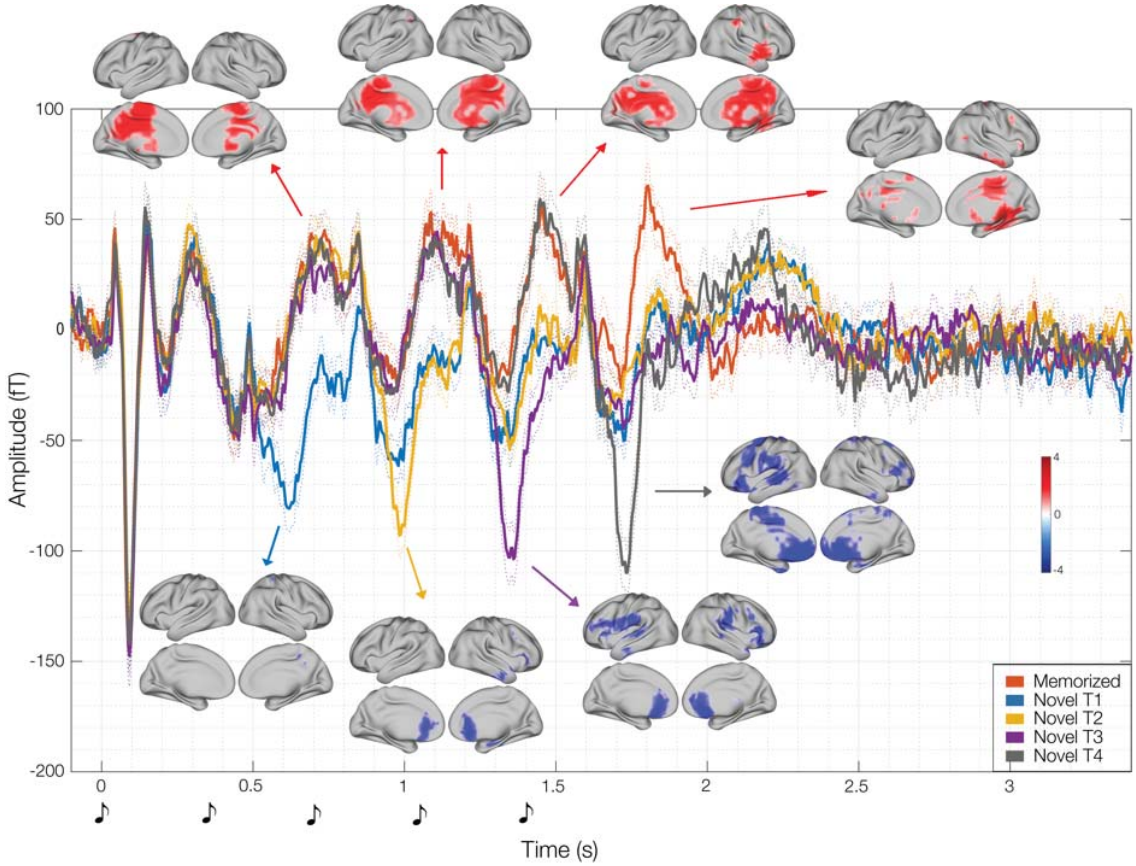


Figure 3. Brain activity recorded at a prototypical magnetometer channel (MEG 0211) and source reconstruction of the main components

Brain activity recorded over time by the fronto-temporal left MEG channel 0211 showing the five experimental conditions. The sketch of the musical tones represents the onset of the sounds forming the musical sequences. For each of the main positive components, contrasts between the source reconstruction of M versus NT1 have been computed and corrected for multiple comparisons using cluster-based MCS. Results are reported in the brain template above the waveforms. With regards to the negative component indexing the prediction error associated to the disruption of the original sequences, we computed contrasts between the source reconstruction of M versus each category of N (i.e., M vs NT1, M vs NT2, M vs NT3, M vs NT4, respectively) and corrected for multiple comparisons using cluster-based MCS. Results are reported in the brain template below the waveforms. The colour of the arrows illustrates what contrast was performed (e.g., the blue arrow indicates that we contrasted M versus NT1, while the yellow arrow refers to M versus NT2, etc.). The colorbar shows the t-values obtained from the contrasts.

Functional regions of interests (ROIs)

Based on the previous analyses, we wished to better define the spatial extent of the significant brain areas based on their functional profile and computed their associated time series (see Methods for details).

This procedure returned six broad ROIs which were particularly active during the recognition task: left *(i)* and right auditory cortex *(ii)*, left *(iii)* and right hippocampus-inferior temporal cortex *(iv)*, ventro-medial prefrontal cortex *(v)*, and medial cingulate gyrus *(vi)*. Then, we computed one t-test for each time-point and each combination of M versus Ns. Finally, we corrected for multiple comparisons using one-dimensional (1D) cluster-based MCS (MCS, $\alpha = .05$, MCS *p-value* = .001).

This analysis returned several significant clusters of differential brain activity over time between M and Ns. As shown in **Figure 4**, M versus N was characterized by stronger activity in VMPFC, ACL, and HITR after 350 – 450 ms from the onset of each tone. Similarly, M presented stronger negative activity ($p < .001$) than N in the MC after 400 – 500 ms from the onset of each tone.

Contrast	**ROI**	**Temporal extent of the largest clusters from the 1st tone of the sequence**	**Peak *t*-value**	***P-value***
Positive activity				
M vs NT1	VMPFC	900 – 1190	6.13	< .001
	ACL	900 – 1040	6.26	< .001
	HITR	1660 – 1930	5.45	< .001
M vs NT2	VMPFC	1680 – 1930	5.94	< .001
	ACL	1260 – 1390	3.56	< .001
	HITR	1640 – 1930	6.12	< .001
M vs NT3	VMPFC	1310 – 1540	7.04	< .001
	ACL	1250 – 1390	4.80	< .001
	HITR	1310 – 1530	6.69	< .001
M vs NT4	VMPFC	1680 – 1910	8.92	< .001
	ACL	1640 – 1790	6.93	< .001
	HITR	1680 – 1900	7.45	< .001
Negative activity				
M vs NT1	MC	680 – 860	-5.18	< .001
M vs NT2	MC	980 – 1180	-5.77	< .001
M vs NT3	MC	1360 – 1550	-5.41	< .001
M vs NT4	MC	1730 – 1920	-5.40	< .001

Table 3. Largest clusters of stronger activity of M versus Ns

Largest clusters of significantly stronger activity of M versus Ns computed for the six ROIs considered in the study. The table shows the contrast, the correspondent ROI, the temporal extent (in ms) of the largest cluster, the peak t-value of the cluster and the associated MCS p-value. The MC shows stronger negativity since the polarity of the MC signal was negative. All clusters are reported in detail in ***Table S4****.*

Conversely, late N100 responses localized in AC were stronger for N versus M. For instance, the temporal extent of the larger cluster of ACL for M versus NT1 was 520 – 700 ms from the onset of the first tone of the sequence (peak *t-value* = 6.75, $p < .001$). Moreover, HIT and VMPFC showed a stronger response for N versus M occurring about 250 – 300 ms after altering the original sequences. For instance, the temporal extent of the larger cluster of HITR for M versus NT4 was1680 – 1900 ms from the onset of the first tone of the sequence (peak *t-value* = 7.45, $p < .001$); the temporal extent of the larger cluster of VMPFC for M versus NT4 was 1680 – 1910 ms from the onset of the first tone of the sequence (peak *t-value* = 8.92, $p < .001$). **Table 3** reports the main significant clusters, while complete statistical results are reported in **Table S4**.

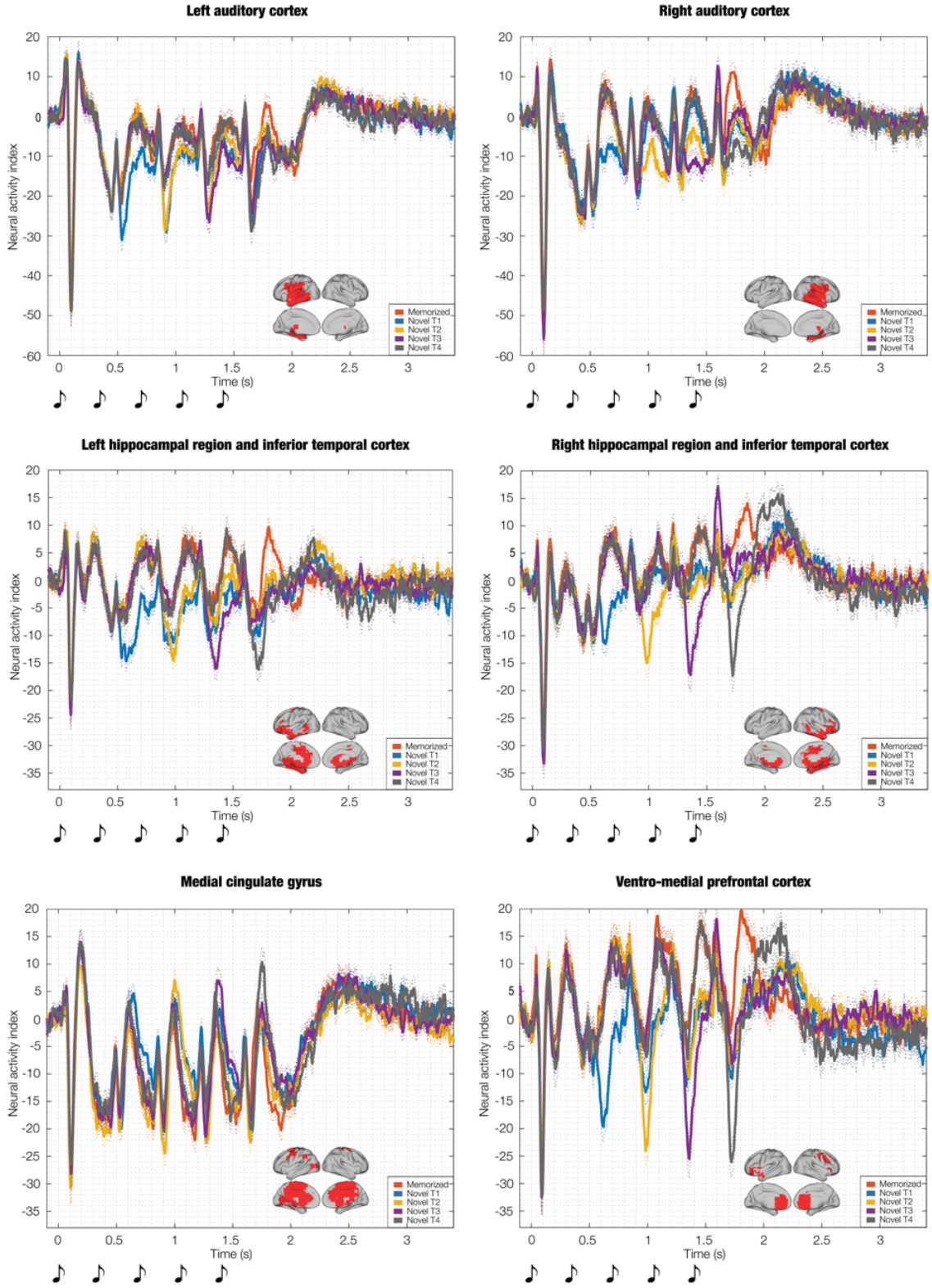


Figure 4. Brain activity (evoked responses) depicted for each ROI and experimental condition

Source localized brain activity illustrated for each experimental condition (M, NT1, NT2, NT3, NT4) and ROI.

The sketch of the musical tones represents the onset of the sounds forming the temporal sequences. The brain templates illustrate the spatial extent of the ROIs, which are also identified by the title above each plot.

Of particular interest is the brain response occurring when the original musical sequence was varied. Here, AC presented a rapid, sharp signal (150 ms after the altered tone), while HIT and VMPFC responded slightly later (around 200-250 ms from the altered tone). As an example, **Figure 5** depicts the cross-correlation between the time series of ACL and VMPFC averaged over participants, highlighting the 80 ms lag between the two signals.

