



**HAL**  
open science

## **Brain dynamics of the interplay between auditory selective attention and working memory during melody encoding**

Aurélien de la Chapelle, Salomé Serres-Blain, Alma Elshafei, Denis P Schwartz, Sébastien Daligault, Julien Jung, Perrine Ruby, Anne Caclin, Aurélie Bidet-Caulet

### ► **To cite this version:**

Aurélien de la Chapelle, Salomé Serres-Blain, Alma Elshafei, Denis P Schwartz, Sébastien Daligault, et al.. Brain dynamics of the interplay between auditory selective attention and working memory during melody encoding. *Scientific Reports*, 2025, 15 (1), pp.22453. <10.1038/s41598-025-03378-w>. <hal-05139162v2>

**HAL Id: hal-05139162**

**<https://hal.science/hal-05139162v2>**

Submitted on 4 Jul 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



## OPEN Brain dynamics of the interplay between auditory selective attention and working memory during melody encoding

Aurélien de la Chapelle<sup>1,7</sup>, Salomé Serres-Blain<sup>1,7</sup>, Alma ElShafei<sup>1,2</sup>, Denis Schwartz<sup>3</sup>, Sébastien Daligault<sup>3</sup>, Julien Jung<sup>1,4</sup>, Perrine Ruby<sup>1</sup>, Anne Caclin<sup>1,7</sup> & Aurélie Bidet-Caulet<sup>1,5,6,7</sup>✉

Working memory and attention are jointly needed in most everyday life tasks and activities. They have however mostly been studied separately. Here we investigate how auditory working memory—in a delayed-matching-to-sample task—and selective attention interact using a recently introduced paradigm (MEMAT, for MEMory and ATtention) and MEG (Magneto-encephalographic) recordings in twenty-two participants. We manipulate the difficulty of the memory task and attentional filtering. *When memory task difficulty increases, accuracy decreases, CNV (Contingent Negative Variation) amplitude in anticipation of the melody to encode increases, and the decrease in alpha power during encoding and maintenance in a left fronto-temporal network is reduced. When attentional filtering difficulty increases, accuracy decreases, the amplitude of the sustained evoked response during encoding increases, whereas the differential processing of relevant and irrelevant sounds in auditory areas is less pronounced, and frontal theta power during encoding and maintenance is higher. In the left auditory cortex, we directly observe the result of the interaction between auditory memory and attention: the facilitation of relevant sound processing in the easy filtering condition reduces when the memory task difficulty increases. This pattern mirrors the observed behavioural effects: a smaller difference in accuracy between attention conditions when the memory task difficulty increases. Overall brain dynamics highlight reciprocal influences of working memory and selective attention processes, in keeping with shared cognitive resources between them.*

Attention and working memory are constantly at play in cognitive tasks and have been the subject of intense experimental work. Yet only a fraction of these studies examines the interplay between these two major cognitive processes. We focus here on auditory selective attention, working memory, and their interactions as revealed by brain imaging studies.

### Auditory selective attention and working memory: brain networks

Numerous studies have shown that auditory selective attention and working memory rely on distributed brain networks, encompassing auditory temporal areas, prefrontal, and parietal cortex. On the one hand, using fMRI or PET, auditory attention has been shown to enhance activation in the auditory cortices in the superior temporal gyri<sup>1–3</sup>. Similarly to the visual modality, top-down auditory attention is found to be supported by the dorsal frontoparietal network of attention (dorsal attention network, DAN), including the frontal eye fields, the inferior and middle frontal gyri, the anterior cingulate cortices, and parietal regions, with greater activation in the right hemisphere<sup>2,4,5</sup>. However, in the case of auditory spatial selective attention (attention to left or right ear), activation of this DAN and of the auditory cortices is greater contralaterally to the to-be-attended side<sup>2,6,7</sup>. On the other hand, auditory working memory networks have also been studied with fMRI and PET,

<sup>1</sup>Université Claude Bernard Lyon 1, INSERM, CNRS, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500 Bron, France. <sup>2</sup>Donders Institute for Brain, Cognition, and Behavior, Radboud University, 6525 EN Nijmegen, The Netherlands. <sup>3</sup>MEG Department, CERMEP–Imagerie du Vivant, Lyon, France. <sup>4</sup>Department of Functional Neurology and Epileptology, Hospices Civils de Lyon, Lyon, France. <sup>5</sup>Inserm, INS, Institut de Neurosciences Des Systèmes, Campus Timone, 27 BVD Jean Moulin, 13005 Marseille, France. <sup>6</sup>Aix-Marseille Univ, Marseille, France. <sup>7</sup>Aurélien de la Chapelle, Salomé Serres-Blain, Anne Caclin and Aurélie Bidet-Caulet have contributed equally to this work. ✉email: aurelie.bidet-caulet@inserm.fr

mostly using verbal stimuli (for a review see<sup>8</sup>) but also with melodic (tonal) material<sup>9,10</sup>. During sound sequence encoding, these neuroimaging studies have shown that activity increases in auditory cortices and the prefrontal cortex, with a left-hemisphere bias especially in prefrontal areas for verbal stimuli (meta-analysis in<sup>11</sup>), and more bilateral activity patterns for tonal stimuli<sup>12,13</sup>. The maintenance (rehearsal) of auditory information recruits prefrontal areas, especially in the left hemisphere for verbal stimuli<sup>11,13</sup> and bilaterally for tonal stimuli<sup>13</sup>, but also other areas including the hippocampus<sup>14</sup>, premotor and motor areas<sup>10,15</sup>, cerebellum<sup>12</sup>, and the Sylvian-parietal-temporal (Spt) area for verbal rehearsal (meta-analysis in<sup>11</sup>). The Intra-Parietal Sulcus (IPS) is involved in tonal working memory, especially when manipulation of the stored information is requested<sup>16,17</sup>.

Neuroimaging studies that directly compare the networks subtending selective attention and working memory are scarce, especially in the auditory modality. A notable exception is the study by Huang et al.<sup>18</sup>, where they manipulated the level of attentional interference and memory load in an auditory Continuous Performance Task. Both manipulations resulted in activity changes in a fronto-parietal network, which were largely overlapping, with however some regions being sensitive only to memory load (notably in the anterior dorsolateral prefrontal cortex) and others only to attentional interference (notably in the Inferior Frontal Cortex).

### Auditory selective attention and working memory: network dynamics

Auditory selective attention can operate on different features, notably space (e.g.,<sup>19</sup>), promoting the processing of information from a specific location; pitch (e.g.,<sup>20</sup>); and more generally on auditory objects<sup>21,22</sup>. The brain dynamics of auditory spatial selective attention have been extensively investigated using non-invasive EEG and MEG and the dichotic paradigm introduced by Hillyard and collaborators<sup>23</sup>. By comparing the evoked responses to the same sounds when task-relevant or irrelevant, attention has been shown to enhance transient responses as early as 20 ms after sound onset (e.g.,<sup>23–25</sup>; see<sup>26</sup> for a review) and sustained responses<sup>27,28</sup>. Importantly, recordings directly from the auditory cortices using intracortical EEG in epileptic patients have shown that auditory selective attention is supported by both an enhancement of evoked responses to relevant sounds and a suppression of evoked responses to irrelevant information, in primary and secondary auditory areas<sup>28</sup>. It suggests that both facilitatory and inhibitory mechanisms are at play (see also<sup>29–35</sup>). They are modulated by task difficulty: on one hand, the difficulty of the task at hand (e.g., more difficult discrimination of to-be-attended sounds) seems to increase the brain responses to to-be-attended sounds while not affecting the brain responses to to-be-ignored sounds<sup>36</sup>. On the other hand, the difficulty of distractor filtering was found to reduce the processing of to-be-attended sounds<sup>37–39</sup> or to increase the brain responses to to-be-ignored sounds<sup>40</sup>. Alpha oscillations have also been found to be involved in top-down auditory attention, both for spatial (e.g.,<sup>19</sup>) and object-based (e.g.,<sup>41</sup>) attention. During anticipation and processing of a target sound, the power of alpha oscillations decreases in task-relevant auditory regions and increases in task-irrelevant occipital regions<sup>19,42–47</sup>, in line with an inhibitory role of the alpha rhythm<sup>48–52</sup>. Interestingly, distinct frequency peaks in the alpha band have been found for the power increase in the occipital regions (around 11 Hz) and the power decrease in the auditory regions (around 9 Hz)<sup>45,46</sup>, emphasizing the importance of considering the alpha frequency peak in different brain regions (see also<sup>53</sup>). Beyond these correlative studies on the link between alpha oscillations and top-down attention, evidence for a causal involvement of alpha oscillations was obtained in the visual modality, thanks to training participants to modulate such oscillations with neurofeedback<sup>54</sup>.

EEG and MEG studies of auditory working memory have revealed a modulation of evoked responses by memory load and specific oscillatory signatures of working memory processes in various frequency bands. The amplitude of the N1 evoked response is related to memory performance: the N1 is reduced (and delayed) during encoding of short melodies in participants with congenital amusia who have a specific deficit of pitch short-term memory<sup>55</sup>. When memory load increases, the amplitude of sustained evoked responses during the maintenance in memory increases, with a specific component recorded over frontal areas reflecting auditory maintenance, the Sustained Anterior Negativity (SAN)<sup>56,57</sup>. Sustained evoked responses are also observed during the encoding of sound sequences, showing an increase in amplitude when the difficulty of the memory task increases<sup>55</sup>, which is very similar to the effect of attention observed during sound sequence processing<sup>28</sup>. Oscillatory activities have been largely studied during working memory tasks, in particular during the maintenance period, when there is no actual stimulus presented. In a recent systematic review, Pavlov and Kotchoubey<sup>58</sup> highlighted that the most consistent findings were an increase of frontal midline theta oscillations during both verbal and visuo-spatial working memory tasks, as well as an occipito-parietal alpha increase during verbal memory tasks. A majority of studies report that frontal theta oscillations increase with memory load, and these oscillations have been linked with the maintenance of items' order<sup>59</sup>, through a coupling with higher-frequency (gamma) activity<sup>60,61</sup>. Regarding alpha oscillations, the occipito-parietal alpha power increases with increasing memory load during verbal short-term memory tasks<sup>62</sup> and could reflect the inhibition of the task-irrelevant visual areas<sup>48</sup>, to protect the memory traces. EEG and MEG studies of oscillatory activities during working memory for non-verbal auditory material are far less numerous than studies about verbal working memory. However, for example during a melody short-term memory task, an increase of alpha power in task-irrelevant regions, namely the left fronto-temporal pathway and bilateral visual areas, has been observed<sup>63</sup>, suggesting similar principles of organization across auditory verbal and non-verbal domains.

### Interactions between selective attention and working memory

The studies reviewed above suggest: (1) largely overlapping brain networks involved in auditory selective attention and working memory, and (2) some similarities in how filtering out distractors and memory load impact on brain network dynamics. In particular, sustained evoked responses increase when selecting the relevant information in attention tasks and when memory load increases in working memory tasks. In both attention and working memory tasks, alpha oscillations increase in non-relevant brain areas. Some studies have tested more directly interactions between selective attention and working memory.

Such interactions have been studied extensively in the visual modality. One prominent finding is increased distraction effects at the behavioral and cerebral levels (increased responses to distractors) under high memory load in dual task-paradigms (one WM task and one attention task), in agreement with the Cognitive Load theory<sup>64–66</sup>. However, in studies using single tasks (one WM task with attentional filtering of distractors), reduced distraction has been found under high memory load<sup>67–69</sup>. Similarly, in the few studies exploring the effect of memory load on auditory selective attention, inconsistent results have been found, with evidence for both enhanced<sup>32</sup> or reduced distraction effects<sup>70–72</sup> with increasing memory load. The other way around, auditory distraction stemming from irrelevant speech or non-speech sounds has been shown to impede on WM performance, mostly assessed with visually-presented verbal material (e.g.<sup>73</sup>; review in<sup>74</sup>). Interestingly, modulations of both memory load and attentional filtering were found to be reflected in activity patterns in the DAN using fMRI<sup>75</sup>.

Such studies either manipulated memory load and assessed its impact on distractors processing, or manipulated filtering difficulty and assessed its impact on memory processes. A joint investigation of both types of manipulation, together with neurophysiological measurements, would however be needed to further understand the interplay between working memory and selective attention. Notably, none of the previous studies examined brain activity during encoding (and rehearsal) in memory of sound sequences when participants have to filter out distracting sounds during encoding (see<sup>70</sup>, for behavioral data with an auditory n-back task).

### The current study

In the current study, to shed light on the links between selective attention and working memory networks, we take advantage of a newly designed paradigm, MEMAT (for MEMory and ATtention)<sup>76</sup>, which allows to manipulate both selective attention and working memory processes during sound sequence encoding and rehearsal, using melodies (hence non-verbal auditory material). In MEMAT, participants perform a delayed-matching-to-sample task (DMST) using melodies presented in one (cued) ear. They have to indicate whether two consecutive sequences of four notes, separated by a two-second silent retention delay, are identical or not. During encoding of the first sequence, an interleaved-to-be-ignored melody is presented in the other ear. The to-be-ignored melody can be more or less similar to the to-be-attended melody, hence allowing to manipulate the difficulty to filter out distractors (selective attention manipulation), and the memory task require more or less precision during encoding, depending on blocks, thus allowing to manipulate the memory load<sup>76</sup>. In previous reports<sup>76,77</sup>, we have shown that behavioral results obtained with the MEMAT paradigm are in favor of some shared cognitive resources between working memory and selective attention in the auditory non-verbal domain, which can be interpreted within the Cognitive Load theory framework<sup>65,66</sup>, i.e., inhibition of distractors would be easier under low memory load. However the results also point to independent working memory and selective attention processes, as revealed by slightly different patterns of results in various participant groups: on the one hand, musicians exhibit greatly enhanced working memory abilities for musical sequences but only limited advantages if any for selective attention compared to non-musicians<sup>76</sup>, and on the other hand, participants with high dream recall frequency are more sensitive to the presence of difficult-to-filter-out distractors than participants with a low dream recall frequency, whereas the impact of the memory task difficulty manipulation was similar in these two participant groups<sup>77</sup>. These between-groups differences reveal that attention and working memory are differently sensitive to expertise effects (for the effect of musical expertise on auditory working memory, see for example<sup>78–81</sup>) and psychophysiological traits (for the effect of such traits on auditory attention, see<sup>82–87</sup>). Overall, behavioral results obtained so far with the MEMAT paradigm revealed both shared and separated processes subtending auditory selective attention and working memory. Here we use MEG recordings during performance of MEMAT to tackle the interplay between selective attention and working memory at the neurophysiological level, using non-verbal auditory sequences. Based on the literature reviewed above, we focus our investigation on both transient and sustained evoked responses, as well as on oscillatory activities in the theta and alpha range, in auditory, visual, and frontal areas. We could expect that, during sound sequence processing, (1) attentional filtering will be associated with a difference in N1 amplitude between targets and distractors, with larger differences revealing better filtering, (2) task difficulty would modulate sustained activity, with larger sustained activity in more difficult conditions, but previous studies did not allow to predict whether this would apply to the manipulation of memory and/or attention, and (3) task difficulty would modulate alpha oscillations in auditory networks, with lower alpha oscillations in difficult conditions, again without a specific hypothesis whether this would apply to memory and/or attention manipulation. Furthermore, in anticipation of the sound sequences, we expect to observe a CNV (magnetic counterpart of the Contingent Negative Variation) that could be modulated by the memory task difficulty, as well as modulations of theta and alpha oscillations compared to baseline. Finally, we analysed sustained activity and theta and alpha oscillations during the silent retention delay, which we also expected to be modulated by difficulty of the memory task and/or of attention filtering.

### Methods

The experimental protocol is adapted from the recent behavioral-only MEMAT study<sup>76</sup>. The main difference is that the current data is based on a much larger number of trials, to achieve a good signal-to-noise ratio in the MEG data analysis.

### Participants

Twenty-two non-musician participants (20 right-handed, 11 men and 11 women, aged 20–33 years, mean education level  $\pm$  standard deviation, SD: 15.5  $\pm$  2.0 years) participated in the experiment. Participants were considered as non-musicians when they practiced less than 1 year of instrument or singing outside compulsory educative programs (mean music education level  $\pm$  SD: 0.4  $\pm$  0.7 years). Note that these participants were selected

to have low or high frequency of dream recall, in order to be able to study the relationship between dream recall frequency (DRF), attention, and short-term memory in analyses not presented here. Behavioral data collected in the present study were analyzed as a function of participants' DRF in Blain, de la Chapelle et al.<sup>77</sup>. Detailed neuropsychological testing and questionnaire data can be found in Blain, de la Chapelle et al.<sup>77</sup>. The present report focuses on MEG data combined from all participants.

All participants were free from neurological or psychiatric disorder and had normal hearing and normal vision. All subjects gave written informed consent to participate. Experimental procedures and methods were approved by the appropriate regional ethics committee (CPP Sud-Est III, ID-RCB: 2014-A01289-38) and were performed in accordance with the Declaration of Helsinki.

## Stimuli

240 four-sound-long melodies were created to be used as S1 melodies (see below) thanks to combinations of eight-harmonic synthetic sounds from the C major scale, spanning four octaves between 65 and 1046 Hz. The maximal interval between any two sounds of a melody was 7 semi-tones. Sounds were 250-ms long and the interval between sounds (offset to onset) was 250 ms. In each melody, there were no consecutive identical sounds. Every melody contained at least an ascending and a descending interval.

## Paradigm

### *Attention manipulation*

The to-be-attended melody (S1) and the to-be-ignored melody (DIS) were played one in each ear. The tones of the two melodies were interleaved, thus not played simultaneously. S1 was played in the ear indicated by an arrow on the screen, and DIS was played in the other ear, each tone one after the other. The first sound to be played was the first tone of S1, and the last tone to be played was the last tone of DIS (see Fig. 1A).

The DIS melodies could be rather easy (easyDIS) or hard (hardDIS) to filter. An easyDIS melody was composed of sounds in a frequency range 6–7 semi-tones higher (or lower) than the corresponding S1 melody frequency range. A hardDIS melody frequency range was the same as the corresponding S1 melody frequency range. DIS melodies were constructed with the same rules as for the S1 melodies, as described above. Presentation of easy- and hard-to-ignore distractors was randomized within a given block therefore participants were not aware in advance of their difficulty.

### *Memory task*

For each trial (see Fig. 1A), the participant was informed of the ear of interest by an arrow (all visual contents are seven-centimeter-wide black elements displayed individually at the center of a white screen). 800 ms after the arrow presentation onset, S1 and DIS were played in the ear of interest and in the other ear, respectively. After a 2000-ms pause, the melody (S2) to compare to S1 was played in the ear of interest. At the end of S2 presentation, the indicative arrow was replaced by a central fixation cross. The participant had then 2000 ms to answer the question "Is S2 identical or different from S1?" using two response buttons (index and middle fingers of the right hand for identical and different respectively). When S2 differed from S1, only one sound out of four differed, and it could be either the second, third, or fourth one, never the first one which would be too salient. The answer period started at the end of the S2 melody. At the end of the 2-s-long answering period, there was a pause lasting randomly from 850 to 1250 ms before the start of the next trial.

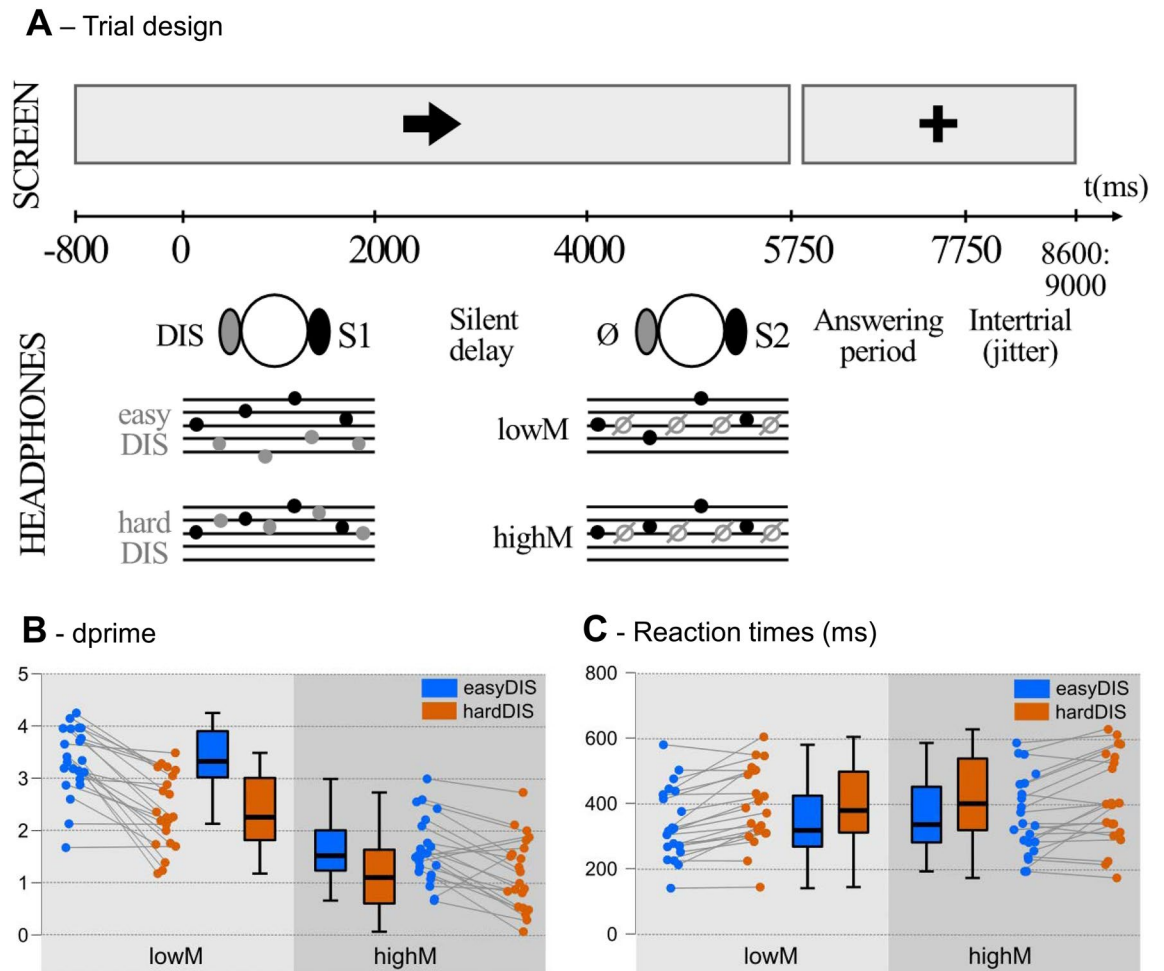
In the low memory task (lowM), when S2 differed from S1, one sound was replaced by another which is 5 to 6 semi-tones apart (within the C-major scale, ascending or descending) and induced a change in the melody contour. In the high memory task (highM), when S2 differed from S1, the changed sound differed from the original one only by 1 or 2 semi-tones (within the C-major scale) and did not induce a change in the melody contour. The difficulty of the memory task (MEMdiff) was fixed within a block, participants were notified of this difficulty at the beginning of each block.

### *Procedure*

First, participants filled in neuropsychological and demographical tests and questionnaires. Then, they were installed in the MEG in a seated position, in a sound-attenuated, magnetically-shielded recording room, at 50-cm distance from the screen. All stimuli were delivered using Presentation software (Neurobehavioural Systems, Albany, CA, USA). Sounds were delivered through air-conducting plastic ear tubes. Detection thresholds for each ear were measured with an adaptive tracking procedure for A3 (220 Hz, which is at the middle of the range of notes used). The auditory stimuli were subsequently presented at a comfortable listening level (60 dB SL). Then brain activity was recorded during a 5-min resting state period (eyes opened with a fixation cross provided to the participant). Participants then performed the main task. They were instructed to answer as accurately as possible with their right hand and to keep fixating the cross during the recordings. Prior to the actual task, they performed two short training blocks (12 trials) for each memory difficulty level. Training melodies (S1, DIS, and S2) were specific to the training sessions. Finally, they performed the ten experimental blocks (48 trials each), 5 under each memory difficulty (low or high). They were aware in advance of the difficulty of the memory task but not of the difficulty of the distractors, which was randomized within each block. The order of the blocks was randomized between participants. Trials lasted 9400 to 9800 ms, leading to eight-minute-long blocks.

### *Balancing—within blocks*

Each of the ten blocks contained 48 trials. All the combinations of ear of interest (Left or Right), distractor difficulty (easyDIS or hardDIS), associated answer (same or different) were equiprobable within block (6 trials/block each). For each combination with an expected answer "different", the direction of the change was



**Fig. 1.** Trial design & behavioral results. (A). The same paradigm is described in Blain, de la Chapelle et al. 77. The ear of interest was indicated by an arrow on the screen during the trial. The to-be-attended melody S1 played in this ear must be encoded in memory while a to-be-ignored melody DIS was played in the other ear, interleaved with S1 tones. After a 2s retention delay, a melody S2 was played in the ear of interest and participants indicated whether S1 and S2 were identical or different. The melody DIS could be either easy (easyDIS) or hard (hardDIS) to ignore based on their frequency similarity with S1. When S2 was different from S1, one note was changed of either 5–6 semi-tones, entailing a contour change in the melody (low difficulty memory task; lowM blocks) or 1–2 semi-tones without contour change (high difficulty memory task; highM blocks). Performance: (B).  $d'$  and (C). reaction times. Effect of the difficulty of the task (lowM: low difficulty, highM: high difficulty) and of the difficulty of the distractor (easyDIS: easy to ignore, hardDIS: hard to ignore). Whisker plots show the median  $\pm$  quartile and the minimum/maximum values. Lines in the strip plots connect data from the same participant.

equiprobably ascending or descending in frequency (3 trials/block each). This within-block balancing allowed to control for side, expected answer, and change direction.

#### Balancing—between blocks

The same S1 sequences were used in lowM and highM. For half of the participants in each group, melodies S1 and DIS were inverted and became respectively DIS and S1. Association between one S1 melody and another DIS melody changed across participants, enabling us to limit the impact in the results from any unfortunate easy or difficult match.

#### Behavioral data

##### Trial rejection

Trials in which participants gave no answer, several answers, or did answer at an inappropriate moment (*i.e.*, outside of the indicated period) were excluded from analysis. Average number  $\pm$  SD of trials excluded per participant was 4 trials  $\pm$  4 (out of 480).

### Measurements

Data were analyzed using Signal Detection Theory measures.  $d'$  (prime  $d$ ) was the difference between normalized “Hits” and normalized “False alarms”. Hits were the proportion of correct “different” answer over all “different” trials, and False alarms were the proportion of incorrect “different” answer over all “same” trials. Criterion ( $c$ ) indicated the response bias. Positive values indicated the tendency of answering “same” and negative values reflected the tendency of answering “different”. Median Reaction Times (RTs) were computed out of correctly answered trials and corresponded to the time between the end of the S2 and the button press.

### Statistical analyses for the behavioral data

All analyses were conducted using Bayesian statistics, which allow to test the similarity between measures and to estimate a degree of logical support or belief regarding specific results, as implemented in the JASP software (v0.16.3)<sup>88–91</sup>. We reported Bayes Factor ( $BF_{10}$ ) as a relative measure of evidence (compared to the null model). To interpret the strength of evidence against the null model, we considered a BF between 1 and 3 as weak evidence, a BF between 3 and 10 as positive evidence, a BF between 10 and 100 as strong evidence, and a BF higher than 100 as a decisive evidence<sup>92</sup>. Similarly, to interpret the strength of evidence in favor of the null model, we considered a BF between 0.33 and 1 as weak evidence, a BF between 0.01 and 0.33 as positive evidence, a BF between 0.001 and 0.01 as strong evidence and a BF lower than 0.001 as decisive evidence. For clarity, we reported information about the best model only, unless several models account almost as well for the data (as assessed when comparing models to the best model). Additionally, we reported the results of an analysis of effects that compares models that contain a given effect (of a factor or an interaction) to equivalent models stripped of the effect and was considered as a relative measure of evidence supporting the inclusion of a factor ( $BF_{inclusion}$ ).

$d'$ , criterion ( $c$ ), and RTs were submitted to a Bayesian repeated-measure ANOVA with MEMdiff (difficulty of the task, two levels: lowM, highM), DISdiff (difficulty of the DIStractor, two levels: easyDIS, hardDIS) as within-participant factors. Post-hoc comparisons for main effects or interactions were conducted using Bayesian  $t$ -tests. Criterion ( $c$ ) for the four conditions (2 levels of MEMdiff  $\times$  2 levels of DISdiff) were submitted to a Bayesian one-sample  $t$ -test comparison to zero.

### Brain recordings

MEG data were continuously recorded using a 275-channel whole-head axial gradiometer system (CTF-275 by VSM MedTech Inc., Vancouver, Canada) with a sampling rate of 600 Hz and a 0.016–150 Hz bandpass filter. Head movements were continuously monitored using 3 coils placed at the nasion and the two preauricular points.

For each participant, a T1-weighted 3D MRI was obtained using a 1.5 T or 3 T whole-body scanner (Magnetom Siemens Erlangen, Germany) after the MEG session. At the end of MEG session, locations of the nasion and the two preauricular points were marked using fiducial markers visible on the T1 acquisitions. The T1 images were used for reconstruction of individual head shapes to create forward models for the source reconstruction procedures.

### Data pre-processing

MEG data were pre-processed offline using the software package for electrophysiological analysis (ELAN Pack) developed at the Lyon Neuroscience Research Center<sup>93</sup>. A third-order spatial gradient noise cancellation was applied to the data. The recorded signal was rejected when head movements exceeded the threshold of  $\pm 1$  cm around the median position. Sensor jumps ( $> 10$  picoTesla) were detected and corrected with custom software. Blinks, saccades, and heartbeats were removed from the signal thanks to an Independent Component Analysis (ICA), using eeglab MATLAB toolbox<sup>94</sup>. This ICA was computed on band-pass filtered signal (0.5–45 Hz). Eye-movements and heartbeat-related components were determined by visual inspection of component topographies and time courses and removed through an ICA inverse transformation applied to the non-filtered signal. The ICA-corrected signals were then band-stop-filtered between 47 and 53 Hz, 97 and 103 Hz, and 147 and 153 Hz (zero-phase shift Butterworth filter, order 3) to remove power-line artifacts. Data segments contaminated with any residual artifact were excluded using a threshold of 2.5 picoTesla. 92% ( $\pm 11$ ) of trials remained in the analyses after artifact rejection (resulting in 55 ( $\pm 6$ ) trials per presentation side of S1 and condition).

### ERF analyses

#### ERF at sensor-level

For each participant and condition, event-related fields (ERFs) were obtained by averaging band-stopped ICA-corrected MEG data locked to the onset of S1 (time-window:  $-1500$  to  $7000$  ms, averaging epochs thus start 250 ms prior to cue onset, see Figure S1). To study the Contingent Negative Variation (CNV) after the visual cue, in anticipation of the melody to encode, ERFs were filtered with a 30-Hz low-pass filter and baseline corrected to the mean amplitude of the 250-ms period before cue onset ( $-1050$  to  $-800$  ms before S1 onset). To study transient evoked responses to tones of S1/DIS melodies and sustained responses during S1/DIS melodies or the retention delay, ERFs were filtered with a 2–30-Hz band-pass filter and with a 2-Hz low-pass filter, respectively. ERFs were then baseline corrected to the mean amplitude of the  $-250$  to  $0$  ms period before S1 onset (see Figure S1 and Table 1).

Because of saliency effects for the first note, we further separated the analysis of the first note and the following ones: on the one hand, ERFs were computed locked to the first note onset, and on the other hand, we computed the average of the ERFs locked to the second, third, and fourth notes of the melodies, with no further baseline correction, separately for S1 and DIS in both cases. We also computed the difference between S1 and DIS (average across notes 2, 3, 4).

Event of interest	Filtering	Studied ERF	STATISTICS			
			ROI [number of vertices]	Time-window for analysis (ms)	Factors tested	Main result
S1	lp 30Hz	CNV	Left PP [32]	-400 to -200	Cued side & MEMdiff	-
				-200 to 0		-
			Right PP [40]	-400 to -200		-
				-200 to 0		HighM > lowM after left cue
S1: note1	bp 2-30Hz	N1	Left temporal [55]	90 to 140	MEMdiff	-
			Right temporal [66]			-
			Left frontal [54]			-
			Right frontal [60]			-
DIS: note1	bp 2-30Hz	N1	Left temporal [55]	100 to 150	MEMdiff & DISdiff	HardDIS > easyDIS
			Right temporal [66]			-
			Left frontal [54]			-
			Right frontal [60]			-
S1 & DIS: notes 2,3,4	bp 2-30Hz	N1	Left temporal [68]	100 to 150	Stimulus Type, MEMdiff & DISdiff	S1 > DIS; lowM & S1 only: easyDIS > hardDIS
			Right temporal [65]			S1 > DIS; S1 only: easyDIS > hardDIS
			Left frontal [27]			S1 > DIS
			Right frontal [24]			S1 > DIS
S1 & DIS: notes 2,3,4	bp 2-30Hz	P2	Left PP [46]	200 to 250	Stimulus Type, MEMdiff & DISdiff	DIS only: hardDIS > easyDIS
			Right PP [44]			S1 > DIS; DIS only: hardDIS > easyDIS
S1—DIS: notes 2,3,4	bp 2-30Hz	Late differential response	Right temporal [78]	225 to 275	Stimulus_Side, MEMdiff, & DISdiff	EasyDIS > hardDIS
S1	lp 2Hz	Sustained response	Left PP [51]	500 to 2000	S1 side, DISdiff, & MEMdiff	HardDIS > easyDIS
			Right PP [44]			HardDIS > easyDIS
			Left PT [37]			-
			Right PT [11]			-
			Left PP [51]	2500 to 4000		-
			Left PT [37]			-

**Table 1.** Summary of the ERF analyses: event and filtering used, studied ERF, ROI information, time-window used for statistical analysis, factors of the Bayesian repeated-measure ANOVA used, and results of the statistical analysis. For CNV and sustained responses, ROIs were defined based on activation from all stimuli between -400 and 0 ms and between 500 and 2000 ms respectively, whereas for N1, P2, and late responses, ROIs were defined based on activation from contralateral stimuli only (see text for details). When computing the difference between S1 and DIS, both S1 and DIS stimuli were contralateral (i.e. they are taken from different trials). For N1 and P2 for notes 2, 3, 4, as activity is largely overlapping for S1 and DIS, the ROIs were drawn from the emergence test of S1. We report in the table only effects for which at least positive evidence (i.e., BF10 or BFInclusion > 3) was observed (see text for the full results). “>” means larger ERF (in absolute value) in a given condition compared to another one. “-” means no positive or greater evidence for an effect. S1: to-be-attended melody, DIS: to-be-ignored melody; lp: low pass; bp: band-pass; PP Planum Polare, PT Planum Temporale, MEMdiff memory task difficulty (lowM vs. highM), DISdiff distractor filtering difficulty (easyDIS vs. hardDIS).