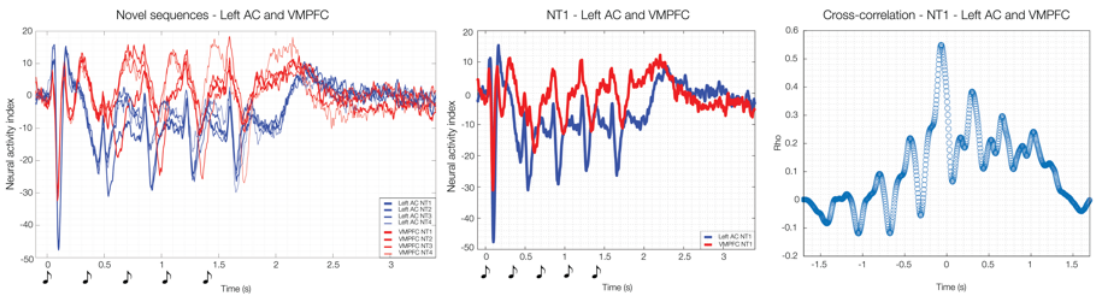


Figure 5. Comparison and cross-correlation between left AC and VMPFC

The left plot shows the source localized brain activity illustrated for each category of N (i.e., NT1, NT2, NT3, NT4) and both left auditory cortex (AC) and ventro-medial prefrontal cortex (VMPFC). Of particular interest it is the sharp peak occurring after the onset of each tone where left AC precedes VMPFC. Moreover, while the strength of the signal increases over time for VMPFC, this does not happen for left AC. The middle plot shows only condition NT1 for both left AC and VMPFC. Finally, the right plot illustrates a cross-correlation computed between the time series averaged over participants of left AC and VMPFC. The plot shows that the strongest cross-correlation happened with a time-lag of approximately 80 ms. The sketch of the musical tones below the first two plots represents the onset of the sounds forming the temporal sequences.

Time-frequency analysis for induced responses

We computed time-frequency analysis using complex Morlet wavelet transform [48] (from 1 to 60 Hz with 1-Hz intervals). This analysis was conducted for induced responses. First, we estimated independent time-frequency decompositions for each voxel of the six ROIs and for each trial. Then, the computed power spectra were averaged over voxels within each ROI and over trials. Finally, in line with the previous analyses, we calculated four contrasts (M versus NT1, M versus NT2, M versus NT3, and M versus NT4). Specifically, we computed a t-test for each frequency and timepoint and corrected for multiple comparisons using 2D cluster-

based MCS. As shown in **Figure 6**, results were similar across ROIs and displayed a generalized increased gamma power for M versus N ($p < .001$). Conversely, N versus M presented a stronger power between 2 and 20 Hz (corresponding to, approximately, theta, alpha, and beta bands), in the time window 1.0 – 3.0 seconds ($p < .001$). Detailed statistical results about this procedure are extensively reported in **Table S5**.

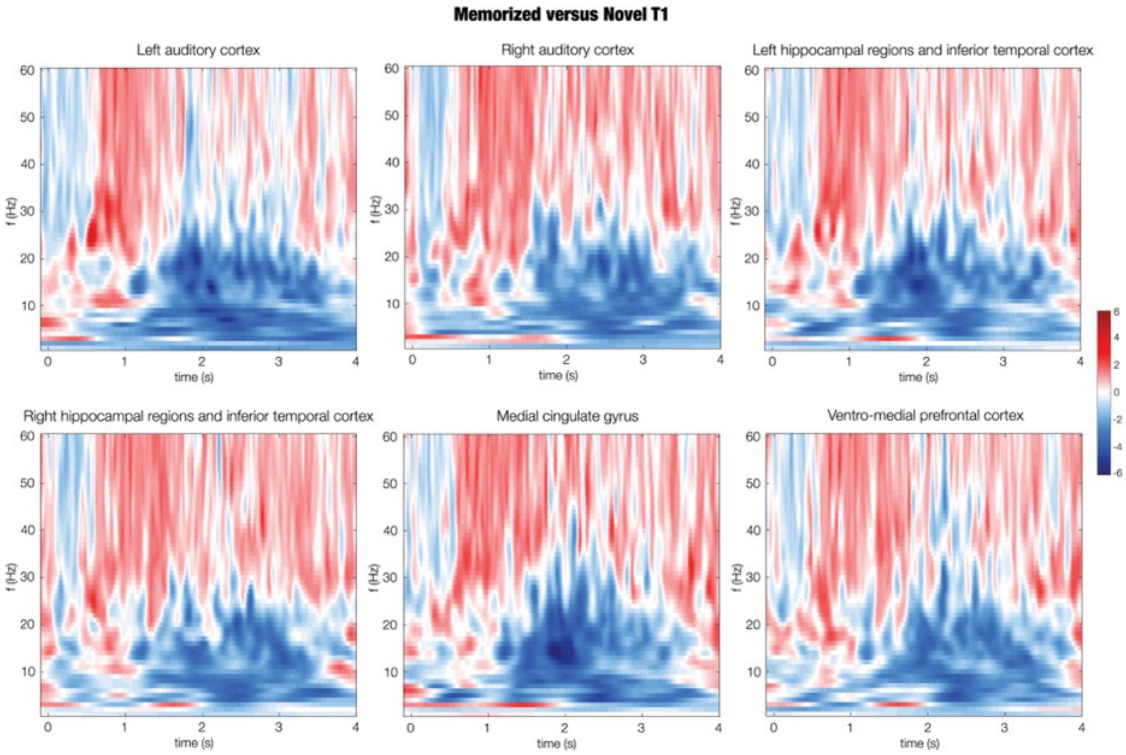


Figure 6. Source localized induced responses – M versus NT1

Contrasts between the source localized induced responses of M versus NT1. The plots indicate a stronger power for gamma in M. Moreover, theta was overall stronger for M versus NT1 during the presentation of the sounds, while alpha, beta and theta were stronger for NT1 versus M after the end of the temporal sequences. The colorbar indicates the t-values obtained by contrasting M versus NT1.

Discussion

In this study, we investigated the hierarchical dynamics of the brain underlying the recognition of previously memorized auditory sequences and the conscious prediction error originated when varying these sequences.

We demonstrated that the neural activity underlying processing of the temporal sequences could discriminate M versus N. This started when the original sequences were disrupted, and it extended until the end of the sequences. Evoked-response analyses returned a large network of approximately the same brain areas involved in both recognition of M and detection of their variations. Remarkably, the functional dynamics of those brain areas was different for M and N.

On the one hand, we observed a hierarchical pathway which characterized the recognition of M. The auditory cortex presented a positive component peaking after 300 ms of each tone forming the sequence. These peaks were followed by a similar component originated in ventro-medial prefrontal cortex, hippocampus and inferior temporal cortex which achieved its maximum 400 ms after each tone. In addition, the medial cingulate presented a negative peak 400 – 450 ms after each tone. All these components were either absent or strongly reduced in the varied sequences. To be noted, while the response of the medial cingulate was the same to all notes, the activity of ventro-medial prefrontal cortex, hippocampus and inferior temporal cortex increased over time.

On the other hand, our results showed a different hierarchical pathway for conscious prediction error. In this case, we observed a fast response of auditory cortex peaking approximately 100 – 150 ms after the onset of each tone. These peaks corresponded to the well-known N100, were stronger for N versus M and peaked for each tone of the sequence. On the contrary, a later response in the ventro-medial prefrontal cortex and hippocampus occurred only after the tone that disrupted the original sequences. Here, a first sharp negative peak appeared 250 – 300 ms after the tone altering the sequence, while a second positive component peaked 500 – 550 ms after the tone. The strength of the responses was progressively greater depending on the tone used to introduce the variation (i.e., a variation in tone five [NT4] elicited stronger responses than a variation in tone two [NT1]).

Induced-response analysis showed that activity in the gamma band was stronger for M versus N, especially during the presentation of the sequence. Conversely, after the end of the sequence, activity in alpha and beta bands was stronger for N versus M. These results were quite similar across the brain network highlighted by the evoked-response analyses.

Finally, accuracy and reaction times in the behavioral task were nearly equal for M and NT1, NT2, NT3. Notably, when varying only the last tone of the sequence (NT4), the reaction times were suddenly larger and the accuracy lower.

Our results are coherent with the ones reported by Bonetti et al. [42-46], which showed a large network of brain areas involved in the recognition of previously memorized auditory sequences. Specifically, they highlighted the crucial role of the auditory cortex, insula, hippocampus, cingulate gyrus, basal ganglia, frontal operculum, and orbitofrontal cortices. In these studies, the authors suggested that the cingulate preceded the hippocampus, insula, and inferior temporal cortices during the recognition process. However, they did not reveal the actual hierarchical brain processing occurring while the sequence was unfolding.

Conversely, the current study showed that the auditory cortex presented a positive component peaking after 300 ms of each tone forming the sequence. Notably, this peak was absent after the sequence was varied, suggesting its relevance for the recognition process. After the auditory cortex, a similar component originated in the ventro-medial prefrontal cortex, hippocampus, and inferior temporal cortex, achieving its maximum 400 ms after each tone. This phenomenon was observed after each tone forming the previously memorized sequence and increased in strength over time. Further, it completely disappeared after the sequence was altered. This finding suggests that the recognition of the auditory sequence was built over time through a rapid hierarchical pathway going from the auditory cortex to the ventro-medial prefrontal cortex, hippocampus, and inferior temporal cortex. This result is coherent with a number of studies suggesting a hierarchical organization of the brain [1, 5, 19]. It also expands their significance by providing novel information on the spatiotemporal dynamics of the hierarchical organization of the auditory system in connection to crucial memory processes. Moreover, our findings support previous studies which showed the auditory cortex as the primary neural core for processing auditory information [49, 50]. Interestingly, although the stimuli used in this study were musical, the recognition of the temporal sequences required the involvement of the inferior temporal cortex and hippocampus, which were previously associated mainly to linguistic elaborations [51, 52] and abstract memory [53, 54]. Similarly, we found that the recognition of the auditory sequences required the involvement of the ventro-medial prefrontal cortex, whose role in auditory processing has not been clearly established yet. Although it is not possible to make definite claims at this stage, we argue that the ventro-medial prefrontal cortex, together with the hippocampus, may track the progress of the sequence to make the evaluation about the

sounds. In this view, it may be the primary responsible of extracting the meaning of the sequence.

Another relevant brain region playing a prominent role was the medial cingulate gyrus, which presented a negative peak 400 – 450 ms after each tone. Its activity was overall stronger for M versus N, but it did not disappear after the sequence was altered. Moreover, its strength did not increase over time. This suggests that the medial cingulate may not be involved in the actual recognition of the sequences. Instead, in accordance with previous works [55], it may be required for allocating attentional resources. Alternatively, in a bolder perspective, it could be argued that the medial cingulate might be specifically involved in auditory processing, joining the musical tones, and providing the perception of the musical sequences as a whole pattern and not a sequential group of single tones. This interpretation would be coherent with the specific involvement of medial cingulate in a variety of musical tasks, as highlighted by a recent meta-analysis [56], but it calls for future studies specifically designed to test this hypothesis.

A relevant difference between Bonetti et al. [43, 44] and Fernández Rubio et al. [45, 46] and our findings is related to the frequency of the evoked responses. While these studies showed that the recognition of the temporal sequences was associated with activity in the cingulate, hippocampus, insula, and inferior temporal cortex in a very slow frequency (0.1 – 1 Hz) accompanying the whole duration of the sequences, here we observed a distinct, faster response to each tone of the sequence. The frequency of such responses was of approximately 1 – 1.5 Hz. This new finding shows that the recognition of the sequence is associated with responses to each of its constituent items (the musical tones). This also suggests that the slow signal observed by Bonetti et al. [43, 44] and Fernández Rubio et al. [45, 46] was the summation of the responses to each sound of the sequences. Since the stimuli used in those studies was faster than ours (the duration of each tone was 250 ms versus 350 ms), the single responses were closer in time and added up building a slower and stronger signal (0.1 – 1 Hz). Future studies should modulate the speed of the musical sequences to detect whether 1 – 1.5 Hz is a specific rhythm of the brain associated to recognition or whether it is driven by the speed of the stimuli.