Grand-average ERF was obtained by averaging the same computed ERF across all 22 participants.

#### *ERF source reconstruction*

Segmentation of the individual T1-weighted MRIs was conducted using the FreeSurfer software package (<http://surfer.nmr.mgh.harvard.edu>)<sup>95</sup>. Segmentation was visually inspected and then imported (15,002 vertices) in the Brainstorm toolbox (version 3.191209—09-Dec-2019—on Matlab R2015a; <http://neuroimage.usc.edu/brainstorm>)<sup>96</sup> with which further source analyses were conducted. The white matter–grey matter boundary segmented by FreeSurfer was used as a source space for subsequent weighted-minimum-norm estimation. A noise covariance matrix was computed on resting state data, which had been preprocessed with the same pipeline as data collected during task performance. After coregistration between the individual anatomy and MEG sensors, the forward model was computed with the OpenMEEG software (<https://openmegg.github.io/>)<sup>97</sup>. Cortical currents were estimated using a distributed model consisting of 15,002 currents dipoles from the time series of the 275 gradiometer signals, using a linear inverse estimator (weighted minimum-norm current estimate, signal-to-noise ratio of 3, whitening PCA, depth weighting of 0.5). Sources orientations were constrained to be orthogonal to the grey-white matter boundary of the individual MRIs. The results were then projected on a standard brain model anatomy (ICBM152; 15,002 vertices) for group analysis purposes.

#### *ERF source analysis*

ERF analyses were performed following five steps: (1) the main ERFs were identified at the sensor level, and time-windows of interest were defined for each ERF (see Table 1 for a list of ERFs studied and the associated time-windows); (2) at the source level, emergence of ERFs (all four conditions collapsed) was tested at the whole brain level using t-tests comparing signal amplitude (on average across the time-window of interest) to zero at each vertex (FDR-correction for multiple comparisons, corrected  $p < 0.05$ ); (3) regions of interest (ROIs) were drawn using Brainstorm based on these emergence tests with a procedure allowing to grow from a seed voxel to match the shape of the activation patch on the cortical surface; (4) ROI time-courses of activity were extracted for each participant, each condition, each presentation side; (5) mean amplitudes in time-windows of interest were computed for each participant, each condition, each presentation side; (6) Bayesian ANOVA (see Results and Table 1 for details about factors, and the Statistical methods for the behavioral data section above for details about Bayesian ANOVA) for each studied ROI and each time-window were performed (see Table 1 for ROIs and time-windows). Emergence of ERFs (step 2) was scrutinized for stimuli presented at left or right ear separately, and if similar patterns were observed, data were further collapsed across stimulus presentation side for ROI definition (see Table 1). Emergence of ERFs was also assessed in the difference between S1 and DIS (notes 2, 3, 4). In the context of neuroimaging studies, Bayesian analysis offer the advantage to alleviate the need for correction for multiple comparisons<sup>98,99</sup>.

### **Time–frequency analysis**

#### *TF at sensor-level*

To investigate the dynamics of theta and alpha power during the task, a time–frequency analysis was performed using the Fieldtrip toolbox (v20190626)<sup>100</sup> on Matlab R2017A. Time–frequency analysis was computed from –2500 to +8000 ms relative to S1 onset, using Morlet wavelet decomposition with a width of four cycles per wavelet ( $m = 7$ )<sup>101</sup>, for frequencies from 4 to 20 Hz, with steps of 1 Hz and 50 ms. For each single trial, the corresponding ERF was subtracted before wavelet decomposition to focus on induced (and not evoked) activity. The relative change in power was then obtained by subtracting then dividing each power value with the power in a pre-cue baseline period (–1800 to –1000 ms relative to S1 onset) in the corresponding frequency range. The main power modulations were identified at the sensor level for all conditions collapsed and used to define time–frequency windows of interest for source-level analyses.

#### *TF: head model*

For source reconstruction, the processing of the T1 MRI was performed using CTF software (CTF Systems Inc., VSM Medtech Inc., Vancouver, Canada). For each participant, an anatomically realistic single shell headmodel based on the cortical surface was generated from individual head shapes<sup>102</sup> and imported in the Fieldtrip toolbox (v20190626)<sup>100</sup>. Then, a grid with 0.5-cm resolution was normalized on a MNI template and morphed into the brain volume of each participant.

#### *TF: whole brain source analyses*

To elucidate the possible brain regions underlying the sensor-level modulations in the theta and alpha bands, we performed whole brain analysis on time–frequency windows of interest.

Several time-windows relative to S1 onset were chosen based on the protocol design: one during the cue presentation (–800:0 ms); two during the presentation of S1 and DIS melodies (400:1200 ms & 1200:2000 ms); and two during the delay (2400:3200 ms & 3200:4000 ms). Three different frequency bands (4 to 7 Hz, 7 to 11 Hz & 11 to 15 Hz) were chosen based on the observations at the sensor level and previous findings<sup>45,46</sup>.

We used the frequency-domain adaptive spatial technique of Dynamical Imaging of Coherent Sources (DICS)<sup>103</sup>. For each participant, data from all conditions was concatenated, and the cross-spectral density (CSD) matrix (–2.5 to +7 s, relative to S1 onset, lambda 5%) was calculated using the multitaper method with a target frequency of 9.5 ( $\pm 5.5$ ) Hz. Leadfields for all grid points along with the CSD matrix were used to compute a common spatial filter that was in turn used to estimate the spatial distribution of power for all time–frequency windows of interest.

Afterward, for left and right S1 presentation separately (irrespective of MEMdiff and DISdiff), power in each time–frequency window of interest was contrasted against the power in a corresponding baseline pre-cue

window (−1800 to −1000 ms relative to S1 onset) for frequencies around 5.5 (+/−1.5), 9 (+/−2) and 13 (+/−2) Hz using a nonparametric cluster-based permutation analysis<sup>104</sup> across participants, controlling for multiple comparisons in the source space dimension.

#### Reconstruction of source-level single-trial activity in ROIs

In order to get a time–frequency resolved estimation of source activity, we used an approach based on regions of interest (ROIs) and we reconstructed single trial activity at the level of ROIs using the linearly constrained minimum variance (LCMV) beamformer<sup>105</sup> in each subject.

Regions of interest (ROIs), in each hemisphere, were defined from the Brodmann atlas based on the group whole brain activation (see previous paragraph): the *auditory ROI (Aud)* was composed of Brodmann areas 22, 41, and 42, the *visual ROI (Vis)* was composed of Brodmann areas 17, 18, and 19, the *anterior prefrontal cortex ROI* was composed of Brodmann areas 9 and 10, the *Broca ROI* was composed of Brodmann areas 44 and 45, and the *dorso-lateral prefrontal (DLPFC) ROI* was composed of Brodmann area 46.

Spatial filters were constructed from the covariance matrix of the averaged single trials at sensor level (−2.5:7 s, relative to S1 onset, 1–20 Hz, lambda 5%) and the respective leadfield for all grid points. Then, spatial filters were multiplied by the sensor-level single-trial data, thus obtaining the time course activity of each voxel of interest within the defined ROIs. Single-trial activity was averaged across all voxels within each ROI in each hemisphere. For each ROI, the evoked response (i.e., the signal averaged across all trials) was subtracted from each trial.

#### ROI peak frequency

To extract the frequency peak in each ROI (source-level activity reconstructed with LCMV, see previous paragraph), for each time-window of interest (−1800:−1000; −800:0; 400:1200; 1200:2000; 2400:3200; 3200:4000 ms), we computed the power spectral density using Fourier decomposition based on Hann windows and extracted the rhythmic activity from the concurrent 1/f modulations of the signal using the Fitting oscillations and one-over-F (FOOOF) toolbox implemented in Fieldtrip (v20221004)<sup>100,106</sup>, for all conditions collapsed. The resulting FFT of the baseline was subtracted from the resulting FFT of each following time-window.

The theta peak corresponded to the highest local maximum in the resulting spectrum between 3 and 7 Hz; while the peaks for alpha decrease and increase corresponded to the lowest local minimum and highest local maximum between 7 and 15 Hz, respectively.

In each ROI, the group frequency range was then defined from the median peak frequency across subjects +/−1.5 Hz, averaged over time-windows, resulting in 4–7 Hz, 8–11 Hz and 9–12 Hz to be used in subsequent analyses for theta, low alpha and high alpha respectively.

#### ROI statistical analysis

In each ROI, the time–frequency power was calculated for each trial using Morlet wavelet decomposition with a width of four cycles per wavelet ( $m = 7$ ) at frequencies between 4 and 15 Hz, in steps of 1 Hz and 50 ms. The relative change in power was then obtained by averaging single trial TF power in each condition (for each level of memory difficulty, distractor filtering difficulty, and side of S1 presentation) and by subtracting then dividing each averaged power value with the power in a pre-cue baseline period (−1800 to −1000 ms relative to S1 onset) in the corresponding frequency range.

For each frequency range (4–7 Hz, 9–12 Hz and 8–11 Hz) and each time-window of interest (−800:0; 400:1200; 1200–200; 2400:3200; 3200:4000 ms), the relative change in power was extracted in each subject and for each ROI of interest based on DICS analysis. To investigate the impact of the memory task and distractor difficulties, these averaged values were submitted to a Bayesian repeated-measure ANOVA with factors MEMdiff, DISdiff, and S1side, for each ROI and each corresponding time-fr.

## Results

### Behavioral results

For  $d'$  (Fig. 1B, see also Figure S2 for percentages of correct responses), the best model explaining the data was the one with factors MEMdiff, DISdiff, and the interaction between MEMdiff and DISdiff ( $BF_{10} = 2.3e + 16$ ), in keeping with Blain, Talamini et al.<sup>76</sup> and Blain, de la Chapelle et al.<sup>77</sup>. The analysis of effects across models revealed decisive evidence for MEMdiff ( $BF_{inclusion} = 3.7e + 10$ ), for DISdiff ( $BF_{inclusion} = 6729.8$ ), and for the MEMdiff x DISdiff interaction ( $BF_{inclusion} = 116.0$ ).  $d'$  was higher for easyDIS compared to hardDIS, and higher in lowM compared to highM. The difference between hardDIS and easyDIS was smaller in highM compared to lowM (or put differently, the difference between lowM and highM was smaller for hardDIS than easyDIS), as revealed by a post-hoc Bayesian t-test ( $BF_{10} = 123.2$ ).

For the criterion (c), in agreement with previous studies using delayed-matching-to-sample tests, there was weak to decisive evidence for positive c values in all conditions, underlining that participants tended to miss differences between melodies (see Table S5 in<sup>77</sup>). The best model explaining criterion results was the one with factors MEMdiff, DISdiff, and the interaction between MEMdiff and DISdiff ( $BF_{10} = 4.4e + 9$ ). There was decisive evidence for MEMdiff ( $BF_{inclusion} = 1.4e + 7$ ), strong evidence for DISdiff ( $BF_{inclusion} = 73.7$ ), and positive evidence for the MEMdiff x DISdiff interaction ( $BF_{inclusion} = 7.0$ ). Criterion was higher under highM compared to lowM and is higher for easyDIS compared to hardDIS. Post-hoc Bayesian paired-samples *t*-tests revealed strong evidence for a difference in criterion between easyDIS and hardDIS for lowM ( $BF_{10} = 90.7$ ) but only weak evidence for such a difference in highM ( $BF_{10} = 1.5$ ).

For RTs (see Fig. 1C), the best model explaining results was the one with factors MEMdiff and DISdiff ( $BF_{10} = 671.5$ ). The analysis of effects revealed decisive evidence for DISdiff ( $BF_{inclusion} = 151.2$ ) and positive

evidence for MEMdiff ( $BF_{\text{inclusion}} = 4.6$ ). RTs were longer for hardDIS compared to easyDIS, and they are longer for highM compared to lowM.

### Event-related fields results

To ease the reading of the ERF results, the key findings are highlighted at the end of each section, and a summary of the main results obtained for ERF analyses is provided in Table 1.

#### Anticipation of the sequence to encode: contingent negative variation (CNV)

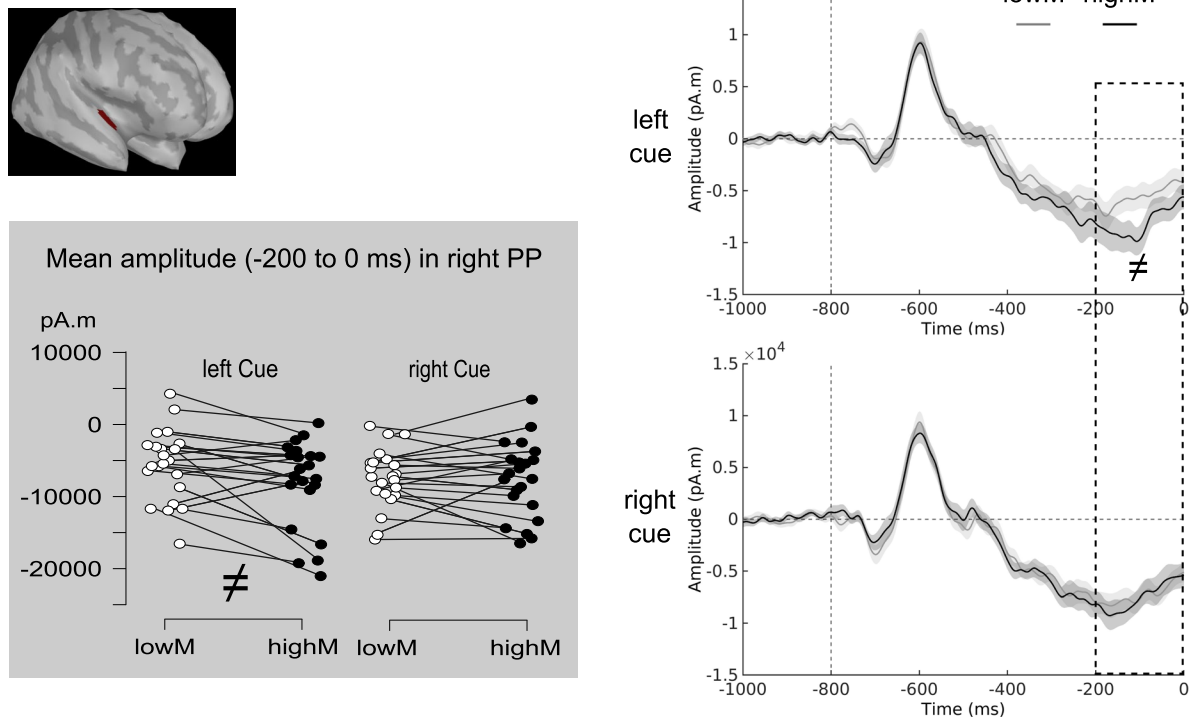
The time course and topographies at the sensor level of the CNV developing after the visual cue are shown in Figure S3. Emergence tests at the source level (time window:  $-400$  to  $0$  ms relative to S1 onset) revealed activity in temporal regions before S1 onset, extending bilaterally along the Planum Polare, irrespective of the cued ear. Activity in these temporal regions was examined with Bayesian repeated-measure ANOVAs with Cued Side and MEMdiff as factors, during early ( $-400$  to  $-200$  ms relative to S1 onset) and late ( $-200$  to  $0$  ms relative to S1 onset) CNV (see Table 1).

In the right Planum Polare ROI (see Fig. 2), the best model explaining the data just before S1 presentation (time-window  $-200$  to  $0$  ms relative to stimulus onset) was the model with the MEMdiff factor ( $BF_{10} = 1.2$ ), immediately followed by the full model including the interaction between Cued Side and MEMdiff ( $BF_{10} = 1.1$ ). Note that these models were thus almost as likely as the null model. The analysis of effects revealed weak evidence for the interaction between MEMdiff and Cued Side ( $BF_{\text{inclusion}} = 2.6$ ) and for MEMdiff ( $BF_{\text{inclusion}} = 1.2$ ). Post-hoc Bayesian paired-sample t-tests revealed positive evidence for a difference between lowM and highM with contralateral stimuli ( $BF_{10} = 3.7$ ) and for no difference between lowM and highM with ipsilateral stimuli ( $BF_{10} = 0.23$ ). When the left ear was cued, larger CNV amplitudes were observed in the right Planum Polare for the highM condition compared to lowM.

The best model explaining CNV amplitudes in the left and right Planum Polare ROIs between  $-400$  and  $-200$  ms prior to S1 onset, as well as in the left Planum Polare ROI between  $-200$  and  $0$  ms prior to S1 onset (see Figure S4), is the null model in all three cases ( $BF_{10}$  for all other models  $< 0.56$ ).

**Key CNV result** the CNV in anticipation of sound sequences ( $-200$  to  $0$  ms prior to sequence onset) presented in the left ear was larger in the high memory load condition (highM), compared to the low memory load condition (lowM), in the right (i.e., contralateral) Planum Polare.

### visual cue | CNV | right Planum Polare



**Fig. 2.** Contingent Negative Variation (CNV) in the source space. Time course of the ERF to the visual cue onset, for each MEMdiff and Cued Side condition, in the right Planum Polare (PP) ROI represented on the MNI template. On the time course plots, shadowed areas represent standard errors of the mean. The strip plot shows individual data averaged by condition in the dotted period ( $-200$  to  $0$  ms) showing positive or greater evidence for an effect of the MEMdiff factor (see Table 1 for time-windows and results). Lines connect data from the same participant.

### Processing of the sound sequences: Transient Evoked Responses (N1) to first tones of S1 and DIS

The time course of the transient evoked responses after each note in S1 and DIS melodies as recorded at the sensor level is shown in Figure S5. ERFs to the first notes of S1 and DIS were larger than ERFs to subsequent notes and were thus analyzed separately (see Figs. 3, S5 & S6). The first note of S1 and DIS evoked a large N1 response in contralateral temporal (Heschl Gyrus, HG, and around) and frontal (Inferior Frontal Gyrus, IFG) regions. Smaller responses with longer latencies were observed ipsilaterally. Therefore, the amplitude of the N1 was analyzed separately for each presentation side in the contra-lateral ROIs with Bayesian paired-value t-tests for the first note of S1 (S1-L and S1-R separately), to assess the effect of the MEMdiff factor, and with Bayesian repeated-measure ANOVAs with MEMdiff and DISdiff as factors for the first note of DIS (DIS-L and DIS-R separately).

For the N1 elicited in response to the first presented tone (S1, time window 90–140 ms post S1 onset), no evidence was found for an effect of MEMdiff (all  $BF_{10} \leq 0.65$ ) in the four ROIs investigated (Table 1 and Figure S6).

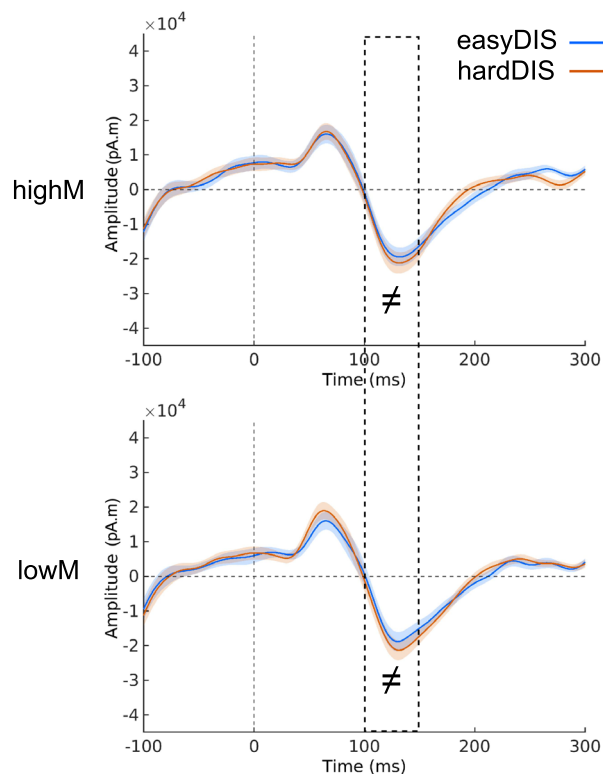
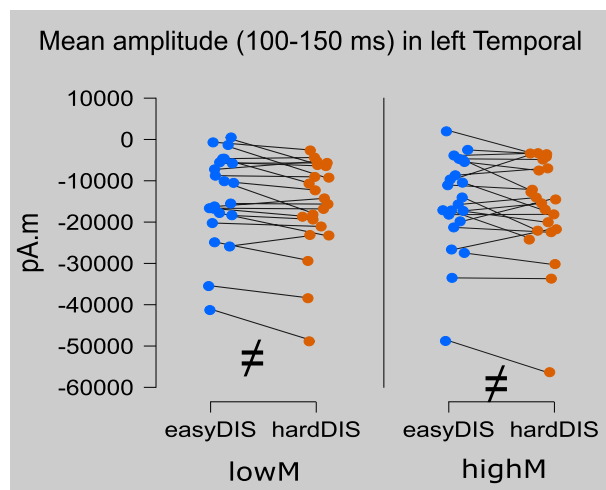
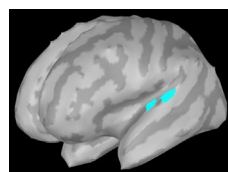
For the N1 elicited in response to the first to-be-ignored tone (DIS, time window 100–150 ms post DIS onset), the best model explaining the data in the left Temporal ROI (i.e., for DIS stimuli presented in the right ear) was the model with the DISdiff factor ( $BF_{10} = 3.6$ ), followed by the model with DISdiff and MEMdiff ( $BF_{10} = 2.2$ ), with the analysis of effects confirming only positive evidence for DISdiff ( $BF_{inclusion} = 3.7$ ). A larger N1 was observed in response to the first DIS tone in hardDIS compared to easyDIS (Fig. 3). In the left IFG ROI, the best model explaining the data (N1 to the first DIS tone) is the full model including the interaction between MEMdiff and DISdiff ( $BF_{10} = 1.9$ ), however it is almost as likely as the null model and further Bayesian paired-value t-tests comparing N1 amplitudes across conditions revealed only weak evidence for differences between conditions, if any (all  $BF_{10} < 1.7$ ). In the right hemisphere ROIs (temporal and IFG), for DIS stimuli presented in the left ear, the best models explaining the N1 amplitudes were the null model (for all other models,  $BF_{10} < 0.9$ ).

**Key result for the N1 for first tones** In the left temporal ROI, the N1 for the first to-be-ignored tone (DIS) was larger for hard-to-ignore distractors (hardDIS) than easy-to-ignore distractors (easyDIS), for distractors presented in the right ear (i.e., contralaterally).

### Processing of the sound sequences: transient evoked responses (N1) to tones 2, 3, 4 (S1 and DIS)

Sensor-level ERFs for subsequent tones (S1 and DIS) are illustrated in Figure S7. Emergence tests revealed significant activity in contralateral temporal (HG and around) and frontal (IFG) regions at the latency of the N1

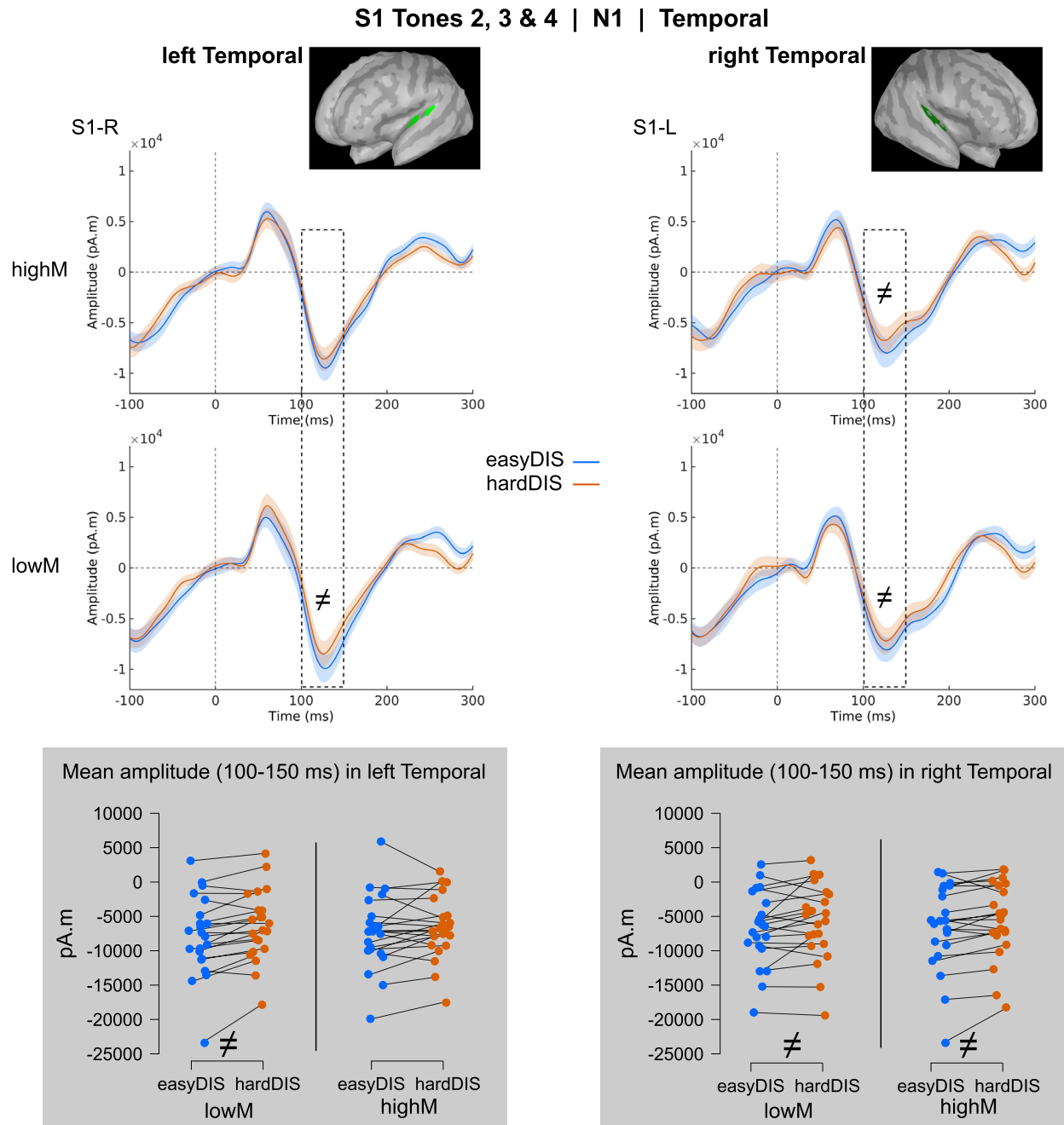
## DIS-R Tone1 | N1 | left Temporal



**Fig. 3.** Transient evoked responses to the first tone in the source space. Time course of the ERF to the first tone of DIS melodies presented to the right (DIS-R), for each DISdiff (DIS represented only) and each MEMdiff condition, in the left Temporal ROI represented on the MNI template. On the time course plots, shadowed areas represent standard errors of the mean. The strip plot shows individual data averaged by condition in the gray period showing positive or greater evidence for a DISdiff effect (see Table 1 for time-windows and results). Lines connect data from the same participant.

(100–150 ms post tone onset). Smaller responses with longer latencies were observed ipsilaterally. N1 amplitudes were analyzed in contralateral ROIs with Bayesian ANOVAs with StimulusType (S1 or DIS), MEMdiff, and DISdiff as factors (see Table 1 and Figs. 4 and S8).

In the left Temporal ROI, the best model explaining N1 amplitude was the full model, including the triple interaction between StimulusType, MEMdiff, and DISdiff ( $BF_{10} = 222,797.5$ ). The analysis of effects revealed decisive evidence for StimulusType ( $BF_{inclusion} = 17,470.0$ ), with larger N1 for S1 compared to DIS, and strong evidence for the triple interaction ( $BF_{inclusion} = 13.4$ ), which we further explored in separate analyses by Memory difficulty (for all other effects,  $BF_{inclusion} < 2.4$ ). Under low Memory difficulty, in a Bayesian ANOVA with



**Fig. 4.** Transient evoked responses to S1 tones 2, 3, and 4 in the Temporal ROIs (source space). Time course of the averaged ERF to the tones 2, 3, and 4 of the to-be-attended melody S1 (tone onset at 0ms), for each DISdiff and MEMdiff condition, in the corresponding ROI represented on the MNI template. On the time course plots, shadowed areas represent standard errors of the mean. The strip plots show individual data averaged by condition in the dotted periods showing positive or greater evidence for a DISdiff effect or for the DISdiff-by-MEMdiff interaction (see Table 1 for time-windows and results). Lines connect data from the same participant. Only contralateral tones are analyzed here: S1-R for left sources, and S1-L for right sources. S1-L or S1-R: S1 melodies presented to the left or right ear, respectively.

StimulusType and DISdiff as factors, the best model explaining N1 amplitude was the full model including the StimulusType by DISdiff interaction ( $BF_{10} = 118,659.4$ ). The analysis of effects revealed decisive evidence for StimulusType ( $BF_{inclusion} = 1692.7$ ) and for the StimulusType by DISdiff interaction ( $BF_{inclusion} = 108.4$ ), and weak evidence against DISdiff ( $BF_{inclusion} = 0.65$ ). Bayesian t-tests exploring this interaction revealed strong evidence for an effect of DISdiff on N1 amplitude for S1 ( $BF_{10} = 51.3$ , larger N1 amplitudes for the easyDIS condition), but weak evidence for an absence of effect of DISdiff on N1 amplitude for DIS ( $BF_{10} = 0.36$ ). *Under high Memory difficulty*, in a Bayesian ANOVA with StimulusType and DISdiff as factors, the best model explaining N1 amplitude was the model with only the StimulusType effect ( $BF_{10} = 104,513.0$ ), followed by the model with StimulusType and DISdiff effects ( $BF_{10} = 56,277.9$ ). The analysis of effects revealed decisive evidence for StimulusType ( $BF_{inclusion} = 103,549.6$ ) but weak evidence against DISdiff ( $BF_{inclusion} = 0.54$ ).

In the right Temporal ROI, the best model explaining N1 amplitude was the model with StimulusType, DISdiff, and the interaction between StimulusType and DISdiff ( $BF_{10} = 4173.9$ ), immediately followed by the model with the same effects and interaction and the effect of MEMdiff in addition ( $BF_{10} = 2399.1$ ). The analysis of effects revealed decisive evidence for the effect of StimulusType ( $BF_{inclusion} = 1520.4$ ), positive evidence for the interaction between StimulusType and DISdiff ( $BF_{inclusion} = 6.0$ ), and weak evidence for an absence of an effect of MEMdiff ( $BF_{inclusion} = 0.69$ ) and of DISdiff ( $BF_{inclusion} = 0.50$ ). Further post-hoc Bayesian t-test to explore the StimulusType-by-DISdiff interaction revealed that, for S1, there was positive evidence for larger N1 amplitude for easyDIS than for hardDIS ( $BF_{10} = 4.2$ ), whereas for DIS, there was weak evidence for no difference in N1 amplitude between easyDIS and hardDIS ( $BF_{10} = 0.63$ ).

In the left Frontal ROI, the best model explaining N1 amplitude was the model with only the effect of StimulusType ( $BF_{10} = 5266.8$ ). Other models had similar  $BF_{10}$ , but the analysis of effects revealed only decisive evidence for StimulusType ( $BF_{inclusion} = 3088.3$ ; for all other effects and interactions,  $BF_{inclusion} < 2.1$ ).

In the right Frontal ROI, the best model explaining N1 amplitude was the model with StimulusType ( $BF_{10} = 22.3$ ), followed by the model with StimulusType, DISdiff, and the interaction between StimulusType and DISdiff ( $BF_{10} = 17.1$ ), the model with StimulusType and DISdiff ( $BF_{10} = 11.9$ ), and the model with StimulusType, MEMdiff and the interaction between StimulusType and MEMdiff ( $BF_{10} = 8.4$ ). The analysis of effects confirmed strong evidence for the effect of StimulusType ( $BF_{inclusion} = 22.5$ ), and only weak evidence for the interactions between StimulusType and DISdiff ( $BF_{inclusion} = 1.4$ ) and between StimulusType and MEMdiff ( $BF_{inclusion} = 1.6$ ); there was weak to positive evidence against all other effects ( $BF_{inclusion} < 0.55$ ).