Our results on temporal sequences recognition are coherent with the global neuronal workspace (GNW) hypothesis proposed by Changeux [57]. He hypothesized that processing privileged categories of stimuli such as meaningful temporal sequences activate the brain areas comprised in the GNW. The GNW was defined as a network of brain areas responsible for consciously processing information in terms of attention, memory, and valence, and

subsequently for making it available to the whole brain [58, 59]. Following Changeux's hypothesis, the recognition of memorized versus novel musical sequences led to stronger activity in putative regions of the GNW such as the cingulate gyrus, hippocampus, and ventro-medial prefrontal cortex. These areas could be necessary to extract a meaningful representation of the sequences and match them with previously acquired memory traces. Importantly, this study did not only support the hypothesis proposed by Changeux, but largely expanded it by revealing the hierarchical dynamics between these brain regions while participants recognized memorized and novel temporal sequences.

Our findings also widely support the well-known predictive coding theory (PCT) [60, 61], which states that the brain is constantly updating internal models to predict external information and stimuli. Recently, this theory has been connected to complex cognitive processes, finding a remarkable example in the neuroscience of music [60, 62]. In their work, Vuust and colleages [62] and Koelsch and colleagues [60] suggested that, while processing music, the brain constantly generates hypothesis about the upcoming development of musical sentences. Our results support and expand PCT for recognition of both previously memorized and novel sequences. On the one hand, when the upcoming sound was matched with the predicted sound based on the previously stored memory trace, first auditory cortex and then hippocampus, inferior temporal cortex and ventro-medial prefrontal cortex responded. These responses increased over time, showing stronger neural activity after each successful prediction of the upcoming sounds. On the other hand, we showed conscious prediction errors when the sequences were altered. This indicates that when the upcoming sound was incoherent with the prediction made by the brain, a network of hierarchical areas was recruited, with the information flowing from auditory cortex to ventro-medial prefrontal cortex and hippocampal regions. Notably, this brain network was similar to the one employed for the recognition of previously memorized sequences, but their temporal dynamics sharply differed.

This latter finding is also coherent with the plethora of studies investigating automatic prediction error in audition indexed by N100 and mismatch negativity (MMN) [63-65]. Previous works revealed the primary involvement of auditory cortex in the generation of the prediction error signal [63-65], reporting a complementary yet much reduced recruitment of the medial cingulate, frontal and hippocampal areas [66-68]. Conversely, in our study we investigated the conscious prediction error and revealed the hierarchical organization of the brain which recruited first auditory cortex (100 – 150 ms) and then, with a stronger activity, the ventro-medial prefrontal cortex and hippocampus (250 – 500 ms). Moreover, our results showed that

auditory cortex discriminated M versus N but did not distinguish the strength of the errors (i.e., errors happening later in the sequence). For instance, the response to the variation inserted at tone number two was the same to the variation at tone number five. In addition, the auditory cortex responded to novel tones for the entire duration of the sequence in a similar manner (i.e., for NT1, the auditory cortex responded in the same way to tones number two, three, four, and five). Remarkably, the prediction error observed in the hippocampal regions and ventro-medial prefrontal cortex showed a different strength depending on *when* the sequence was altered (e.g., the response to the variation inserted at tone number five was much stronger than to the variation at tone number two). Moreover, these areas responded primarily to the first tone where we introduced the variation and very little to the subsequent sounds. These findings suggest that the brain signature underlying the awareness of the variation may be represented by the responses recorded in the ventro-medial prefrontal cortex and hippocampus and their specific temporal dynamics.

Along this line, our findings showed a potential relationship between reaction times and accuracy in the recognition task and the second response of the prediction error occurring in the right hippocampus and inferior temporal cortex and in the ventro-medial prefrontal cortex.

Indeed, both reaction times and accuracy were approximately the same for M, NT1, NT2, and NT3. However, accuracy was significantly reduced, and reaction times increased for NT4. Similarly, while the second component of the prediction errors (occurring after approximately 500 ms from the onset of the varied tone) was rather sharp for NT1, NT2, and NT3, its frequency was much slower for NT4. There are at least two possible explanations for this phenomenon. *(i)* The variation of the last tone of the sequence elicited a slower prediction error, both at a neural and behavioral level. *(ii)* Alternatively, a bolder and more intriguing hypothesis relates to musical chunking and the beat used to present the stimuli. We argue that the lower accuracy and higher reaction times for NT4 was due to the chunking occurring when listening to the musical stimuli presented with a beat every four tones. In this view, after listening to four tones of the original sequence (corresponding to a full beat), the perception that the sequence belonged to the M category was very strong and, especially much stronger than after only three tones. For this reason, we did not observe a linear increase in reaction times and accuracy but only a strong difference between all categories and NT4 and a much slower prediction error only for NT4 compared to the other categories of N. Currently, we do not have enough data to make definitive claims and future studies are called to vary the length of the sequences (e.g., having sequences with a beat every four tones

and a length of seven, eight or nine tones). Similarly, future studies should use musical stimuli with a beat every three tones and thus reveal which interpretation is correct.

Finally, the induced-response analysis showed that the gamma band was stronger for M versus N, especially during the presentation of the sequences. Conversely, after the end of the sequence, alpha and beta bands were stronger for N versus M. This result is coherent with previous studies which reported increased gamma power and phase-locking between different brain areas during recognition of target stimuli [32, 33, 40, 41]. In addition, our findings expand on previous literature by providing evidence that bursts of gamma activity are associated with recognition of temporal sequences built upon musical sounds. The induced-response analysis also showed stronger power for alpha and beta in N versus M after the end of the sequence. Arguably, this result may represent the higher processing required by the brain after listening to novel temporal sequences, possibly to store the new information carried by the unfamiliar sounds. Future studies are necessary to further clarify this interpretation. Moreover, further research employing MEG and additional tools such as intracranial EEG (iEEG) should conduct cross-frequency coupling analysis, testing whether gamma-theta coupling is connected to recognition of temporal sequences.

To conclude, this study unravels the hierarchical dynamics of the brain underlying conscious processing of auditory sequences extended over time, providing novel insights on the neural basis of memory recognition as well as the brain mechanisms underlying conscious prediction error.

Data availability

The codes are available at the following link: https://github.com/leonardob92/LBPD-1.0.git. The multimodal neuroimaging data related to the experiment is available upon reasonable request.

Acknowledgements

The Center for Music in the Brain (MIB) is funded by the Danish National Research Foundation (project number DNRF117). Additionally, we thank the Fundación Mutua Madrileña for the economic support provided to author GFR.

LB is supported by Carlsberg Foundation (CF20-0239), Lundbeck Foundation (Talent Prize 2022), Center for Music in the Brain, Linacre College of the University of Oxford, and Society for Education and Music Psychology (SEMPRE's 50th Anniversary Awards Scheme).

MLK is supported by Center for Music in the Brain and Centre for Eudaimonia and Human Flourishing, which is funded by the Pettit and Carlsberg Foundations.

Author contributions

LB, GFR, and MLK conceived the hypotheses. LB designed the study. LB, MLK, and PV recruited the resources for the experiment. LB, GFR, and FC collected the data. LB, GFR, and FC performed pre-processing and statistical analysis. DP provided MATLAB codes for decoding analysis. DP, MLK, and PV provided essential help to interpret and frame the results within the neuroscientific literature. GFR and LB wrote the first draft of the manuscript. LB and FC prepared the figures. All the authors contributed to and approved the final version of the manuscript.

Competing interests' statement

The authors declare no competing interests.

References

1. Mesulam, M.M. From sensation to cognition. *Brain* **121 (Pt 6)**, 1013-52; 10.1093/brain/121.6.1013 (1998).
2. Li, Z. A saliency map in primary visual cortex. *Trends in Cognitive Sciences* **6**, 9-16; 10.1016/s1364-6613(00)01817-9 (2002).
3. Zhaoping, L. A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology* **58**, 1-10; 10.1016/j.conb.2019.06.001 (2019).
4. Jessell, T.M., Kandel, E.R., & Schwartz, J.H. 27. Central visual pathways. *Principles of neural science*, 533-540; (2000).
5. Grossberg, S., Mingolla, E., & Ross, W.D. Visual brain and visual perception: how does the cortex do perceptual grouping? *Trends in Neurosciences* **20**, 106-111; (1997).
6. Moore, J.K. Organization of the human superior olivary complex. *Microsc Res Tech* **51**, 403-12; 10.1002/1097-0029(20001115)51:4<403::AID-JEMT8>3.0.CO;2-Q (2000).
7. Demanez, J.P., & Demanez, L. Anatomophysiology of the central auditory nervous system: basic concepts. *Acta Oto-Rhino-Laryngologica Belgica* **57**, 227-36; (2003).
8. Mendoza, J.E., *Trapezoid Body*, in *Encyclopedia of Clinical Neuropsychology*, J.S. Kreutzer, J. DeLuca, and B. Caplan, Editors. 2011, Springer New York, NY. p. 2549.
9. Pandya, D.N. Anatomy of the auditory cortex. *Revue Neurologique* **151**, 486-94; (1995).
10. Kaas, J.H., Hackett, T.A., & Tramo, M.J. Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology* **9**, 164-70; 10.1016/s0959-4388(99)80022-1 (1999).
11. Kaas, J.H., & Hackett, T.A. Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11793-9; 10.1073/pnas.97.22.11793 (2000).
12. Scott, S.K., & Johnsrude, I.S. The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences* **26**, 100-7; 10.1016/S0166-2236(02)00037-1 (2003).
13. Addis, D.R., Barense, M., & Duarte, A., *The Wiley handbook on the cognitive neuroscience of memory*. 2015: John Wiley & Sons.

14. Whitwell, R.L., Milner, A.D., & Goodale, M.A. The Two Visual Systems Hypothesis: New Challenges and Insights from Visual form Agnosic Patient DF. *Frontiers in Neurology* **5**, 255; 10.3389/fneur.2014.00255 (2014).
15. Goodale, M.A., & Milner, A.D. Separate visual pathways for perception and action. *Trends in Neurosciences* **15**, 20-5; 10.1016/0166-2236(92)90344-8 (1992).
16. Weiller, C., Reisert, M., Peto, I., Hennig, J., Makris, N., Petrides, M., Rijntjes, M., & Egger, K. The ventral pathway of the human brain: A continuous association tract system. *NeuroImage* **234**, 117977; 10.1016/j.neuroimage.2021.117977 (2021).
17. Arbib, M.A. Dorsal and ventral streams in the evolution of the language-ready brain: Linking language to the world. *Journal of Neurolinguistics* **43**, 228-253; (2017).
18. Mahjoory, K., Schoffelen, J.-M., Keitel, A., & Gross, J. The frequency gradient of human resting-state brain oscillations follows cortical hierarchies. *Elife* **9**, e53715; (2020).
19. Deco, G., & Kringelbach, M.L. Hierarchy of information processing in the brain: a novel 'intrinsic ignition'framework. *Neuron* **94**, 961-968; (2017).
20. Deco, G., Vidaurre, D., & Kringelbach, M.L. Revisiting the global workspace orchestrating the hierarchical organization of the human brain. *Nature human behaviour* **5**, 497-511; (2021).
21. Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J.R., van der Helm, P.A., & van Leeuwen, C. A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin* **138**, 1218-52; 10.1037/a0029334 (2012).
22. Allison, T., Ginter, H., McCarthy, G., Nobre, A.C., Puce, A., Luby, M., & Spencer, D.D. Face recognition in human extrastriate cortex. *Journal of Neurophysiology* **71**, 821-5; 10.1152/jn.1994.71.2.821 (1994).
23. Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. How face perception unfolds over time. *Nature communications* **10**, 1-10; (2019).
24. Tsao, D.Y., & Livingstone, M.S. Mechanisms of face perception. *Annual Review of Neuroscience* **31**, 411-37; 10.1146/annurev.neuro.30.051606.094238 (2008).
25. Boccara, C.N., Sargolini, F., Thoresen, V.H., Solstad, T., Witter, M.P., Moser, E.I., & Moser, M.B. Grid cells in pre- and parasubiculum. *Nature Neuroscience* **13**, 987-94; 10.1038/nn.2602 (2010).