*Key results for the N1 for tones 2, 3, 4* In the four ROIs, the N1 was larger for to-be-attended sounds (S1) than for to-be-ignored sounds (DIS), revealing efficient filtering. In the left temporal ROI, the N1 to S1 was larger in the presence of easy-to-ignore distractors (easyDIS) than in the presence of hard-to-ignore distractors (hardDIS), under low memory load (lowM) only. In the right temporal ROI, the N1 to S1 is larger in easyDIS than in hardDIS, irrespective of the memory task difficulty.

#### *Processing of the sound sequences: transient evoked responses (P2) to tones 2, 3, 4 (S1 and DIS)*

Activity was observed in the contralateral Planum Polare in the P2 latency range (200–250 ms) for tone 2,3,4 (S1 and DIS). As for the N1, P2 amplitudes were analyzed with Bayesian repeated-measure ANOVAs with StimulusType, MEMdiff, and DISdiff as factors (see Table 1 and Figs. 5 & S9).

In the left Planum Polare, the best model explaining P2 amplitude was the model with StimulusType, DISdiff, and the interaction between StimulusType and DISdiff ( $BF_{10} = 32.1$ ). The analysis of effects confirmed positive evidence for the interaction between StimulusType and DISdiff ( $BF_{inclusion} = 5.6$ ) and for the effect of DISdiff ( $BF_{inclusion} = 5.0$ ), and weak evidence for the effect of StimulusType ( $BF_{inclusion} = 1.1$ ). Further post-hoc Bayesian t-test to explore the StimulusType-by-DISdiff interaction revealed that, for S1, there was positive evidence for no difference in P2 amplitude between easyDIS and hardDIS ( $BF_{10} = 0.23$ ), whereas for DIS, there was strong evidence for a larger P2 amplitude in hardDIS compared to easyDIS ( $BF_{10} = 22.8$ ).

In the right Planum Polare, the best model explaining P2 amplitude was the model with StimulusType, DISdiff, and the interaction between StimulusType and DISdiff ( $BF_{10} = 855.6$ ), followed by the model with StimulusType and DISdiff ( $BF_{10} = 732.2$ ). The analysis of effects confirmed an effect of StimulusType with decisive evidence ( $BF_{inclusion} = 123.0$ ; P2 was larger for S1 than DIS), of DISdiff with positive evidence ( $BF_{inclusion} = 6.9$ ), and of the interaction between StimulusType and DISdiff with weak evidence ( $BF_{inclusion} = 1.2$ ). Further post-hoc Bayesian t-test to explore the StimulusType-by-DISdiff interaction revealed that, for S1, there was positive evidence for no difference in P2 amplitude between easyDIS and hardDIS ( $BF_{10} = 0.25$ ), whereas for DIS, there was strong evidence for a larger P2 amplitude in hardDIS compared to easyDIS ( $BF_{10} = 69.9$ ).

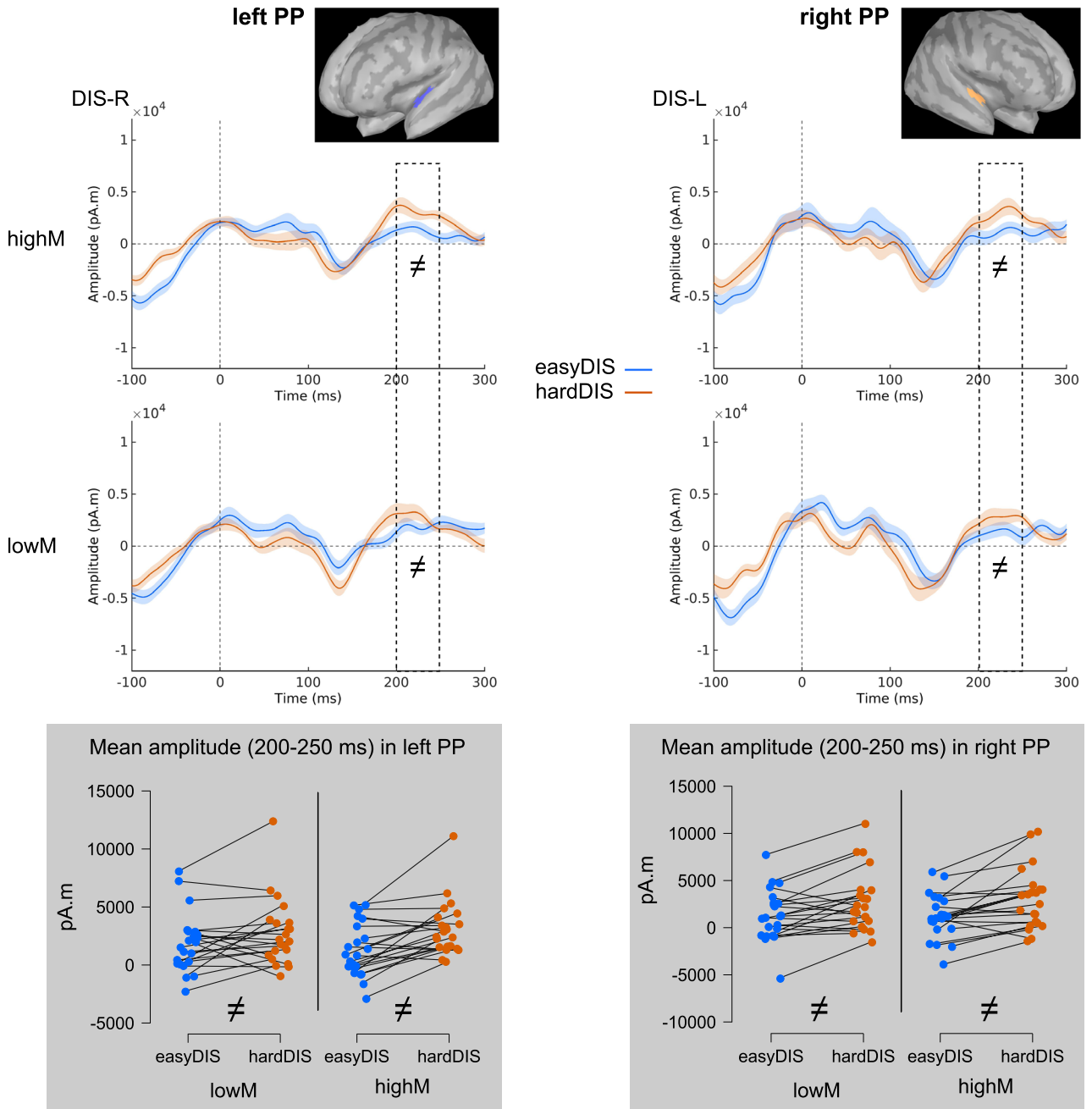
*Key results for the P2 for tones 2, 3, 4* in the two ROIs (left and right Planum Polare), P2 was larger for to-be-attended sounds (S1) than for to-be-ignored sounds (DIS), revealing efficient filtering. Also in the two ROIs, for to-be-ignored sounds (DIS), P2 was larger for hard-to-ignore distractors (hardDIS) than for easy-to-ignore distractors (easyDIS).

#### *Processing of the sound sequences: transient evoked responses, late differential response to tones 2, 3, 4 (S1–DIS)*

The ERF evoked by S1 and DIS not only differs during N1 and P2, as described above, but they also diverge later, between 225 and 275 ms, in right temporal areas around Heschl's gyrus, irrespective of stimulation side. This late differential response (S1–DIS amplitude) is examined with Bayesian repeated-measure ANOVAs with Stimulus\_Side, DISdiff, and MEMdiff as factors (see Table 1 and Fig. 6).

In the right Temporal ROI, the best model explaining the late differential response (S1–DIS amplitude) is the model with the factors Stimulus\_Side, DISdiff, and the interaction between Stimulus\_Side and DISdiff ( $BF_{10} = 91.2$ ), followed by the model with only the effect of DISdiff ( $BF_{10} = 75.5$ ) and the model with the effects of Stimulus\_Side and DISdiff ( $BF_{10} = 36.4$ ). The analysis of effects confirms strong evidence for DISdiff

DIS Tones 2, 3 & 4 | P2 | Planum Polare

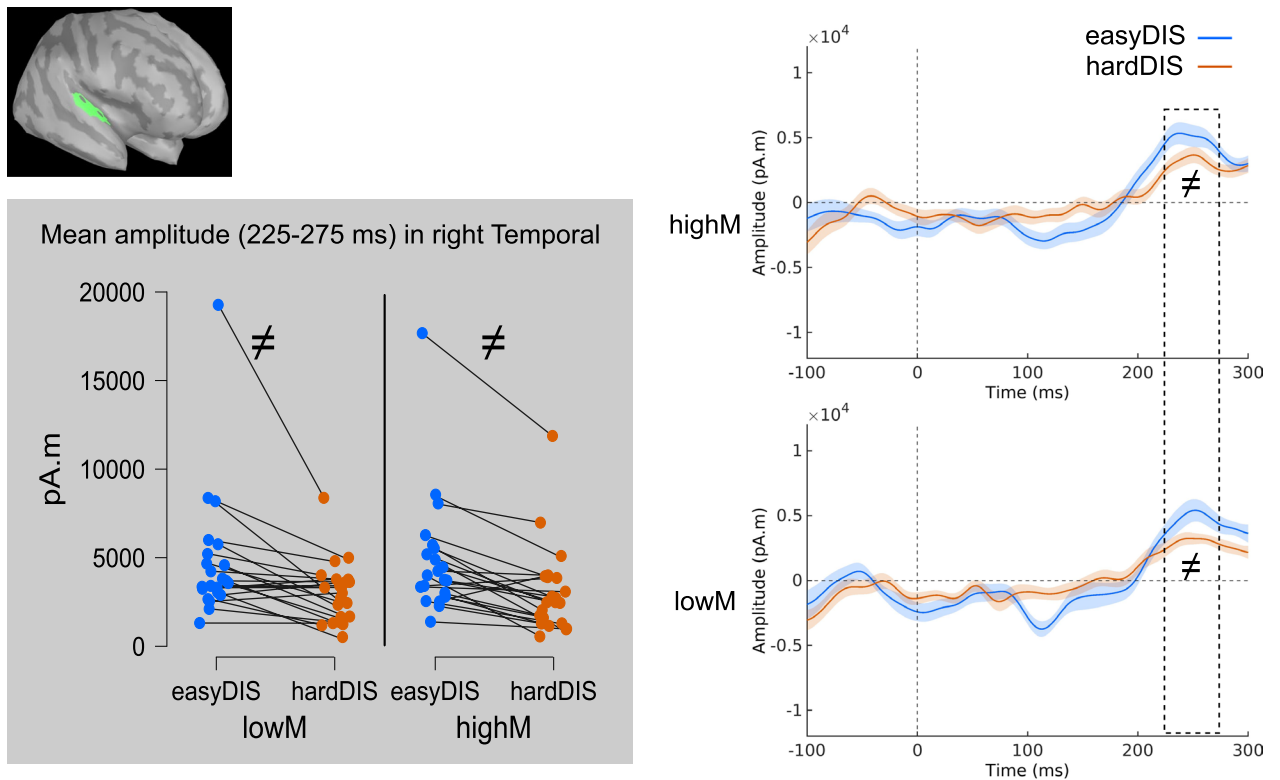


**Fig. 5.** Transient evoked responses to DIS tones 2, 3, and 4 in the Planum Polare ROIs (source space). Time course of the averaged ERF to the tones 2, 3, and 4 of the to-be-ignored melody DIS (tone onset at 0ms), for each DISdiff and MEMdiff condition, in the corresponding ROI represented on the MNI template. On the time course plots, shadowed areas represent standard errors of the mean. The strip plots show individual data averaged by condition in the dotted periods showing positive or greater evidence for a DISdiff effect (see Table 1 for time-windows and results). Lines connect data from the same participant. Only contralateral tones are analyzed here: DIS-R for left sources, and DIS-L for right sources. DIS-L or DIS-R: DIS melodies presented to the left or right ear, respectively. leftPP: left Planum Polare, right PP: right Planum Polare.

( $BF_{inclusion} = 42.9$ ) and weak evidence for the interaction between Stimulus\_Side and DISdiff ( $BF_{inclusion} = 2.0$ ), while there is weak evidence against an effect of Stimulus\_Side ( $BF_{inclusion} = 0.49$ ).

*Key result for the late differential response (tones 2, 3, 4)* The difference between the to-be-attended (S1) and the to-be-ignored (DIS) ERF amplitude is larger for easy-to-ignore distractors (easyDIS) than hard-to-ignore distractors (hardDIS).

## difference S1-DIS Tones 2, 3 & 4 | late differential response | right Temporal



**Fig. 6.** Transient evoked responses to the tones 2, 3, and 4, late differential response, in the right Temporal ROI (source space). Time course of the difference (S1-DIS) between averaged ERF to the tones 2, 3, and 4 of the to-be-attended melody S1 and of the to-be-ignored melody DIS (tone onset at 0ms), for each DISdiff and MEMdiff condition, in the right Temporal ROI represented on the MNI template. On the time course plots, shadowed areas represent standard errors of the mean. The strip plots show individual data averaged by condition in the dotted periods showing positive or greater evidence for a DISdiff effect (see Table 1 for time-windows and results). Lines connect data from the same participant.

### Processing of the sound sequences: sustained evoked responses

During S1 and DIS presentation, irrespective of S1 presentation side, emergence tests (time window 500–2000 ms after S1 onset) revealed activity in bilateral temporal areas, in two distinct clusters in the Planum Polare and the Planum Temporale. Activity in these temporal ROIs was examined with Bayesian repeated-measure ANOVAs with S1side, DISdiff, and MEMdiff as factors (see Table 1 and Figs. 7, S10 & S11).

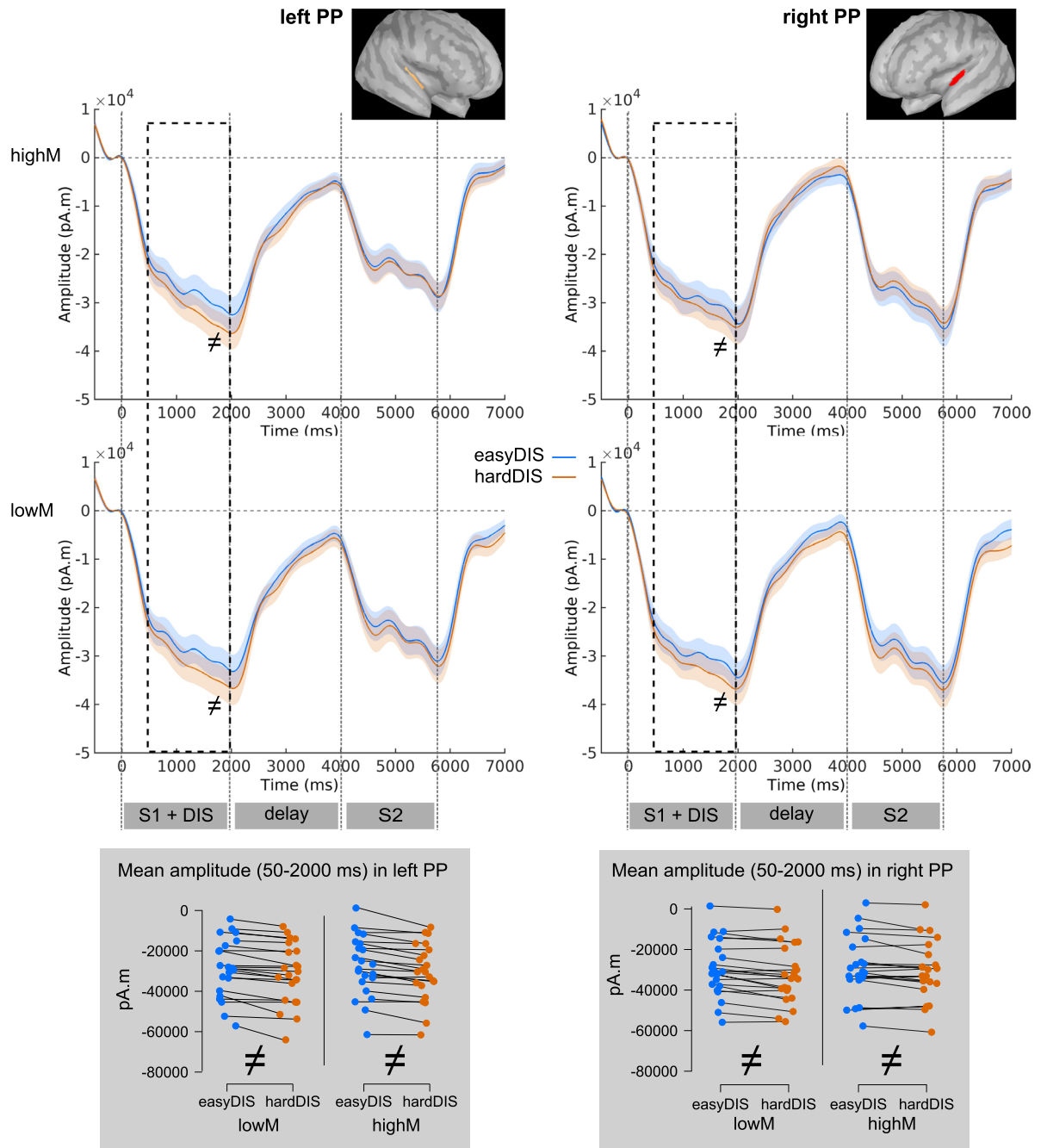
In the left Planum Polare, the best model explaining results during presentation of S1 and DIS melodies was the one with the factors DISdiff and S1side ( $BF_{10} = 129,061.0$ ), immediately followed by the model with only DISdiff ( $BF_{10} = 94,853.0$ ), and several other models with similar evidence. The analysis of effects only revealed decisive evidence for the effect of DISdiff ( $BF_{inclusion} = 89,087.1$ ) and weak evidence for the effect of S1side ( $BF_{inclusion} = 1.4$ ), whereas there was weak to positive evidence against all other effects ( $BF_{inclusion} < 0.63$ ). Sustained evoked responses in the left Planum Polare during S1 and DIS were larger in condition hardDIS compared to easyDIS.