26. Hartley, T., Lever, C., Burgess, N., & O'Keefe, J. Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20120510; (2014).
27. Theves, S., Fernandez, G., & Doeller, C.F. The Hippocampus Encodes Distances in Multidimensional Feature Space. *Current Biology* **29**, 1226-1231 e3; 10.1016/j.cub.2019.02.035 (2019).
28. Furl, N., Garrido, L., Dolan, R.J., Driver, J., & Duchaine, B. Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of Cognitive Neuroscience* **23**, 1723-40; 10.1162/jocn.2010.21545 (2011).
29. George, N., Dolan, R.J., Fink, G.R., Baylis, G.C., Russell, C., & Driver, J. Contrast polarity and face recognition in the human fusiform gyrus. *Nature Neuroscience* **2**, 574-80; 10.1038/9230 (1999).
30. Bosking, W.H., Crowley, J.C., & Fitzpatrick, D. Spatial coding of position and orientation in primary visual cortex. *Nature Neuroscience* **5**, 874-82; 10.1038/nn908 (2002).
31. Cichy, R.M., Pantazis, D., & Oliva, A. Resolving human object recognition in space and time. *Nature Neuroscience* **17**, 455-62; 10.1038/nn.3635 (2014).
32. Gruber, T., Trujillo-Barreto, N.J., Giabbiconi, C.-M., Valdés-Sosa, P.A., & Müller, M.M. Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. *NeuroImage* **29**, 888-900; (2006).
33. Slobounov, S., Tutwiler, R., Slobounova, E., Rearick, M., & Ray, W. Human oscillatory brain activity within gamma band (30–50 Hz) induced by visual recognition of non-stable postures. *Cognitive Brain Research* **9**, 177-192; (2000).
34. Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. Relating structure to function: Heschl's gyrus and acoustic processing. *Journal of Neuroscience* **29**, 61-9; 10.1523/JNEUROSCI.3489-08.2009 (2009).
35. Naatanen, R., & Picton, T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* **24**, 375-425; 10.1111/j.1469-8986.1987.tb00311.x (1987).
36. Näätänen, R., & Kreegipuu, K., *The Mismatch Negativity (MMN)*, in *The Oxford handbook of event-related potential components*, S.J. Luck and E.S. Kappenman, Editors. 2012, Oxford university press. p. 145-158.

37. Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. *NeuroImage* **19**, 1417-26; 10.1016/s1053-8119(03)00224-6 (2003).
38. Kumar, S., Joseph, S., Gander, P.E., Barascud, N., Halpern, A.R., & Griffiths, T.D. A Brain System for Auditory Working Memory. *J Neurosci* **36**, 4492-505; 10.1523/JNEUROSCI.4341-14.2016 (2016).
39. Albouy, P., Weiss, A., Baillet, S., & Zatorre, R.J. Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. *Neuron* **94**, 193-206 e5; 10.1016/j.neuron.2017.03.015 (2017).
40. Lenz, D., Schadow, J., Thaerig, S., Busch, N.A., & Herrmann, C.S. What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *International Journal of Psychophysiology* **64**, 31-38; (2007).
41. Han, J., Mody, M., & Ahlfors, S.P. Gamma phase locking modulated by phonological contrast during auditory comprehension in reading disability. *Neuroreport* **23**, 851-6; 10.1097/WNR.0b013e32835818e1 (2012).
42. Bonetti, L., Brattico, E., Carlomagno, F., Donati, G., Cabral, J., Haumann, N.T., Deco, G., Vuust, P., & Kringelbach, M.L. Rapid encoding of musical tones discovered in whole-brain connectivity. *NeuroImage* **245**, 118735; 10.1016/j.neuroimage.2021.118735 (2021).
43. Bonetti, L., Brattico, E., Carlomagno, F., Cabral, J., Stevner, A., Deco, G., Whybrow, P.C., Pearce, M., Pantazis, D., & Vuust, P. Spatiotemporal brain dynamics during recognition of the music of Johann Sebastian Bach. *bioRxiv*, (2020).
44. Bonetti, L., Brattico, E., Bruzzone, S.E.P., Donati, G., Deco, G., Pantazis, D., Vuust, P., & Kringelbach, M.L. Temporal pattern recognition in the human brain: a dual simultaneous processing. *bioRxiv*, (2021).
45. Fernández Rubio, G., Brattico, E., Kotz, S.A., Kringelbach, M.L., Vuust, P., & Bonetti, L. The spatiotemporal dynamics of recognition memory for complex versus simple auditory sequences. *bioRxiv*, 2022.05.15.492038; 10.1101/2022.05.15.492038 (2022).
46. Fernández Rubio, G., Carlomagno, F., Vuust, P., Kringelbach, M.L., & Bonetti, L. Associations between abstract working memory abilities and brain activity underlying long-term recognition of auditory sequences. *bioRxiv*, (2022).
47. Bruzzone, S.E.P., Lumaca, M., Brattico, E., Vuust, P., Kringelbach, M.L., & Bonetti, L. Dissociated brain functional connectivity of fast versus slow frequencies underlying

individual differences in fluid intelligence: a DTI and MEG study. *Scientific Reports* **12**, 4746; 10.1038/s41598-022-08521-5 (2022).

48. Daubechies, I., *Ten lectures on wavelets*. 1992: SIAM.
49. Brattico, E., & Pearce, M. The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts* **7**, 48; (2013).
50. Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., & Ball, T. A rapid sound-action association effect in human insular cortex. *PloS One* **2**, e259; 10.1371/journal.pone.0000259 (2007).
51. Zatorre, R.J., Belin, P., & Penhune, V.B. Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences* **6**, 37-46; 10.1016/s1364-6613(00)01816-7 (2002).
52. Limongi, R., Sutherland, S.C., Zhu, J., Young, M.E., & Habib, R. Temporal prediction errors modulate cingulate-insular coupling. *NeuroImage* **71**, 147-57; 10.1016/j.neuroimage.2012.12.078 (2013).
53. Hasselmo, M.E., *A Handbook for Modeling Hippocampal Circuits*. 2011, Frontiers Research Foundation.
54. Squire, L.R., Dronkers, N., & Baldo, J., *Encyclopedia of neuroscience*. Vol. 2. 2009: Elsevier Amsterdam, The Netherlands:.
55. Merkley, T.L., Larson, M.J., Bigler, E.D., Good, D.A., & Perlstein, W.M. Structural and functional changes of the cingulate gyrus following traumatic brain injury: relation to attention and executive skills. *Journal of the International Neuropsychological Society* **19**, 899-910; 10.1017/S135561771300074X (2013).
56. 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. *Scientific Reports* **11**, 1-15; (2021).
57. Changeux, J.P., *Artistic Creativity: A Neuronal Hypothesis*, in *Secrets of creativity: What neuroscience, the arts, and our minds reveal*, S. Nalbantian and P.M. Matthews, Editors. 2019, Oxford University Press. p. 21–P1.N50.
58. Dehaene, S., & Changeux, J.P. Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biology* **3**, e141; 10.1371/journal.pbio.0030141 (2005).
59. Dehaene, S., & Changeux, J.P. Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200-27; 10.1016/j.neuron.2011.03.018 (2011).

60. Koelsch, S., Vuust, P., & Friston, K. Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences* **23**, 63-77; 10.1016/j.tics.2018.10.006 (2019).
61. Friston, K. Predictive coding, precision and synchrony. *Cognitive Neuroscience* **3**, 238-9; 10.1080/17588928.2012.691277 (2012).
62. Vuust, P., Heggli, O.A., Friston, K.J., & Kringelbach, M.L. Music in the brain. *Nature Reviews: Neuroscience*, 10.1038/s41583-022-00578-5 (2022).
63. Brattico, E., Näätänen, R., & Tervaniemi, M. Context effects on pitch perception in musicians and nonmusicians: Evidence from event-related-potential recordings. *Music Perception* **19**, 199-222; (2001).
64. Naatanen, R., Paavilainen, P., Rinne, T., & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology* **118**, 2544-90; 10.1016/j.clinph.2007.04.026 (2007).
65. Bonetti, L., Bruzzone, S.E.P., Sedghi, N.A., Haumann, N.T., Paunio, T., Kantojarvi, K., Kliuchko, M., Vuust, P., & Brattico, E. Brain predictive coding processes are associated to COMT gene Val158Met polymorphism. *NeuroImage* **233**, 117954; 10.1016/j.neuroimage.2021.117954 (2021).
66. Bonetti, L., Carlomagno, F., Kliuchko, M., Gold, B.P., Palva, S., Haumann, N.T., Tervaniemi, M., Huotilainen, M., Vuust, P., & Brattico, E. Whole-brain computation of cognitive versus acoustic errors in music. *bioRxiv*, (2022).
67. Kastner, S., & Ungerleider, L.G. Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience* **23**, 315-41; 10.1146/annurev.neuro.23.1.315 (2000).
68. Korzyukov, O.A., Winkler, I., Gumenyuk, V.I., & Alho, K. Processing abstract auditory features in the human auditory cortex. *NeuroImage* **20**, 2245-58; 10.1016/j.neuroimage.2003.08.014 (2003).

Materials and methods

Participants

The participant sample consisted of 83 volunteers (33 males and 50 females) aged 19 to 63 years old (mean age: 28.76 ± 8.06 years). The sample was recruited in Denmark and participants came from Western countries. All participants were healthy and reported normal hearing. Their educational background was homogeneous.

The project was approved by the Institutional Review Board (IRB) of Aarhus University (case number: DNC-IRB-2020-006). The experimental procedures were carried out in compliance with the Declaration of Helsinki – Ethical Principles for Medical Research. All participants gave the informed consent before starting the experimental procedure.

Experimental stimuli and design

In this study, we used an old/new paradigm auditory recognition task [1-5] during magnetoencephalography (MEG) recordings. First, participants listened to a short (approximately 25 seconds long) musical piece twice and were asked to memorize it as much as possible. The musical piece consisted of the first four measures of the right-hand part of Johann Sebastian Bach's Prelude No. 2 in C Minor, BWV 847. A MIDI version of the piece was created using Finale (MakeMusic, Boulder, CO). Each tone of the piece had the same duration (approximately 350 ms). Second, participants were presented with 135 five-tone musical excerpts that lasted 1750 ms each. Participants were requested to state whether each excerpt belonged to the original music ('memorized' sequence [M], old) or were a varied musical sequence ('novel' sequence [N], new) (**Figure 1a**). Twenty-seven excerpts were drawn from the original musical piece and 108 were variations of the original melodies (**Figure S1** shows all the sequences used in the study). The two categories of stimuli (M and N) were created as follows. The M sequences were comprised by the first five tones of the first three measures of the musical piece. These sequences were presented nine times each, for a total of 27 trials. The N sequences were created through systematic variations of the three M sequences (**Figure 1b**). This procedure consisted of changing every musical tone of the sequence after the first (NT1), second (NT2), third (NT3) or fourth (NT4) tone. We created nine variations for each of the original M sequences and each of the four categories of N. This resulted in 27 N sequences for each category, and 108 N in total. To be noted, the variations were created according to the following rules:

- Inverted melodic contours (used twice): the melodic contour of the variation was inverted with respect to the original M sequence (i.e., if the M sequence had the following melodic contour: down-down-up-down, the N sequence would be: up-up-down-up)
- Same tone scrambled (used three times): the remaining tones of the M sequence were scrambled (e.g., M sequence: C-E-D-E-C, was converted into NT1 sequence: C-C-E-E-D). When this was not possible (e.g., in the case of NT4, where only the last tone is different from the M sequence), we substituted the last tone of the M sequence with a random tone.
- Same tone (used three times): the same tone was repeatedly used, in some cases varying only the octave (e.g., M sequence: C-E-D-E-C, was transformed into NT1 sequence: $C\text{-}E^{8-}\ E^{8-}\ E_{8}^{-}\ E_{8}$).
- Scrambling intervals (used once): the intervals between the tones were scrambled (e.g., M sequence: 6^{th}m - 2^{nd}m – 2^{nd}m – 3^{rd}m, was adapted to NT1 sequence: 2^{nd}m, 6^{th}m, 3^{rd}m, 2^{nd}m).