In the right Planum Polare, the best model explaining sustained evoked response amplitudes during S1 and DIS was the one with only the factor DISdiff ( $BF_{10} = 207.1$ ). Other models explained almost as well the data, however the analysis of effects brought only decisive evidence for an effect of DISdiff in the right Planum Polare ( $BF_{inclusion} = 202.2$ ). As in the left hemisphere, response amplitudes were larger in condition hardDIS compared to easyDIS.

In the left Planum Temporale, the best model explaining sustained evoked response amplitude during S1 and DIS was the one with factors DISdiff, MEMdiff, S1side, and the interaction between DISdiff and MEMdiff ( $BF_{10} = 1.6$ ). Several other models, including the null model, explained almost as well the data. The analysis of effects only revealed weak evidence for the effect of DISdiff ( $BF_{inclusion} = 1.9$ ) and S1side ( $BF_{inclusion} = 2.2$ ), with differences in the same direction as in the Planum Polare, and weak evidence for no effect of MEMdiff ( $BF_{inclusion} = 0.90$ ) and of the interaction between DISdiff and MEMdiff ( $BF_{inclusion} = 0.47$ ).

In the right Planum Temporale, the best model explaining sustained evoked response amplitude during S1 and DIS is the null model (for all other models,  $BF_{10} < 0.65$ ).

S1 + DIS | Sustained responses | Planum Polare



**Fig. 7.** Sustained evoked responses in the Planum Polare ROIs (source space). Time course of the ERF for each DISdiff and each MEMdiff condition, in the corresponding ROIs represented on the MNI template. On the time course plots, shadowed areas represent standard errors of the mean. The strip plot shows individual data averaged by condition in the dotted periods showing positive or greater evidence for a DISdiff effect (see Table 1 for time-windows and results). Lines connect data from the same participant. leftPP: left Planum Polare, right PP: right Planum Polare.

**Key result for sustained evoked responses during sound sequence processing** In the bilateral Planum Polare, sustained evoked responses were larger when distractors were hard to ignore (hardDIS) compared to when distractors were easy to ignore (easyDIS).

#### *Rehearsing of the sound sequences: sustained evoked responses during the retention delay*

During the retention delay (time window 2500–4000 ms), sustained evoked responses were detected only in the left hemisphere, irrespective of S1 side, in the same areas as during S1 and DIS presentation. Activity in these left temporal ROIs was examined with Bayesian repeated-measure ANOVAs with S1side, DISdiff, and MEMdiff as factors.

In the left Planum Polare, the best model explaining sustained response amplitude was the null model (for all other models,  $BF_{10} < 0.57$ ).

In the left Planum Temporale, the best model explaining sustained response amplitude was the model with only the S1side effect ( $BF_{10} = 1.13$ ). Other models, including the null model, explained almost as well the data, and the analysis of effects revealed only weak evidence for the effect of S1side ( $BF_{inclusion} = 1.13$ ), with larger amplitudes ipsilaterally.

## Time–frequency results

### *Sensor-level activation*

The time–frequency power (relative change) plots computed on sensor-level data (see Figure S12) with a baseline from  $-1.8$  to  $-1$  s relative to S1 onset revealed a decrease in alpha power after cue onset at occipital sensors. A decrease in alpha power (7–11 Hz) was observed during the presentation of the S1 and DIS melodies and during the delay at left temporal sensors; while an increase in alpha power (11–15 Hz) was observed at occipital sensors. These two alpha sub-bands are named low and high alpha in the following. Finally, an increase in theta power (4–7 Hz) was observed during the retention delay at frontal sensors.

### *Source-level whole brain activation*

Sources of theta and alpha activities were estimated using DICS beamformer in different time–frequency windows of interest, and contrasted to a pre-cue baseline period using non-parametric cluster-based permutation testing, for S1 presented to the left or right ear separately (Figure S13).

The main power modulations were:

1. An increase in theta power in anterior prefrontal regions (BA 9, 10) during cue, melody presentation, and delay, more pronounced in the right hemisphere (Figure S13, top row);
2. A decrease in low alpha power in auditory regions in superior temporal cortices (BA 22, 41, 42), Broca area (BA 44, 45), and dorsolateral prefrontal cortex (DLPFC, BA 46) during melody presentation in both hemispheres, and during the delay in the left hemisphere (Figure S13, middle row);
3. A decrease in low alpha power in visual regions in occipital cortices (BA 17, 18, 19) during cue presentation (Figure S13, middle row) followed by an increase in high alpha power during melody presentation and delay (Figure S13, bottom row).

### *Source-level ROI activation*

According to the analysis of the individual peak frequency in each ROI, we investigated the following frequency bands: 4–7 Hz in anterior prefrontal regions; 8–11 Hz for decrease in alpha power in the auditory, visual, Broca and DLPFC ROIs; and 9–12 Hz for increase in alpha power in the visual ROI. To investigate the impact of the memory task and distractor difficulties, power values in time–frequency windows of interest were examined using Bayesian repeated-measure ANOVAs with S1side, DISdiff, and MEMdiff as factors (see Table 2 and Figs. 8 & S14).

**Anterior prefrontal theta.** In the anterior prefrontal regions, the best model explaining results in the 4–7 Hz band after cue onset and before S1 onset was the one with the factors DISdiff and S1side in the left hemisphere ( $BF_{10} = 5.0$ ) and with the factor MEMdiff in the right hemisphere ( $BF_{10} = 1.2$ ). The analysis of effects, in the left hemisphere, revealed positive evidence for the effect of S1side ( $BF_{inclusion} = 3.7$ ) and weak evidence for the effect of DISdiff ( $BF_{inclusion} = 1.4$ ). The increase in theta power was larger after the left cue presentation. The best model explaining results in the 4–7 Hz band after S1 onset was the one with the factor DISdiff in the left hemisphere ( $BF_{10} = 1.9$  between 400 and 1200 ms;  $BF_{10} = 3.3$  between 1200 and 2000 ms;  $BF_{10} = 4.5$  between 3200 and 4000 ms) and in the right hemisphere ( $BF_{10} = 10.6$  between 3200 and 4000 ms); otherwise the null model was the best model (for all other models,  $BF_{10} < 0.89$ ). The increase in theta power was larger in the hardDIS condition.

**Auditory and prefrontal low alpha.** In the left auditory regions, the best model explaining results in the 8–11 Hz band during the presentation of S1 and DIS melodies was the one with the factor MEMdiff only ( $BF_{10} = 3.4$  between 400 and 1200 ms,  $BF_{10} = 1.7$  between 1200 and 2000 ms), and during the delay, it was the null model. The decrease in alpha power was larger in the lowM condition.

In the right auditory regions, the best model explaining results in the 8–11 Hz band during presentation of S1 and DIS melodies was the one with the factor S1side ( $BF_{10} = 1.5$  between 400 and 1200 ms). Between 1200 and 2000 ms, the best model was the null model (for all other models,  $BF_{10} < 0.94$ ).

In the Broca areas, the best model explaining results in the 8–11 Hz band after S1 onset was the null model (for all other models,  $BF_{10} < 0.41$ ), except between 400 and 1200 ms (best model with factor S1side:  $BF_{10} = 1.1$ ) and between 3200 and 4000 ms (best model with factor MEMdiff:  $BF_{10} = 3.2$ ) in the left hemisphere. The decrease in alpha power was larger in the lowM condition.

In the left DLPFC, the best model explaining results in the 8–11 Hz band after S1 onset was the null model (for all other models,  $BF_{10} < 0.54$ ).

ROI	Brodmann areas	Hemisphere	Target frequency	STATISTICS				
				Cue	S1 + DIS presentation		Retention delay	
				-800:0 ms	400:1200 ms	1200:2000 ms	2400:3200 ms	3200:4000 ms
Anterior Prefrontal	9, 10	L	$\theta \nearrow$ 4-7 Hz	$\theta \nearrow \nearrow$ after left cue	-	$\theta \nearrow \nearrow$ in hardDIS	-	$\theta \nearrow \nearrow$ in hardDIS
		R		-	-	-	$\theta \nearrow \nearrow$ in hardDIS	
Auditory	22, 41, 42	L	$\alpha \searrow$ 8-11 Hz		$\alpha \searrow \searrow$ in lowM	-	-	-
		R			-	-		
Broca	44, 45	L	$\alpha \searrow$ 8-11 Hz		-	-	-	$\alpha \searrow \searrow$ in lowM
		R			-	-		
DLPFC	46	L	$\alpha \searrow$ 8-11 Hz		-	-	-	-
		R			-	-		
Visual	17, 18, 19	L	$\alpha \searrow$ 8-11 Hz	-				
		R		-				
		L	$\alpha \nearrow$ 9-12 Hz		-	-	-	-
		R			-	-	-	

**Table 2.** Summary of the TF analysis.

We report in the table only effects for which at least positive evidence (i.e.,  $BF_{10}$  or  $BF_{inclusion} > 3$ ) was observed with a Bayesian repeated-measure ANOVA with factors S1side, DISdiff (distractor filtering difficulty, easyDIS vs. hardDIS), and MEMdiff (memory task difficulty, lowM vs. highM; see text for the full results). Reported effects are relative to the opposite condition. “-” means no positive or greater evidence for an effect. Gray cells were not tested in accordance with the whole-brain source results.  $\nearrow$ : increase in power,  $\searrow$ : decrease in power.

*Visual low and high alpha.* In the bilateral visual regions, the best model explaining results in the 8–11 Hz band during cue presentation was the null model. In the bilateral visual regions, the best model explaining results in the 9–12 Hz band during S1 and DIS melodies and during the delay was the null model, except in the right visual regions between 400 and 1200 ms (best model with factor S1side:  $BF_{10} = 1.0$ ) and between 3200 and 4000 ms (best model with factors DISdiff and S1side:  $BF_{10} = 2.6$ ) (for all other models,  $BF_{10} < 0.77$ ). The decrease after the cue and the increase after S1 onset in alpha power did not seem to be modulated by any of the investigated factors.

*Key TF results* Prefrontal theta power increased bilaterally during sound sequence processing and the retention delay, more strongly so when distractors were difficult to ignore (hardDIS). Alpha power decreased bilaterally during sound sequence processing and the retention delay in auditory and inferior frontal (Broca) areas, more strongly so during the difficult memory task (highM) in the left hemisphere.

## Discussion

In the present study, we investigate auditory working memory and selective attention, in a delayed-matching-to-sample task (DMST) requiring to filter out distractors during encoding of melodic sequences in memory. Behavioral and MEG results allow to further understand working memory, selective attention, and their interactions in the auditory modality, as discussed below.

During the DMST at the core of the MEMAT paradigm, we observe a series of electrophysiological markers that were expected based on prior studies: (1) following the presentation of the visual cue (which itself generated visual ERFs and an alpha decrease in visual areas), in anticipation of the sequence to encode, a slow magnetic field develops bilaterally (magnetic counterpart of the CNV, see below), notably in the Planum Polare, and theta power increases in anterior prefrontal regions; (2) transient and sustained evoked responses are observed during sequence encoding in superior temporal and inferior frontal areas, along with an increase in theta power in anterior prefrontal areas, a decrease in alpha power in a fronto-temporal network, and an increase in alpha power in visual areas; (3) a sustained evoked response persists in superior temporal areas during the retention delay, also along with an increase in theta power in anterior prefrontal areas, a decrease in alpha power in a left fronto-temporal network, and an increase in alpha power in visual areas.

This pattern of transient and sustained ERFs is consistent with previous descriptions in auditory non-verbal DMST<sup>55,57</sup>. Here the addition of a visual cue at a fixed delay prior to S1 onset allows to characterize the slow response developing in anticipation of the sound sequences, that could reflect the sensory contribution to the CNV<sup>107</sup> and strongly resembles the auditory preparatory potential observed with intracranial recordings in epileptic patients slightly more posteriorly in Heschl gyri<sup>108</sup>.

Power in the alpha band is modulated along the course of a trial: a transient decrease in visual areas following the visual cue, and then a sustained increase in visual areas associated with a sustained decrease in fronto-temporal areas during encoding and maintenance (we did not study the retrieval phase). Following the

interpretation of the alpha rhythm as an inhibitory process<sup>48–52</sup>, this pattern suggests a transient activation of the visual system to process the visual cue, followed by an inhibition of task-irrelevant occipital visual areas and an activation of the task-relevant fronto-temporal auditory network during sound encoding and maintenance in memory. In keeping with ElShafei et al.<sup>45,46</sup>, alpha oscillations are centered around a higher frequency in visual areas compared to auditory areas during auditory processing.

Power in the theta band increases in anterior frontal areas during cue processing and the subsequent encoding and maintenance phases. Studies of short-term and working verbal memory with EEG/MEG have almost always uncovered increased frontal theta oscillations, in particular during the maintenance phase (review in<sup>58</sup>). As discussed in Hsieh and Ranganath<sup>109</sup>, the specific role of theta oscillations (and theta-gamma coupling) in short-term/working memory processes is debated, and might further depend on the regions involved (frontal vs. hippocampal areas for example). Midline frontal theta oscillations have been in particular associated with maintaining temporal relationships between items<sup>59</sup>. The present study reveals that frontal theta oscillations are also involved in non-verbal auditory memory, both during encoding and maintenance of pitch sequences.

Here, we manipulate memory task difficulty at the block level, and participants were warned about the difficulty of the upcoming block of trials. At the behavioral level, as expected<sup>76,77</sup>, accuracy decreases and response time increases when the difficulty of the memory task increases. At the neurophysiological level, an increase in the amplitude of the slow ERF (CNV) in anticipation of the target melody is observed in the high difficulty memory task relative to the low difficulty one when the left ear is cued, in the contralateral right auditory cortex (Fig. 2). This result confirms the efficiency of the manipulation of task difficulty, with more pronounced anticipation of the relevant sound when the memory task is more difficult.

Beyond this anticipation effect, a smaller decrease in alpha oscillations was observed in the high difficulty memory condition relative to the low difficulty one, in the left auditory cortex during encoding of the S1 melody in memory (Fig. 8B), and in the left inferior frontal cortex (Broca's area) during the retention delay (Fig. 8C). These results in the alpha band suggest that in the difficult memory condition, the left auditory fronto-temporal network is less recruited than during the easier memory condition. We will return to this point when discussing the interaction between memory and attention.

During sequence encoding in MEMAT, a distracting melody is presented in the other ear, which is either easy or hard to ignore (on different trials within the same blocks). Participants succeed in selecting to-be-attended (S1) over to-be-ignored (DIS) melodies as evidenced by larger ERFs for target tones over distracting tones in temporal and frontal regions (N1, P2, late response for notes 2-3-4, see Figs. 4, 5 and 6 and Table 1; review in<sup>26</sup>). The difficulty of the attentional filtering impedes as expected on performance and reaction times<sup>76,77</sup>, with poorer performance and longer RTs for difficult-to-ignore distractors. This filtering difficulty modulates several of the electrophysiological markers listed above (see Tables 1 and 2), during the encoding phase of the DMST (transient and sustained ERFs, theta oscillations), as well as during the maintenance phase (theta oscillations). These modulations of electrophysiological responses shed light on how selective attention operates during a working memory task.

In response to the first distracting tone, a larger N1 amplitude is observed in response to the hard-to-ignore distractor than to the easy-to-ignore distractor in the left temporal areas (Fig. 3). As in MEMAT we manipulate the frequency similarity between to-be-attended and to-be-ignored sounds to create two levels of filtering difficulty, the hard-to-ignore sounds being closer in frequency to the to-be-attended melody could catch more attention and be more difficult to suppress than the easy-to-ignore sounds, in line with contingent stimulus-driven orienting<sup>110,111</sup>. This suggests a more efficient inhibition of the easy-to-ignore distracting tones as observed in previous EEG studies of auditory selective attention<sup>37,40</sup>.