This procedure allowed us to investigate *(i)* the brain dynamics underlying the recognition of previously memorized auditory sequences and (*ii*) the conscious detection of the sequence variation.

Data acquisition

The MEG recordings were acquired in a magnetically shielded room at Aarhus University Hospital (AUH), Aarhus, Denmark, using an Elekta Neuromag TRIUX MEG scanner with 306 channels (Elekta Neuromag, Helsinki, Finland). The data was recorded at a sampling rate of 1000 Hz with an analogue filtering of 0.1 – 330 Hz. Before the recordings, the head shape of the participants and the position of four Head Position Indicator (HPI) coils were registered with respect to three anatomical landmarks using a 3D digitizer (Polhemus Fastrak, Colchester, VT, USA). This recording was later used to co-register the MEG data with the MRI anatomical scans. For the entire duration of the MEG recordings, the HPI coils registered the continuous localization of the head, which was subsequently employed for movement correction. In addition, two sets of bipolar electrodes were used to record cardiac rhythm and eye movements. This allowed us to remove the electrocardiography (ECG) and electrooculography (EOG) artifacts in a later stage of the analysis pipeline.

The MRI scans were recorded on a CE-approved 3T Siemens MRI-scanner at AUH. The recorded data consisted of structural T1 (mprage with fat saturation) with a spatial resolution of 1.0 x 1.0 x 1.0 mm and the following sequence parameters: echo time (TE) = 2.61 ms, repetition time (TR) = 2300 ms, reconstructed matrix size = 256 x 256, echo spacing = 7.6 ms, bandwidth = 290 Hz/Px.

The MEG and MRI recordings were acquired in two separate days.

Behavioral data

We obtained behavioral data (number of correctly recognized trials and correspondent reaction times) from the experimental task carried out during the MEG recording.

Since the data was not normally distributed, we computed two independent Kruskal-Wallis H tests [6] (non-parametric one-way analysis of variance) to assess whether the five categories of temporal sequences (M, NT1, NT2, NT3, NT4, NT5) differed in terms of correct responses and reaction times. Multiple comparisons were corrected using the Tukey-Kramer correction [7].

MEG data preprocessing

The raw MEG sensor data (204 planar gradiometers and 102 magnetometers) was first pre-processed by MaxFilter [8] to attenuate external interferences. We applied signal space separation (MaxFilter parameters: spatiotemporal signal space separation [SSS], down-sample from 1000Hz to 250Hz, movement compensation using cHPI coils [default step size: 10 ms], correlation limit between inner and outer subspaces used to reject overlapping intersecting inner/outer signals during spatiotemporal SSS: 0.98).

The data was then converted into Statistical Parametric Mapping (SPM) format and further preprocessed and analyzed in MATLAB (MathWorks, Natick, MA, USA) using a combination of in-house-built codes (LBPD, https://github.com/leonardob92/LBPD-1.0.git) and the Oxford Centre for Human Brain Activity (OHBA) Software Library (OSL) [9] (https://ohba-analysis.github.io/osl-docs/), a freely available software that builds upon Fieldtrip [10], FSL [11], and SPM [12] toolboxes.

The continuous MEG data was visually inspected to identify and remove large artifacts using the OSLview tool. The data that was removed was less than 0.1% of the amount of collected data. Independent component analyses (ICA) were used to discard the interference of eyeblinks and heart-beat artefacts from the brain data [13]. First, we decomposed the original signal into independent components. Second, we isolated and discarded the

components that picked up eyeblink and heart-beat activities. Third, the signal was rebuilt using the remaining components. Finally, the signal was epoched in 135 trials (27 M, 27 NT1, 27 NT2, 27 NT3, 27 NT4) and baseline-corrected by removing the mean signal recorded in the baseline from the post-stimulus brain signal. Each trial lasted 4500 ms (4400 ms plus 100 ms of baseline time).

Multivariate pattern analysis (decoding)

We performed multivariate pattern analyses to decode different neural activity associated with the recognition of M versus N. Here, we computed four independent analyses, decoding M from each of the four categories of the novel sequences (i.e., M versus NT1, M versus NT2, M versus NT3, M versus NT4).

We used support vector machines (SVMs) [14] and calculated independent analyses for each participant. The MEG data was rearranged in a 3D matrix (channels x timepoints x trials) and submitted to the SVM algorithm. To avoid overfitting, a leave-one-out cross-validation approach was adopted to train the SVM classifier to decode the two experimental conditions. This procedure divided the trials into N different groups (here $N = 8$). Then, for each timepoint, it assigned $N - 1$ groups to the training set and the remaining N^{th} group to the testing set. After that, the classifier ability to separate the two conditions was evaluated. This process was performed 100 times with random reassignment of the data to training and testing sets. To summarize, the decoding accuracy time series were averaged to obtain a final time series showing the performance of the classifier for each participant.

To test the significance of the decoding results (chance level set at 50%), we employed a sign permutation test against the chance level for each timepoint and then corrected for multiple comparisons using false-discovery rate (FDR) correction ($\alpha = .05$; FDR-adjusted $q < .012$).

To assess whether each pair of conditions were differentiated by neural patterns which were stable over time, we computed four temporal generalization multivariate analyses. The algorithm was the same as the one previously described. However, in this case we utilized each timepoint of the training set to predict not only the same timepoint in the testing set, but all timepoints [15, 16]. Here, the significance was tested using a signed permutation test against the chance level (50%) for each timepoint, as the previous analyses. Then, we corrected for multiple comparisons using two-dimensional (2D) cluster-based Monte-Carlo simulations (MCS, $\alpha = .01$, MCS *p-value* = .001) [1-5, 17]. First, we computed the clusters size of the

continuous, binarized, significant values in time. Second, we made 1000 permutations of these binarized values. For each permutation, we computed the size of the maximum emerging cluster and built a reference distribution using those values. Finally, we considered significant the original clusters that were bigger than the 99.99% of the permuted data maximum cluster sizes.

Source reconstruction

MEG is a powerful tool to detect the whole-brain activity with excellent temporal resolution. However, to obtain a complete picture of the whole-brain activity underlying complex cognitive tasks the spatial component of the brain activity must be also identified. Here, we employed the established beamforming method [18-20], built upon a combination of in-house-built codes and codes available in OSL, SPM, and FieldTrip.

To reconstruct the brain sources that generated the MEG signal, an inverse problem must be solved. The MEG recording shows the activity of the neural signals outside the head but cannot provide information on the specific brain sources which generated it. Thus, we used beamforming algorithms to solve this problem, implementing the two following steps: (*i*) designing a forward model and (*ii*) computing the inverse solution.

The forward model is a theoretical model which considers each brain source as an active dipole (brain voxel). It describes how the unitary strength of each dipole would be reflected over all MEG sensors. Here, we employed magnetometer channels and an 8-mm grid, which returned 3559 dipole locations (voxels) within the whole brain. After co-registering the individual structural T1 data with the fiducial points (i.e., information about head landmarks), we computed the forward model by adopting the widely used method called "Single Shell", which is presented in detail in Nolte [21]. The output of this computation, referred to as "leadfield model", was stored in the matrix L (sources x MEG channels). In the three cases where the structural T1 was not available we performed the leadfield computation using a template (MNI152-T1 with 8-mm spatial resolution).

Then, we computed the inverse solution. As mentioned above, we chose the beamforming, which is one of the most popular and effective algorithms available in the field. This procedure employs a different set of weights which are sequentially applied to the source locations for isolating the contribution of each source to the activity recorded by the MEG channels. This is done for each timepoint of the recorded brain data. The beamforming inverse solution can be summarized by the following main steps.

The data recorded by MEG sensors (*B*) at time *t*, can be described by the following equation (1):

$$B_{(t)} = L * Q_{(n_i,t)} + \qquad (1)$$

where *L* is the leadfield model, *Q* is the dipole matrix carrying the activity of each active dipole (*q*) over time and is noise (see Huang and colleagues [22] for details). To solve the inverse problem, *Q* must be computed. In the beamforming algorithm, weights are computed and then applied to the MEG sensors at each timepoint, as shown for the single dipole *q* in equation (2):

$$q_{(t)} = W^T * B_{(t)} \qquad (2)$$

To obtain *q*, the weights *W* should be computed (the subscript *T* refers to transpose matrix). To this goal, the beamforming relies on the matrix multiplication between *L* and the covariance matrix between MEG sensors (*C*), which is calculated on the concatenated experimental trials. Specifically, for each brain source *n*, the weights W_n are computed as shown in equation (3):

$$W_{(n)} = (L_{(n)}{}^T * C^{-1} * L_{(n)})^{-1} * L_{(n)}{}^T * C^{-1} \qquad (3)$$

To be noted, the computation of the leadfield model was performed for the three main orientations of each brain source (dipole), according to Nolte [21]. Before computing the weights, the orientations were reduced to one using the singular value decomposition algorithm on the matrix multiplication reported in equation (4). This procedure is widely adopted to simplify the beamforming output [9, 22].

$$L = svd(l^T * C^{-1} * l)^{-1} \qquad (4)$$

Here, *l* represents the leadfield model with the three orientations, while *L* is the resolved one-orientation model that was utilized in (3). Finally, the weights were applied to each brain source and timepoint. To be noted, the covariance matrix C was computed on the continuous signal, which was estimated by concatenating the trials of all experimental conditions. The

weights were applied to the brain data associated with each condition and normalized according to Luckhoo et al. [23] for counterbalancing the reconstruction bias towards the centre of the head. The weights were applied to the neural activity averaged over trials for the evoked responses and to the neural activity of each independent trial for the induced responses. This procedure returned a time series for each of the 3559 brain sources (and each trial in the case of induced responses). The sign ambiguity of the evoked responses time series was adjusted for each brain source using its sign in correspondence with the N100 response to the first tone of the auditory sequences [5].

Neural sources of the differential brain activity between M and N

To detect the brain sources underlying the differential signal observed for M and N, we considered both the results from the decoding analyses and the inspection of the MEG data after preprocessing. We calculated this to identify which condition was association with a larger neural signal. If condition one is significantly stronger than condition two at time t = x and then condition two is significantly stronger than condition one at time t = x + 1, the decoding will return an overall significant difference between conditions from t = x until t = x + 1, even if such difference is qualitatively not the same at t = x and at t = x +1. Thus, to define the time windows for inspecting the neural sources, it is good practice to look both at significant results from decoding and the correspondent brain activity in the MEG data after preprocessing.

This procedure, applied to all our four contrasts, returned the following time windows: 0.50 – 0.60 sec; 0.70 – 0.80 sec; 0.98 – 1.02 sec; 1.05 – 1.15 sec; 1.33 – 1.39 sec; 1.45 – 1.55 sec; 1.70 – 1.75 sec; 1.75 – 1.85 sec. For each time window and condition, we averaged the time series of all brain sources over time and computed t-tests contrasting M versus N (t-tests were computed independently for M versus each of the four categories of N). Finally, we corrected for multiple comparisons using a 3D cluster-based MCS (MCS, α = .003, MCS *p-value* = .001). Here, we calculated the sizes of the clusters of neighbouring brain voxels which were significant. Then, we computed 1000 permutations of the original data. For each permutation, we estimated the sizes of the clusters of neighbouring permuted brain voxels which were significant. This returned a reference distribution of the biggest cluster sizes observed in the permutated data. Finally, we considered significant the original clusters that were bigger than the 99.99% of clusters forming the reference distribution. Additional details on the MCS algorithm can be found in [1, 2, 4, 5, 17].