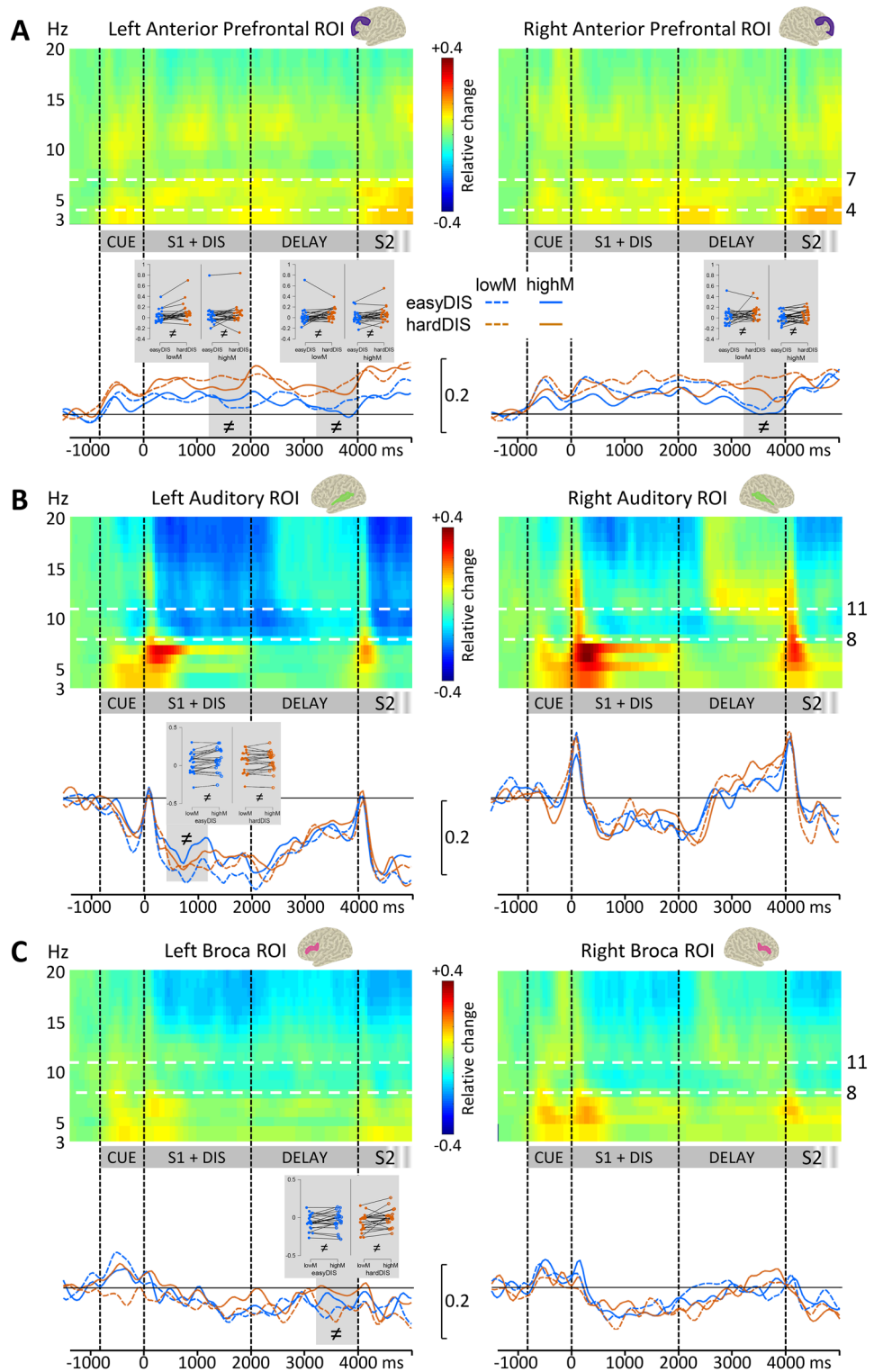
Critically, in bilateral superior temporal regions, in response to subsequent tones, N1 amplitude for the to-be-attended tones is larger when accompanied with easy-to-ignore distracting melodies, whereas the N1 to the to-be-ignored tones is not modulated by filtering difficulty (Fig. 4). It suggests a facilitation of the processing of the target, more efficiently so when the distracting sounds are dissimilar to the target, in line with previous findings<sup>37,38</sup>.

Later on, more anteriorly, in bilateral planum polare, the amplitude of the P2 to the tones of the melody to ignore is smaller when these tones were easy rather than difficult to ignore, in keeping with previous results<sup>37</sup>. This suggests a more efficient inhibition of the distracting tone processing when they are easy to ignore (Fig. 5).

Around 250 ms, a larger ERF amplitude for to-be-attended tones than for to-be-ignored tones in right temporal areas seems to be related to an additional component superimposed to the obligatory P2 (Fig. 6), such as the PN (Processing Negativity)<sup>112,113</sup>. This effect is more pronounced when the to-be-ignored tones are easy to ignore, i.e. more dissimilar to the to-be-attended tones<sup>30</sup> or when the task is easier<sup>40</sup>. This increased component suggests a larger differential processing of relevant and irrelevant tones when distractors are easy to ignore, and has been interpreted as reflecting active suppression of irrelevant signal<sup>30,40</sup>.

Taken together, these findings highlight that distinct attentional mechanisms are at play, with facilitation of target processing preceding inhibition of irrelevant tones, in keeping with previous findings<sup>28–32,37,38,40</sup>. These two effects have distinct localizations: auditory areas in the vicinity of the primary auditory cortex for target facilitation, and more anterior superior temporal areas in the planum polare for distractor inhibition, in line with previous observations<sup>114</sup>. Except for the later differential effect (~250 ms) which was observed only in the right hemisphere, all these effects on transient ERFs were bilateral in agreement with most previous studies. All these mechanisms are less efficient when targets and distractors are more similar<sup>37,38</sup>.

During sound sequence encoding, in addition to transient ERFs following each tone of the target or distracting melodies, a sustained evoked response was observed in bilateral superior temporal areas. In bilateral planum polare, the sustained evoked response is larger when the distractors are hard-to-ignore compared to easy-to-ignore, suggesting that this component reflects the deployment of cognitive resources (listening effort) to solve the difficult filtering task. A similar modulation of the sustained evoked response during S1 encoding



by task difficulty in an auditory DMST was also observed in Albouy et al.<sup>55</sup> It is noteworthy here that sustained evoked responses and target-related transient ERFs are modulated in opposite directions by the difficulty to filter out the distracting melodies, suggesting that the deployment of additional resources when the distractors are hard-to-ignore (reflected in the sustained response) is not sufficient to restore the same level of differential processing between targets and distractors in auditory areas (reflected in target and distractor-related transient ERFs). This failure translates in behavioral data as costs of filtering difficulty, both for  $d'$  and RT.

These effects on the ERFs highlight that during sound sequence encoding in memory, the processing of targets (and distractors) is modulated by attentional processes, possibly resulting in more or less precise memory traces of target sounds, which could impact on performance in the retrieval phase. Furthermore, we observe that theta power during sound sequence encoding and retention is larger when the distracting melodies is hard to ignore (compared to when they were easy to ignore). As described above, frontal theta oscillations have

◀ **Fig. 8.** Time–frequency plots and time profiles in the source space. Time–frequency power (relative change) plots during Cue presentation, S1 & DIS melodies, and the delay (S1 onset at 0ms), for all conditions averaged, in the left and right Anterior Prefrontal regions (A), in the left and right Auditory cortices (B), in the left and right Broca areas (C), in the left and right Dorso-Lateral Prefrontal Cortex (DLPFC, D), and in the left and right Visual cortices (E). Response to S2 is not fully represented here. In comparison to baseline (–1800 to –1000 ms relative to S1 onset), a decrease in power appears in blue and an increase in power appears in red. Regions of interest were defined based on the results from the cluster-based permutation tests performed on the DICS-reconstructed sources for the whole brain. Time course of the power (relative change) for each DISdiff and each MEMdiff condition, in the corresponding ROIs represented on the MNI template: alpha band (8 to 11 Hz) in the Auditory (A) and BROCA (B) ROIs, theta band (4 to 7 Hz) in the Anterior Prefrontal (C) ROIs. The strip plot shows individual data averaged by condition in the gray periods showing positive or greater evidence for a DISdiff effect (see Table 2 for time-windows and results, and figure S14 for a larger version of the strip plots). Lines connect data from the same participant.

been associated with working memory processes including the processing of serial order in sequences<sup>59</sup>. The current result would thus imply that a stronger involvement of these time-based memory processes is necessary when parsing interleaved target and distracting sounds is more demanding. This increase in theta power with increasing filtering difficulty is also in line with a link between theta activity and increased effort<sup>115</sup>.

At the behavioral level, we observe, as in Blain, de la Chapelle et al.<sup>77</sup> and Blain, Talamini et al.<sup>76</sup>, an interaction between filtering difficulty and the memory task difficulty on performance ( $d'$ ), with greater differences between easy and hard-to-ignore distractors in the low difficulty memory task than in the high difficulty memory task. In Blain, Talamini, et al.<sup>76</sup>, based on several lines of evidence derived from behavioral measures in non-musician and musician participants, we have argued that this interaction observed in  $d'$  reflects shared cognitive resources between auditory selective attention and working memory, compatible with the *Cognitive Load theory*<sup>65,66,116</sup>.

In the neurophysiological data, we observe an interaction between filtering difficulty and the memory task difficulty on the amplitude of the N1 to to-be-attended sounds in the left auditory cortex during sequence encoding, with a difference of target N1 amplitude according to filtering difficulty only in the low difficulty memory task. This neurophysiological result exactly mirrors the behavioral data (Figs. 1 and 4). It suggests that under low memory load, attentional filtering is more efficient, resulting in larger response amplitudes for the to-be-attended sounds when the distractors are easy to ignore. Conversely, under high memory load, less resources would be available for attentional filtering, resulting in a reduced facilitation of relevant sound processing when the distractors are easy to ignore. Interestingly this interaction between filtering difficulty and memory task difficulty on the amplitude of the N1 to target sounds is only observed in the left auditory cortex, where a smaller decrease in alpha power is observed under high memory task difficulty. Hence, in the left auditory areas, under high memory load, the level of excitability is lower (larger alpha power) and the attentional facilitation of transient evoked responses to relevant sounds is reduced when distracting sounds are hard-to-ignore. A possible interpretation is that the left auditory cortex plays a critical role in attentional filtering<sup>28</sup>, and that increasing the cognitive load with enhanced memory task difficulty impedes on these processes. In the framework of the cognitive load theory, we could interpret this finding as attributing less cognitive resources to a process subserving selective attention. It is worth emphasizing that in the case of the MEMAT paradigm used here, memory processes and attentional filtering compete for cognitive resources already during sound sequence encoding, in contrast with previous studies targeting the maintenance phase, which could account for some seemingly discrepant findings (e.g.,<sup>32,72</sup>).

Further evidence for an interplay between attentional filtering and memory processes was obtained by analyzing frontal theta oscillations. During both sound sequence encoding and the maintenance delay, the power of theta oscillations which are thought to reflect time-based serial order processes in memory tasks<sup>59</sup> was increased when target selection was difficult. This suggests that time-based processes, which are relevant both for information selection and retention could be shared between attention and memory.

One possible limitation of the present study is the block-wise manipulation of memory load while attention difficulty varied within each block. Although this design could introduce bias, we aimed to minimize anticipatory strategies by implementing a within-block attention manipulation, which has been shown to yield similar behavioral results to a block-wise manipulation of attention<sup>76</sup>.

Here by manipulating the difficulties of an auditory memory task (DMST) and of attentional filtering of distracting sounds, we could shed light on the neurophysiological processes subserving working memory, selective attention, and their interactions in the auditory modality. Attentional selection during sound sequence encoding operates through a combination of target processing *facilitation* and distractor *inhibition* that are reduced with increasing filtering difficulty. Increasing attentional demands increases sustained evoked activity in superior temporal areas, which could favor filtering; while also enhancing theta oscillations in anterior frontal areas, which could reflect the involvement of a general time-based processing system. Conversely, increasing memory demands results in better anticipation of the sound sequences to encode (higher CNV amplitude) but also in a smaller desynchronization of alpha oscillations in the left auditory and frontal areas which could hinder attentional filtering, as reflected in target-related ERFs. In the left auditory cortex, we directly observed the interaction between auditory memory and attention: the facilitation of relevant sound processing in the easy filtering condition decreased as the difficulty of the memory task increased. This pattern mirrors the observed behavioral effects. Overall brain dynamics highlight reciprocal influences of working memory and selective attention processes at early processing stages, in keeping with shared cognitive resources between them. Future work is needed to understand how the interplay between these different processes is orchestrated.

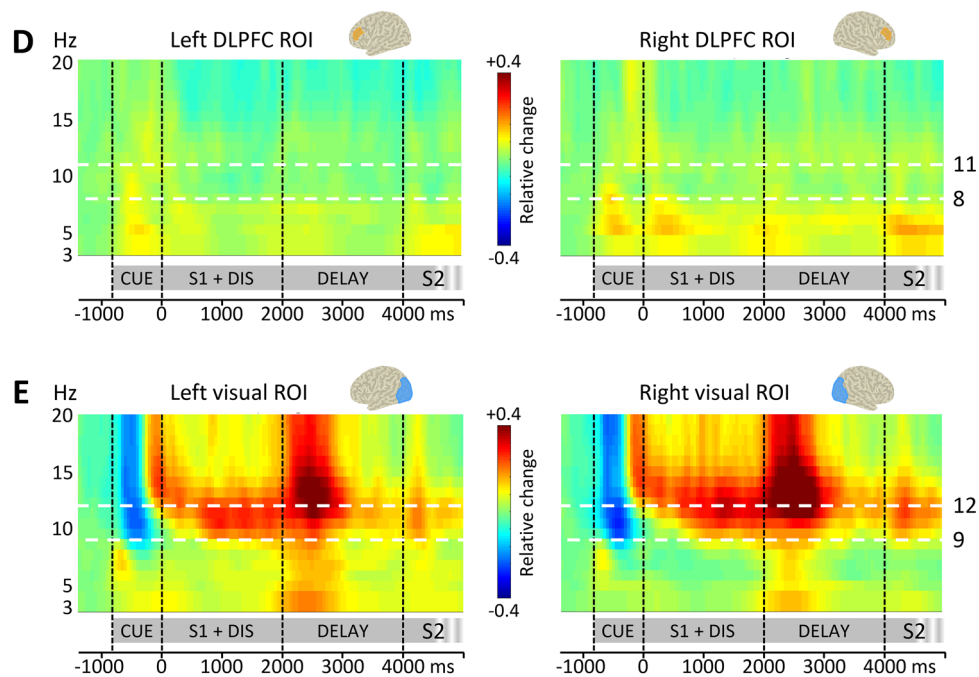


Fig. 8. (continued)

### Data availability

The data that support the findings of this study are not publicly available because there is no approval from the local ethic committee for sharing. Data are however available from the authors upon reasonable request and with permission of the ethic committee.

Received: 1 December 2023; Accepted: 16 May 2025

Published online: 01 July 2025

### References

- O'Leary, D. S. et al. Auditory and visual attention assessed with PET. *Hum. Brain Mapp.* **5**, 422–436 (1997).
- Lipschutz, B., Kolinsky, R., Damhaut, P., Wikler, D. & Goldman, S. Attention-dependent changes of activation and connectivity in dichotic listening. *Neuroimage* **17**, 643–656 (2002).
- Petkov, C. I. et al. Attentional modulation of human auditory cortex. *Nat. Neurosci.* **7**, 658–663 (2004).
- Kawashima, R. et al. Selective visual and auditory attention toward utterances—A PET study. *Neuroimage* **10**, 209–215 (1999).
- Belin, P. et al. The functional anatomy of sound intensity discrimination. *J. Neurosci.* **18**, 6388–6394 (1998).
- Tzourio, N. et al. Functional anatomy of human auditory attention studied with PET. *Neuroimage* **5**, 63–77 (1997).
- Alho, K. et al. Selective tuning of the left and right auditory cortices during spatially directed attention. *Cogn. Brain Res.* **7**, 335–341 (1999).
- Buchsbaum, B. R. & D'Esposito, M. A sensorimotor view of verbal working memory. *Cortex* **112**, 134–148 (2019).
- Zatorre, R. J., Evans, A. C. & Meyer, E. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* **14**, 1908–1919 (1994).
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D. & Koelsch, S. Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* **32**, 771–783 (2011).
- Buchsbaum, B. R. et al. Conduction aphasia, sensory-motor integration, and phonological short-term memory—An aggregate analysis of lesion and fMRI data. *Brain Lang.* **119**, 119–128 (2011).
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L. & Schlaug, G. Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *Neuroimage* **19**, 1417–1426 (2003).
- Albouy, P. et al. Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia. *Hum. Brain Mapp.* **40**, 855–867 (2019).
- Kumar, S. et al. A brain system for auditory working memory. *J. Neurosci.* **36**, 4492–4505 (2016).
- Koelsch, S. et al. Functional architecture of verbal and tonal working memory: An FMRI study. *Hum. Brain Mapp.* **30**, 859–873 (2009).
- Foster, N. E. V. & Zatorre, R. J. A role for the intraparietal sulcus in transforming musical pitch information. *Cereb. Cortex* **20**, 1350–1359 (2010).
- Foster, N. E. V. & Zatorre, R. J. Cortical structure predicts success in performing musical transformation judgments. *Neuroimage* **53**, 26–36 (2010).
- Huang, S., Seidman, L. J., Rossi, S. & Ahveninen, J. Distinct cortical networks activated by auditory attention and working memory load. *Neuroimage* **83**, 1098–1108 (2013).
- Weisz, N., Müller, N., Jatzew, S. & Bertrand, O. Oscillatory alpha modulations in right auditory regions reflect the validity of acoustic cues in an auditory spatial attention task. *Cereb. Cortex* **24**, 2579–2590 (2014).
- McLaughlin, S. A., Larson, E. & Lee, A. K. C. Neural switch asymmetry in feature-based auditory attention tasks. *J. Assoc. Res. Otolaryngol. JARO* **20**, 205–215 (2019).
- de Vries, I. E. J., Marinato, G. & Baldauf, D. Decoding object-based auditory attention from source-reconstructed MEG alpha oscillations. *J. Neurosci.* **41**, 8603–8617 (2021).

22. Marinato, G. & Baldauf, D. Object-based attention in complex, naturalistic auditory streams. *Sci. Rep.* **9**, 2854 (2019).
23. Hillyard, S. A., Hink, R. F., Schwent, V. L. & Picton, T. W. Electrical signs of selective attention in the human brain. *Science* **182**, 177–180 (1973).
24. Woldorff, M., Hansen, J. C. & Hillyard, S. A. Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential. *Electroencephalogr. Clin. Neurophysiol. Suppl.* **40**, 146–154 (1987).
25. Woods, D. L., Alho, K. & Algazi, A. Brain potential signs of feature processing during auditory selective attention. *NeuroReport* **2**, 189–192 (1991).
26. Giard, M.-H., Fort, A., Mouchetant-Rostaing, Y. & Pernier, J. Neurophysiological mechanisms of auditory selective attention in humans. *Front. Biosci.-Landmark* **5**, 84–94 (2000).
27. Picton, T. W., Woods, D. L. & Proulx, G. B. Human auditory sustained potentials. I. The nature of the response. *Electroencephalogr. Clin. Neurophysiol.* **45**, 186–197 (1978).
28. Bidet-Caulet, A. et al. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.* **27**, 9252–9261 (2007).
29. Michie, P. T., Solowij, N., Crawford, J. M. & Glue, L. C. The effects of between-source discriminability on attended and unattended auditory ERPs. *Psychophysiology* **30**, 205–220 (1993).
30. Alho, K., Töttölä, K., Reinikainen, K., Sams, M. & Näätänen, R. Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalogr. Clin. Neurophysiol. Potentials Sect.* **68**, 458–470 (1987).
31. Alho, K., Woods, D. L. & Algazi, A. Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology* **31**, 469–479 (1994).
32. Bidet-Caulet, A., Mikyska, C. & Knight, R. T. Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. *Neuroimage* **50**, 277–284 (2010).
33. Chait, M., de Cheveigné, A., Poeppel, D. & Simon, J. Z. Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia* **48**, 3262–3271 (2010).
34. Donald, M. W. The timing and polarity of different attention-related ERP changes inside and outside of the attentional focus. *Electroencephalogr. Clin. Neurophysiol. Suppl.* **40**, 81–86 (1987).
35. Schroger, E. & Eimer, M. Endogenous covert spatial orienting in audition cost-benefit analyses of reaction times and event related potentials. *Q. J. Exp. Psychol. Sect. A* **50**, 457–474 (1997).
36. Lange, K. & Schnuerch, R. Challenging perceptual tasks require more attention: The influence of task difficulty on the N1 effect of temporal orienting. *Brain Cogn.* **84**, 153–163 (2014).
37. Alain, C. & Woods, D. L. Signal clustering modulates auditory cortical activity in humans. *Percept. Psychophys.* **56**, 501–516 (1994).
38. Alain, C., Achim, A. & Richer, F. Perceptual context and the selective attention effect on auditory event-related brain potentials. *Psychophysiology* **30**, 572–580 (1993).
39. Ponjavic-Conte, K. D., Hambrook, D. A., Pavlovic, S. & Tata, M. S. Dynamics of distraction: competition among auditory streams modulates gain and disrupts inter-trial phase coherence in the human electroencephalogram. *PLoS ONE* **8**, e53953 (2013).
40. Melara, R. D., Chen, S. & Wang, H. Inhibiting change: Effects of memory on auditory selective attention. *Cogn. Brain Res.* **25**, 431–442 (2005).
41. de Vries, I. E. J., Marinato, G. & Baldauf, D. Decoding object-based auditory attention from source-reconstructed MEG alpha oscillations. *J. Neurosci. Off. J. Soc. Neurosci.* **41**, 8603–8617 (2021).
42. Müller, N. & Weisz, N. Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. *Cereb. Cortex* **22**, 1604–1613 (2012).
43. Frey, J. N. et al. Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. *J. Neurosci.* **34**, 6634–6639 (2014).
44. Weise, A., Hartmann, T., Schröger, E., Weisz, N. & Ruhnau, P. Cross-modal distractors modulate oscillatory alpha power: the neural basis of impaired task performance. *Psychophysiology* **53**, 1651–1659 (2016).
45. ElShafei, H. A., Bouet, R., Bertrand, O. & Bidet-Caulet, A. Two sides of the same coin: distinct sub-bands in the  $\alpha$  rhythm reflect facilitation and suppression mechanisms during auditory anticipatory attention. *eneuro* **5**(4), ENEURO.0141-18.2018. <https://doi.org/10.1523/ENEURO.0141-18.2018> (2018).
46. ElShafei, H. A., Fornoni, L., Masson, R., Bertrand, O. & Bidet-Caulet, A. Age-related modulations of alpha and gamma brain activities underlying anticipation and distraction. *PLoS ONE* **15**, e0229334 (2020).
47. Wöstmann, M., Herrmann, B., Maess, B. & Obleser, J. Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc. Natl. Acad. Sci.* **113**, 3873–3878 (2016).
48. Klimesch, W., Sauseng, P. & Hanslmayr, S. EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Res. Rev.* **53**, 63–88 (2007).
49. Foxe, J. J. & Snyder, A. C. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2011.00154> (2011).
50. Jensen, Ole & Mazaheri, Ali. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Human Neurosci* <https://doi.org/10.3389/fnhum.2010.00186> (2010).
51. Thut, G., Nietzel, A., Brandt, S. A. & Pascual-Leone, A.  $\alpha$ -band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* **26**, 9494–9502 (2006).
52. Palva, S. & Palva, J. M. Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front. Psychol.* **2**, 204 (2011).
53. Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J. & Nobre, A. C. Inter- and intra-individual variability in alpha peak frequency. *Neuroimage* **92**, 46–55 (2014).
54. Bagherzadeh, Y. Alpha synchrony and the neurofeedback control of spatial attention.
55. Albouy, P. et al. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* **136**, 1639–1661 (2013).
56. Lefebvre, C. et al. Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia* **51**, 2939–2952 (2013).
57. Grimault, S. et al. Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography. *Neuroimage* **94**, 96–106 (2014).
58. Pavlov, Y. G. & Kotchoubey, B. Oscillatory brain activity and maintenance of verbal and visual working memory: A systematic review. *Psychophysiology* **59**, e13735 (2022).
59. Hsieh, L.-T., Ekstrom, A. D. & Ranganath, C. Neural oscillations associated with item and temporal order maintenance in working memory. *J. Neurosci.* **31**, 10803–10810 (2011).
60. Canolty, R. T. et al. High gamma power is phase-locked to theta oscillations in human neocortex. *Science* **313**, 1626–1628 (2006).
61. Lisman, J. E. & Jensen, O. The theta-gamma neural code. *Neuron* **77**, 1002–1016 (2013).
62. Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A. & Maess, B. Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* **32**, 12376–12383 (2012).
63. Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P. & Caclin, A. Impaired short-term memory for pitch in congenital amusia. *Brain Res.* **1640**, 251–263 (2016).

64. de Fockert, J. W., Rees, G., Frith, C. D. & Lavie, N. The role of working memory in visual selective attention. *Science* **291**, 1803–1806 (2001).
65. Lavie, N., Hirst, A., de Fockert, J. W. & Viding, E. Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* **133**, 339–354 (2004).
66. Lavie, N. Distracted and confused?: Selective attention under load. *Trends Cogn. Sci.* **9**, 75–82 (2005).
67. Sörqvist, P., Stenfelt, S. & Rönnerberg, J. Working memory capacity and visual-verbal cognitive load modulate auditory-sensory gating in the brainstem: Toward a unified view of attention. *J. Cogn. Neurosci.* **24**, 2147–2154 (2012).
68. Sörqvist, P., Nössl, A. & Halin, N. Working memory capacity modulates habituation rate: Evidence from a cross-modal auditory distraction paradigm. *Psychon. Bull. Rev.* **19**, 245–250 (2012).
69. Sörqvist, P. & Marsh, J. E. How concentration shields against distraction. *Curr. Dir. Psychol. Sci.* **24**, 267–272 (2015).
70. Bayramova, R., Toffalini, E., Bonato, M. & Grassi, M. Auditory selective attention under working memory load. *Psychol. Res.* **85**, 2667–2681 (2021).
71. Berti, S. & Schröger, E. Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *Eur. J. Neurosci.* **17**, 1119–1122 (2003).
72. SanMiguel, I., Corral, M.-J. & Escera, C. When loading working memory reduces distraction: Behavioural and electrophysiological evidence from an auditory-visual distraction paradigm. *J. Cogn. Neurosci.* **20**, 1131–1145 (2008).
73. Salamé, P. & Baddeley, A. Effects of background music on phonological short-term memory. *Q. J. Exp. Psychol.* **41**, 107–122 (1989).
74. Banbury, S. P., Macken, W. J., Tremblay, S. & Jones, D. M. Auditory distraction and short-term memory: Phenomena and practical implications. *Hum. Factors* **43**, 12–29 (2001).
75. Majerus, S., Péters, F., Bouffier, M., Cowan, N. & Phillips, C. The dorsal attention network reflects both encoding load and top-down control during working memory. *J. Cogn. Neurosci.* **30**, 144–159 (2018).
76. Blain, S., Talamini, F., Fornoni, L., Bidet-Caulet, A. & Caclin, A. Shared cognitive resources between memory and attention during sound-sequence encoding. *Atten. Percept. Psychophys.* **84**, 739–759 (2022).
77. Blain, S., de la Chapelle, A., Caclin, A., Bidet-Caulet, A. & Ruby, P. Dream recall frequency is associated with attention rather than with working memory abilities. *J. Sleep Res.* **31**, e13557 (2022).
78. George, E. M. & Coch, D. Music training and working memory: An ERP study. *Neuropsychologia* **49**, 1083–1094 (2011).
79. Talamini, F., Carretti, B. & Grassi, M. The working memory of musicians and nonmusicians. *Music Percept.* **34**, 183–191 (2016).
80. Talamini, F., Altoè, G., Carretti, B. & Grassi, M. Musicians have better memory than nonmusicians: A meta-analysis. *PLoS ONE* **12**, e0186773 (2017).
81. Talamini, F. et al. Auditory and visual short-term memory: influence of material type, contour, and musical expertise. *Psychol. Res.* **86**, 421–442 (2022).
82. Ruby, P. et al. Alpha reactivity to first names differs in subjects with high and low dream recall frequency. *Front. Psychol.* **4**, 419 (2013).
83. Eichenlaub, J.-B., Bertrand, O., Morlet, D. & Ruby, P. Brain reactivity differentiates subjects with high and low dream recall frequencies during both sleep and wakefulness. *Cereb. Cortex N. Y. N* **1991**(24), 1206–1215 (2014).
84. Eichenlaub, J.-B. et al. Resting brain activity varies with dream recall frequency between subjects. *Neuropsychopharmacology* **39**, 1594–1602 (2014).
85. Vallat, R., Nicolas, A. & Ruby, P. Brain functional connectivity upon awakening from sleep predicts interindividual differences in dream recall frequency. *Sleep* **43**(12), 116 (2020).
86. Vallat, R., Türker, B., Nicolas, A. & Ruby, P. High dream recall frequency is associated with increased creativity and default mode network connectivity. *Nat. Sci. Sleep* **14**, 265–275 (2022).
87. Ruby, P. et al. High dream recall frequency is associated with an increase of both bottom-up and top-down attentional processes. *Cereb. Cortex N. Y. N* **1991**(32), 3752–3762 (2022).
88. JASP Team. JASP. (2022).
89. Marsman, M. & Wagenmakers, E.-J. Bayesian benefits with JASP. *Eur. J. Dev. Psychol.* **14**, 545–555 (2017).
90. Wagenmakers, E.-J. et al. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychon. Bull. Rev.* **25**, 58–76 (2018).
91. van den Bergh, D., Wagenmakers, E.-J. & Aust, F. Bayesian repeated-measures ANOVA: An updated methodology implemented in JASP. Preprint at <https://doi.org/10.31234/osf.io/fb8zn> (2022).
92. Lee, M. D. & Wagenmakers, E.-J. *Bayesian cognitive modeling: a practical course*. (Cambridge University Press, 2014).
93. Aguera, P.-E., Jerbi, K., Caclin, A. & Bertrand, O. ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Comput. Intell. Neurosci.* **5**(1–5), 11 (2011).
94. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
95. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).
96. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci* **2011**, 1–13. <https://doi.org/10.1155/2011/879716> (2011).
97. Gramfort, A., Papadopoulos, T., Olivi, E. & Clerc, M. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomed. Eng. OnLine* **9**, 45 (2010).
98. Teichmann, L., Moerel, D., Baker, C. & Grootswagers, T. An empirically driven guide on using bayes factors for m/eeg decoding. *Aperture Neuro* **2**, 1–10 (2022).
99. Dienes, Z. How bayes factors change scientific practice. *J. Math. Psychol.* **72**, 78–89 (2016).
100. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* <https://doi.org/10.1155/2011/156869> (2011).
101. Tallon-Baudry, C. & Bertrand, O. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.* **3**, 151–162 (1999).
102. Nolte, G. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.* **48**, 3637 (2003).
103. Gross, J. et al. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proc. Natl. Acad. Sci.* **98**, 694–699 (2001).
104. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
105. Van Veen, B. D., Van Drongelen, W., Yuchtman, M. & Suzuki, A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* **44**, 867–880 (1997).
106. Donoghue, T. et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**, 1655–1665 (2020).
107. Brunia, C. H. M. & Van Boxtel, G. J. M. Wait and see. *Int. J. Psychophysiol.* **43**, 59–75 (2001).
108. Hu, M. et al. Neural processes of auditory perception in Heschl's gyrus for upcoming acoustic stimuli in humans. *Hear. Res.* **388**, 107895 (2020).
109. Hsieh, L.-T. & Ranganath, C. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *Neuroimage* **85**, 721–729 (2014).

110. Folk, C. L., Remington, R. W. & Johnston, J. C. Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 1030–1044 (1992).
111. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).
112. Näätänen, R. Processing negativity: An evoked-potential reflection. *Psychol. Bull.* **92**, 605–640 (1982).
113. Näätänen, R. *Attention and brain function* (Lawrence Erlbaum Associates, 1992).
114. Rif, J., Hari, R., Hämäläinen, M. S. & Sams, M. Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalogr. Clin. Neurophysiol.* **79**, 464–472 (1991).
115. Wisniewski, M. G., Thompson, E. R. & Iyer, N. Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology* **54**, 1916–1928 (2017).
116. Yi, D.-J., Woodman, G. F., Widders, D., Marois, R. & Chun, M. M. Neural fate of ignored stimuli: Dissociable effects of perceptual and working memory load. *Nat. Neurosci.* **7**, 992–996 (2004).

## Acknowledgements

We would like to thank Lesly Fornoni and Anne Cheylus for their technical support.

## Author contributions

S.S.B, A.C and A.B.C. conceptualized the experiment. S.S.B. and S.D. recorded the data. A.D.L.C. and S.S.B. analyzed the data. A.D.L.C., S.S.B, A.E. contributed to programming analysis tools. D.S. and S.D. carried out data curation. D.S., J.J., A.C and A.B.C. coordinated responsibility for the research activity planning and execution. A.D.L.C., A.C. and A.B.C. wrote the manuscript. P.R. reviewed the manuscript. A.B.C. prepared data presentation. P.R., A.C. and A.B.C. supervised the project.

## Funding

This work was supported by a grant from “Agence Nationale de la Recherche” (ANR) of the French Ministry of Research ANR-14-CE30-0001-01 (attributed to ABC). This work was conducted within the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) of Université de Lyon, within the program “Investissements d’avenir” (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR).

## Declarations

### Competing interest

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-03378-w>.

**Correspondence** and requests for materials should be addressed to A.B.-C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025