Functional regions of interests (ROIs)

Our previous analyses highlighted a network mainly comprising four broad brain areas which were involved in the task. These areas roughly corresponded to the bilateral medial cingulate gyrus (MC), right hippocampal area and inferior temporal cortex (HITR), left auditory cortex (ACL), and bilateral ventro-medial prefrontal cortex (VMPFC).

Then, we wished to refine the spatial extent of those areas based on their functional profile and obtain their associated time series. Thus, first we computed *t*-values for each brain voxel and each timepoint contrasting M versus N. Second, we isolated the strongest *t*-value in absolute terms for each of the four broad regions identified in our previous analysis. This allowed us to identify the peaks of differential activity occurring between M and N for each ROI. Third, we used those peaks (averaged in a time window of ± 20 ms) and strict *t*-value thresholds (abs(t) > 3) to isolate the brain voxels that were mainly contributing to discriminate M versus N. This procedure refined the spatial extent of the four broad ROIs that we previously identified. Finally, to cover potential hemispheric differences, we created two more ROIs which mirrored HITR and ACL in the opposite hemisphere (HITL and ACR, respectively). Once we defined these six broad ROIs, we computed the time series showing their activity over time by averaging the time series of each of the brain voxels forming every ROI. To be noted, the spatial accuracy of the reconstructed MEG signal cannot be completely accurate, thus it is good practice employing such broad ROIs [24, 25].

Statistical analysis on ROIs time series

We employed the time series of the previously identified ROIs to compute additional statistics between M and N conditions. Here, we computed one t-test for each timepoint and each combination of M versus Ns (i.e., M versus NT1, M versus NT2, M versus NT3, M versus NT4). Then, we corrected for multiple comparisons using a one-dimensional (1D) cluster-based MCS (MCS, α = .05, MCS *p-value* = .001). First, we identified the clusters of significant continuous values in time. Second, we computed 1000 permutations, randomizing the significant values obtained from the t-tests. For each permutation, we extracted the maximum cluster size and built their reference distribution. Finally, we considered significant the original clusters that were larger than 99.99% of the permuted ones.

Time-frequency analysis for induced responses

We computed a time-frequency analysis using complex Morlet wavelet transform (from 1 to 60 Hz with 1-Hz intervals) [26]. This analysis was conducted for induced responses, independently for the six ROIs previously described and for the four contrasts considered in this study (i.e., M versus NT1, M versus NT2, M versus NT3, M versus NT4). Specifically, the time-frequency decomposition was done independently for each trial, brain voxel, and participant. Then, the power spectrum of each trial and each brain voxel was averaged within each of the six ROIs.

Finally, we computed a t-test for each frequency and timepoint, making four contrasts: M versus NT1, M versus NT2, M versus NT3, M versus NT4. The emerging *p*-values were binarized (α = .05) and then submitted to a 2D MCS (MCS *p*-value = .001). Here, we calculated the clusters size of continuous significant values in time and frequency. Then, we made 1000 permutations of the binarized *p*-values. For each permutation, we measured the size of the maximum emerging cluster and built a reference distribution with one value for each permutation. Finally, the original clusters were considered significant when they were bigger than the 99.99% of the permuted data maximum cluster sizes.

References (Methods only)

69. Bonetti, L., Brattico, E., Carlomagno, F., Donati, G., Cabral, J., Haumann, N.T., Deco, G., Vuust, P., & Kringelbach, M.L. Rapid encoding of musical tones discovered in whole-brain connectivity. *NeuroImage* **245**, 118735; 10.1016/j.neuroimage.2021.118735 (2021).
70. Bonetti, L., Brattico, E., Carlomagno, F., Cabral, J., Stevner, A., Deco, G., Whybrow, P.C., Pearce, M., Pantazis, D., & Vuust, P. Spatiotemporal brain dynamics during recognition of the music of Johann Sebastian Bach. *bioRxiv*, (2020).
71. Fernández Rubio, G., Brattico, E., Kotz, S.A., Kringelbach, M.L., Vuust, P., & Bonetti, L. The spatiotemporal dynamics of recognition memory for complex versus simple auditory sequences. *bioRxiv*, 2022.05.15.492038; 10.1101/2022.05.15.492038 (2022).
72. Fernández Rubio, G., Carlomagno, F., Vuust, P., Kringelbach, M.L., & Bonetti, L. Associations between abstract working memory abilities and brain activity underlying long-term recognition of auditory sequences. *bioRxiv*, (2022).
73. Bonetti, L., Brattico, E., Bruzzone, S.E.P., Donati, G., Deco, G., Pantazis, D., Vuust, P., & Kringelbach, M.L. Temporal pattern recognition in the human brain: a dual simultaneous processing. *bioRxiv*, (2021).
74. Kruskal, W.H., & Wallis, W.A. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* **47**, 583-621; (1952).
75. Tukey, J.W. Comparing individual means in the analysis of variance. *Biometrics* **5**, 99-114; (1949).
76. Taulu, S., & Simola, J. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology* **51**, 1759-68; 10.1088/0031-9155/51/7/008 (2006).
77. Woolrich, M., Hunt, L., Groves, A., & Barnes, G. MEG beamforming using Bayesian PCA for adaptive data covariance matrix regularization. *NeuroImage* **57**, 1466-1479; (2011).
78. Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* **2011**, 156869; 10.1155/2011/156869 (2011).

79. Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., & Smith, S.M. Bayesian analysis of neuroimaging data in FSL. *NeuroImage* **45**, S173-86; 10.1016/j.neuroimage.2008.10.055 (2009).
80. Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., & Nichols, T.E., *Statistical parametric mapping: the analysis of functional brain images*. 2011: Elsevier.
81. Mantini, D., Della Penna, S., Marzetti, L., de Pasquale, F., Pizzella, V., Corbetta, M., & Romani, G.L. A signal-processing pipeline for magnetoencephalography resting-state networks. *Brain Connectivity* **1**, 49-59; 10.1089/brain.2011.0001 (2011).
82. Wilson, M., *Support vector machines*, in *Encyclopedia of Ecology*, P. Robbins, Editor. 2008. p. 3431-3437.
83. Cichy, R.M., Pantazis, D., & Oliva, A. Resolving human object recognition in space and time. *Nature Neuroscience* **17**, 455-462; 10.1038/nn.3635 (2014).
84. King, J.R., & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences* **18**, 203-10; 10.1016/j.tics.2014.01.002 (2014).
85. Bruzzone, S.E.P., Lumaca, M., Brattico, E., Vuust, P., Kringelbach, M.L., & Bonetti, L. Dissociated brain functional connectivity of fast versus slow frequencies underlying individual differences in fluid intelligence: a DTI and MEG study. *Scientific Reports* **12**, 4746; 10.1038/s41598-022-08521-5 (2022).
86. Hillebrand, A., & Barnes, G.R. Beamformer analysis of MEG data. *International Review of Neurobiology* **68**, 149-71; 10.1016/S0074-7742(05)68006-3 (2005).
87. Huang, M.X., Mosher, J.C., & Leahy, R.M. A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. *Physics in Medicine and Biology* **44**, 423-40; 10.1088/0031-9155/44/2/010 (1999).
88. Brookes, M.J., Stevenson, C.M., Barnes, G.R., Hillebrand, A., Simpson, M.I., Francis, S.T., & Morris, P.G. Beamformer reconstruction of correlated sources using a modified source model. *NeuroImage* **34**, 1454-65; 10.1016/j.neuroimage.2006.11.012 (2007).
89. Nolte, G. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology* **48**, 3637; (2003).
90. Huang, M.X., Shih, J., Lee, R., Harrington, D., Thoma, R., Weisend, M., Hanlon, F., Paulson, K., Li, T., & Martin, K. Commonalities and differences among vectorized

beamformers in electromagnetic source imaging. *Brain Topography* **16**, 139-158; (2004).

91. Luckhoo, H.T., Brookes, M.J., & Woolrich, M.W. Multi-session statistics on beamformed MEG data. *NeuroImage* **95**, 330-5; 10.1016/j.neuroimage.2013.12.026 (2014).

92. Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., & Schoffelen, J.M. Good practice for conducting and reporting MEG research. *NeuroImage* **65**, 349-63; 10.1016/j.neuroimage.2012.10.001 (2013).

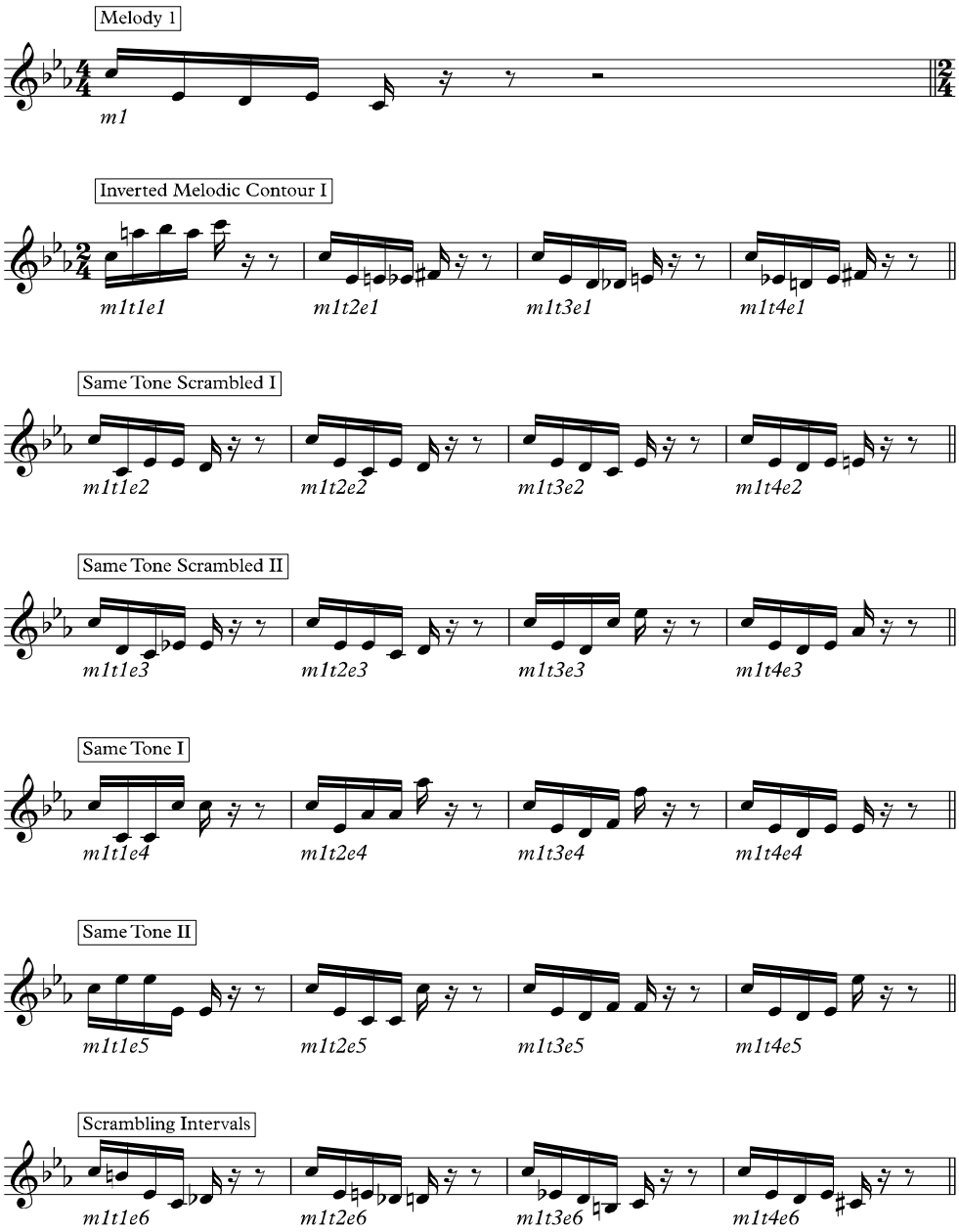
93. Pernet, C., Garrido, M.I., Gramfort, A., Maurits, N., Michel, C.M., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., & Puce, A. Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature Neuroscience* **23**, 1473-1483; 10.1038/s41593-020-00709-0 (2020).

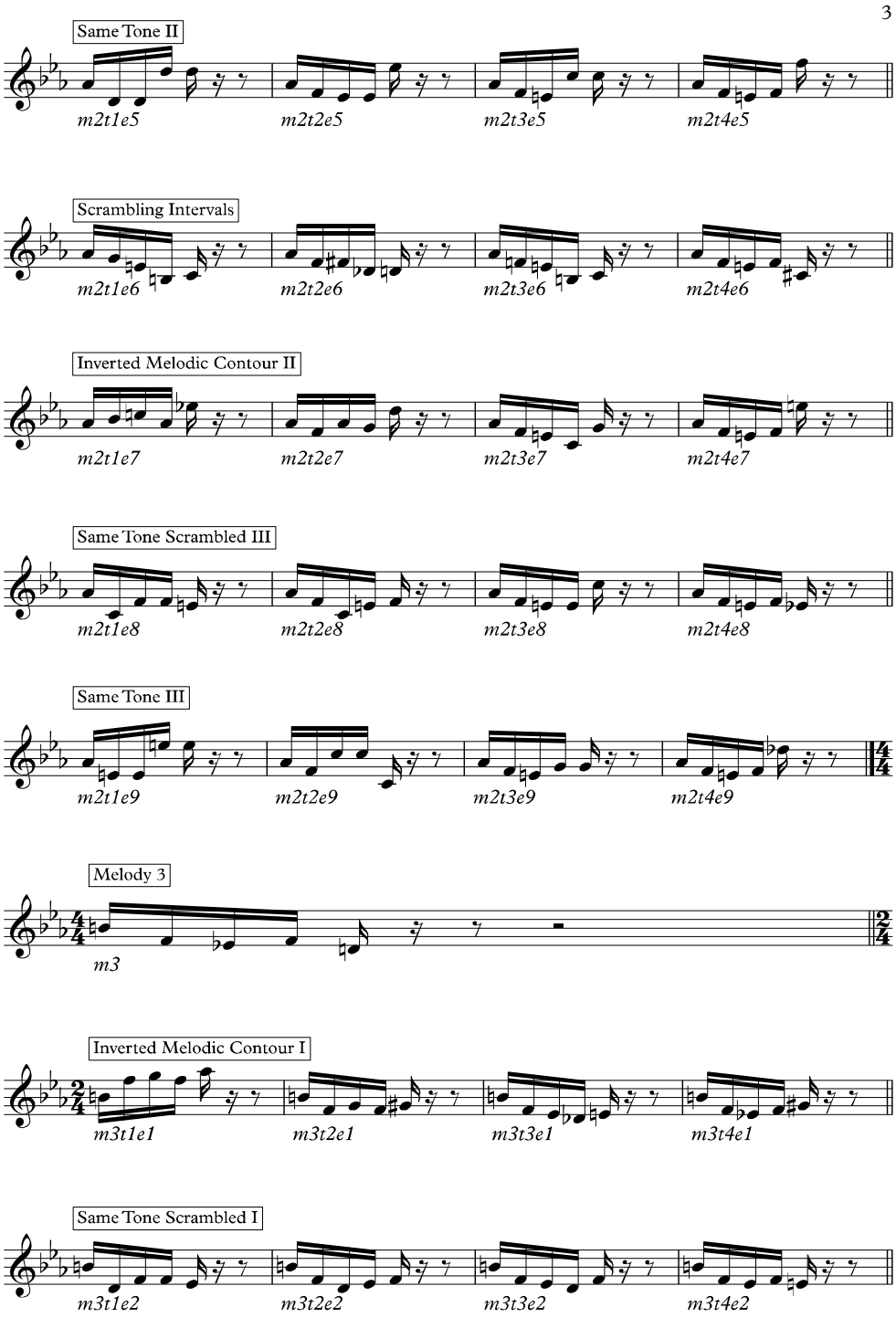
94. Daubechies, I., *Ten lectures on wavelets*. 1992: SIAM.

SUPPLEMENTARY MATERIAL

Supplementary materials related to this study and organized as supplementary figures (*i*) and tables (*ii*). In the cases when the supplementary tables were too large to be reported in the current document, they have been exported to Excel files that can be found at the following link:
https://drive.google.com/drive/folders/1W1w8UpPKnyp0RMjksKmxi3XqC6UwseBY?usp=sharing

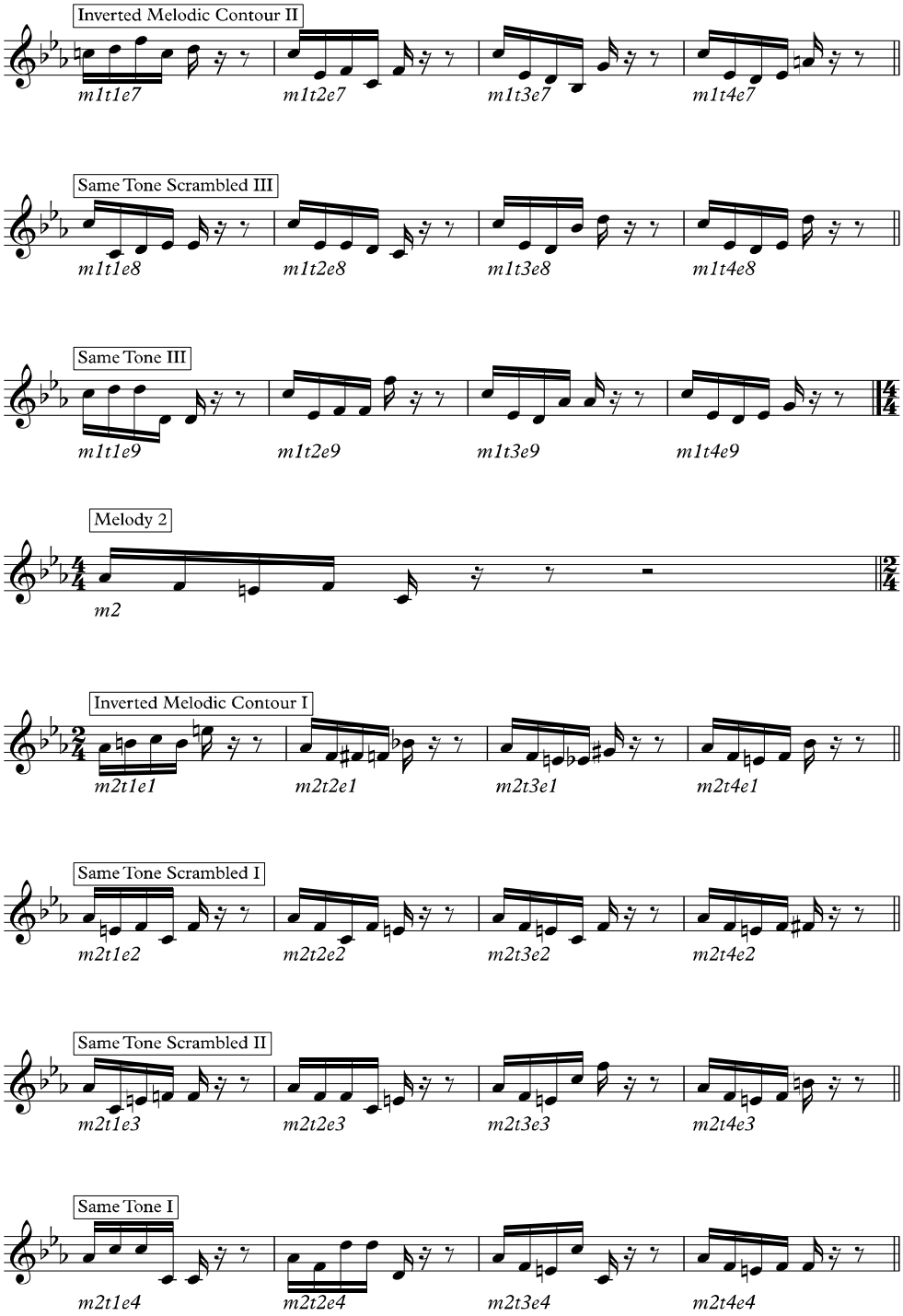
SUPPLEMENTARY FIGURES




3
Same Tone II
m2t1e5
m2t2e5
m2t3e5
m2t4e5
Scrambling Intervals
m2t1e6
m2t2e6
m2t3e6
m2t4e6
Inverted Melodic Contour II
m2t1e7
m2t2e7
m2t3e7
m2t4e7
Same Tone Scrambled III
m2t1e8
m2t2e8
m2t3e8
m2t4e8
Same Tone III
m2t1e9
m2t2e9
m2t3e9
m2t4e9
Melody 3
m3
Inverted Melodic Contour I
m3t1e1
m3t2e1
m3t3e1
m3t4e1
Same Tone Scrambled I
m3t1e2
m3t2e2
m3t3e2
m3t4e2

1180

2


Inverted Melodic Contour II
m1t1e7
m1t2e7
m1t3e7
m1t4e7
Same Tone Scrambled III
m1t1e8
m1t2e8
m1t3e8
m1t4e8
Same Tone III
m1t1e9
m1t2e9
m1t3e9
m1t4e9
Melody 2
m2
Inverted Melodic Contour I
m2t1e1
m2t2e1
m2t3e1
m2t4e1
Same Tone Scrambled I
m2t1e2
m2t2e2
m2t3e2
m2t4e2
Same Tone Scrambled II
m2t1e3
m2t2e3
m2t3e3
m2t4e3
Same Tone I
m2t1e4
m2t2e4
m2t3e4
m2t4e4

1181

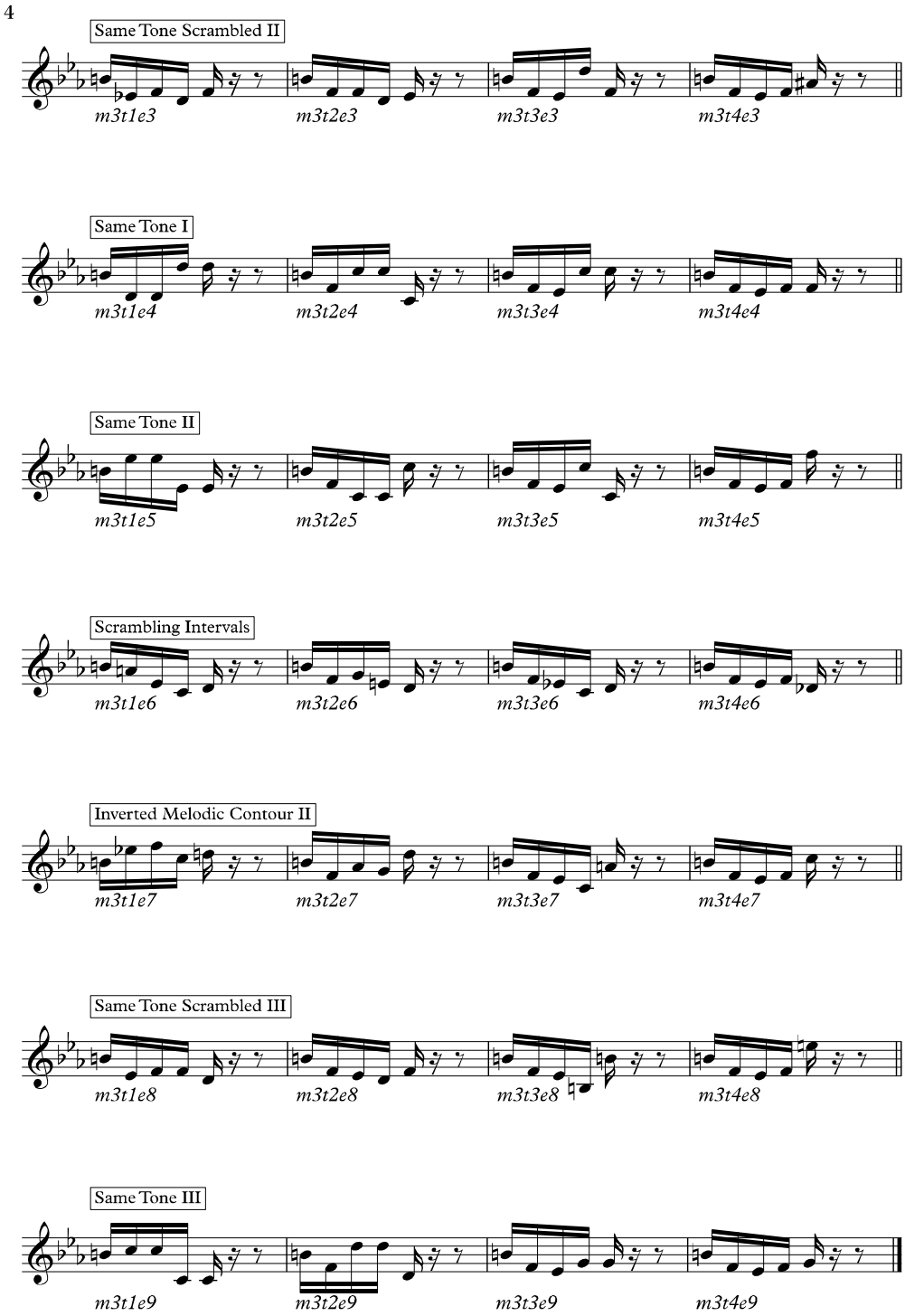

4
Same Tone Scrambled II
m3t1e3
m3t2e3
m3t3e3
m3t4e3
Same Tone I
m3t1e4
m3t2e4
m3t3e4
m3t4e4
Same Tone II
m3t1e5
m3t2e5
m3t3e5
m3t4e5
Scrambling Intervals
m3t1e6
m3t2e6
m3t3e6
m3t4e6
Inverted Melodic Contour II
m3t1e7
m3t2e7
m3t3e7
m3t4e7
Same Tone Scrambled III
m3t1e8
m3t2e8
m3t3e8
m3t4e8
Same Tone III
m3t1e9
m3t2e9
m3t3e9
m3t4e9

Figure S1. Temporal sequences used in the experiment

The figure shows all temporal sequences used in the experiment, providing detailed information on how they were created. The M sequences were three and comprised the first five tones of the first three measures of the musical piece. These three sequences were presented nine times each, for a total of 27 trials. The N sequences were created through systematic variations of the three M sequences. This procedure consisted of changing every musical tone of the sequence after the first (NT1), second (NT2), third (NT3) or fourth (NT4) tone. We created nine variations for each of the original M sequences and each of the four categories of N. This resulted in 27 N sequences for each category, and 108 N in total. To be noted, as shown in this figure, the variations were created according to the following rules: (i) Inverted melodic contours (used twice): the melodic contour of the variation was inverted with respect to the original M sequence (i.e., if the M sequence had the following melodic contour: down-down-up-down, the N sequence would be: up-up-down-up); (ii) Same tone scrambled (used three times): the remaining tones of the M sequence were scrambled (e.g., M sequence: C-E-D-E-C, was converted into NT1 sequence: C-C-E-E-D). When this was not possible (e.g., in the case of NT4, where only the last tone is different from the M sequence), we substituted the last tone of the M sequence with a random tone; (iii) Same tone (used three times): the same tone was repeatedly used, in some cases varying only the octave (e.g., M sequence: C-E-D-E-C, was transformed into NT1 sequence: C-E^{8-} E^{8-} E_8^{-} E_8); (iv) Scrambling intervals (used once): the intervals between the tones were scrambled (e.g., M sequence: 6^{th}m - 2^{nd}m – 2^{nd}m – 3^{rd}m, was adapted to NT1 sequence: 2^{nd}m, 6^{th}m, 3^{rd}m, 2^{nd}m).

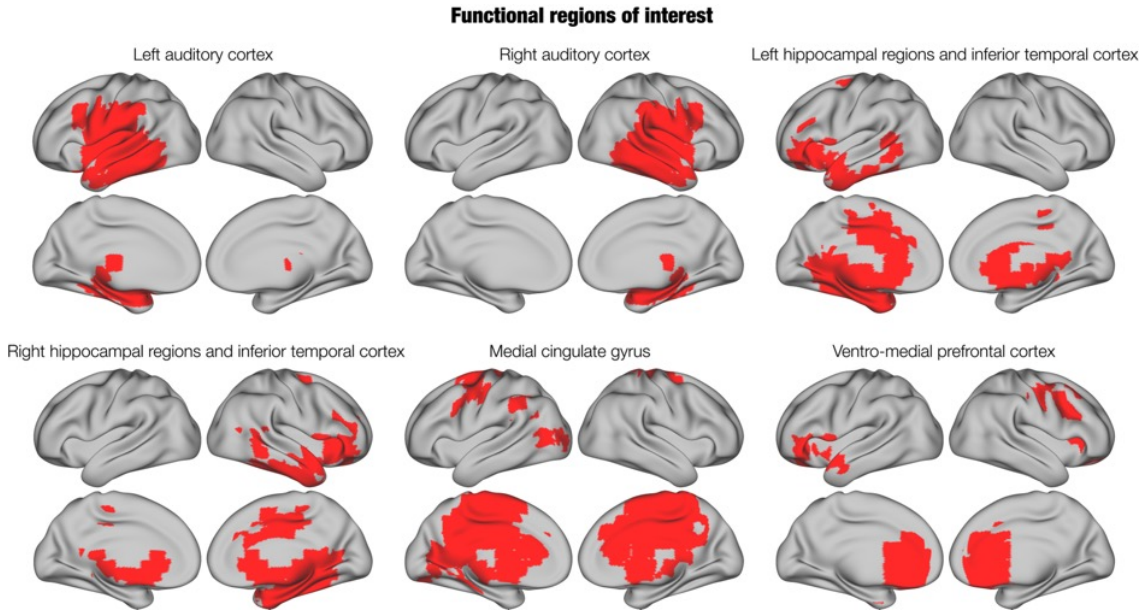


Figure S2. Functional parcels (ROIs) derived from the brain activity underlying the task

The main activity during recognition of the previously memorized and novel auditory sequences gave rise to the following six functional ROIs: left (i) and right auditory cortex (ii); left (iii) and right hippocampal regions and inferior temporal cortex (iv); medial cingulate gyrus (v), and ventro-medial prefrontal cortex (vi).

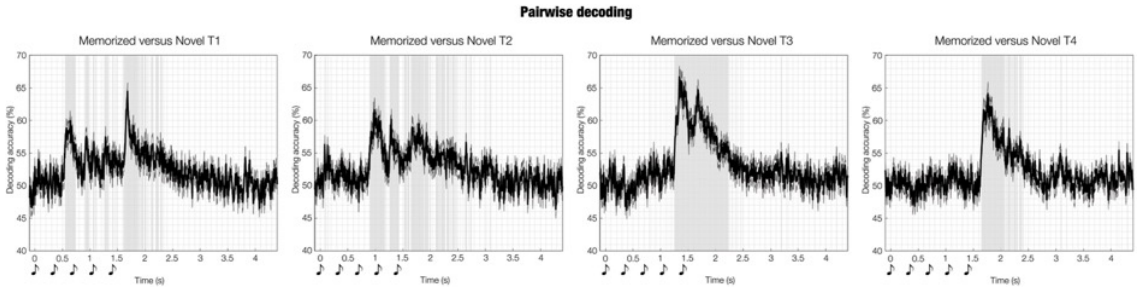


Figure S3. Pairwise decoding time series

Multivariate pattern analysis decoding the different neural activity associated with memorized versus novel musical sequences. Each plot shows the decoding time series for one of the four rounds of pairwise decoding that we computed (M vs NT1, M vs NT2, M vs NT3, M vs NT4). The sketch of the musical tones represents the onset of the sounds forming the temporal sequences.

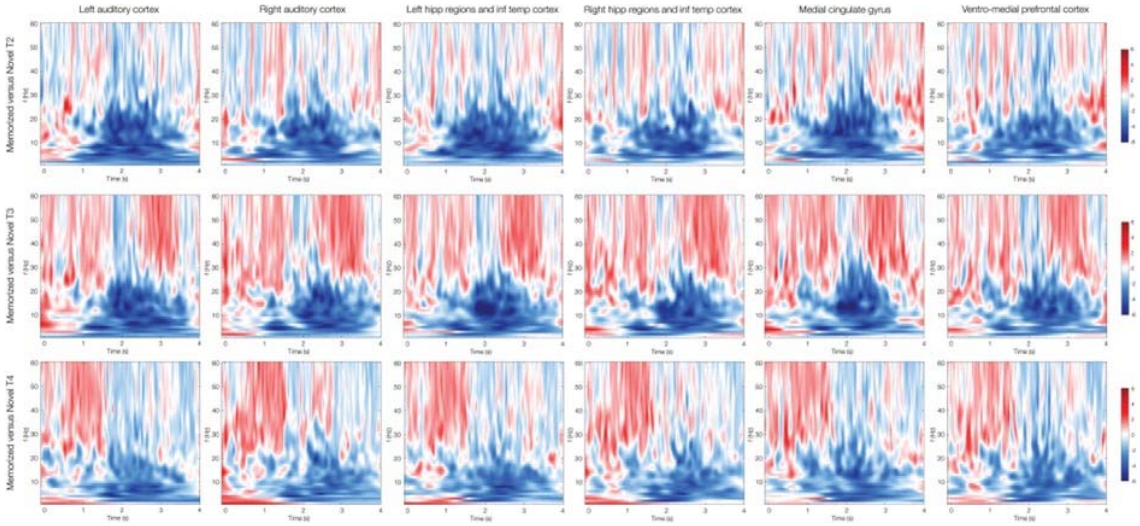


Figure S4. Source localized induced responses – M versus NT2, NT3, NT4

Contrasts between the source localized induced responses of M versus NT2, NT3, NT4, respectively. The plots indicate a stronger power for gamma in M. Moreover, theta was overall stronger for M versus NT1 during the presentation of the sounds, while alpha, beta and theta were stronger for N versus M after the end of the temporal sequences. The colorbar indicates the t-values obtained by contrasting M versus N.

SUPPLEMENTARY TABLES

Table S1. Pairwise decoding

Binary time series showing the FDR-corrected significant timepoints (1s) of the decoding time series (i.e., when the algorithm successfully classified M versus N). The first row shows time (in seconds), while the other rows refer to the four contrasts of this study (M versus NT1, M versus NT2, M versus NT3, M versus NT4).

Table S2. Temporal generalization

Cluster-based MCS on temporal generalization decoding results computed independently for the four following contrasts: M versus NT1, M versus NT2, M versus NT3, M versus NT4. The table shows size, MCS p-value and temporal extent of the cluster (both training and testing sets).

Table S3. Brain source of decoding time windows

Significant brain sources (after cluster based MCS correction for multiple comparisons) of the significant time windows emerged from the decoding analysis. Results are reported with the correspondent AAL label of each of the significant voxel, as well as their hemisphere, t-value and MNI coordinates. Results are provided for the following contrasts: M versus NT1, M versus NT2, M versus NT3, M versus NT4.

Table S4. ROIs time series

Significant clusters of differential brain activity between M and N in the six broad functional ROIs isolated in the previous analyses. Results are reported independently for the six ROIs and for the four contrasts (M versus NT1, M versus NT2, M versus NT3, M versus NT4), and comprise cluster size, p-value, temporal extent of the clusters and peak t-value within the cluster.

Table S5. Time-frequency results of induced responses

Significant clusters of differential power in different frequency bands (1 – 60Hz) computed using complex Morlet wavelet transform. Results are reported independently for the six ROIs and for the four contrasts (M versus NT1, M versus NT2, M versus NT3, M versus NT4), and comprise cluster size, p-value, temporal, and frequency extent of the clusters.