

Neural dynamics underlying successful auditory short-term memory performance

Ulrich Pomper^{1,2}  | Lorenza Zaira Curetti³ | Maria Chait¹ 

¹Ear Institute, University College London, London, UK

²Faculty of Psychology, University of Vienna, Vienna, Austria

³Manchester Centre for Audiology and Deafness, University of Manchester, Manchester, UK

Correspondence

Ulrich Pomper, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria.

Email: ulrich.pomper@univie.ac.at

Funding information

EC Horizon 2020; BBSRC project grant, Grant/Award Number: BB/P003745/1; Austrian Science Fund, Grant/Award Number: ZK66

Edited by: John Foxe

Abstract

Listeners often operate in complex acoustic environments, consisting of many concurrent sounds. Accurately encoding and maintaining such auditory objects in short-term memory is crucial for communication and scene analysis. Yet, the neural underpinnings of successful auditory short-term memory (ASTM) performance are currently not well understood. To elucidate this issue, we presented a novel, challenging auditory delayed match-to-sample task while recording MEG. Human participants listened to ‘scenes’ comprising three concurrent tone pip streams. The task was to indicate, after a delay, whether a probe stream was present in the just-heard scene. We present three key findings: First, behavioural performance revealed faster responses in correct versus incorrect trials as well as in ‘probe present’ versus ‘probe absent’ trials, consistent with ASTM search. Second, successful compared with unsuccessful ASTM performance was associated with a significant enhancement of event-related fields and oscillatory activity in the theta, alpha and beta frequency ranges. This extends previous findings of an overall increase of persistent activity during short-term memory performance. Third, using distributed source modelling, we found these effects to be confined mostly to sensory areas during encoding, presumably related to ASTM contents per se. Parietal and frontal sources then became relevant during the maintenance stage, indicating that effective STM operation also relies on ongoing inhibitory processes suppressing task-irrelevant information. In summary, our results deliver a detailed account of the neural patterns that differentiate successful from unsuccessful ASTM performance in the context of a complex, multi-object auditory scene.

KEYWORDS

auditory scene analysis; MEG; performance; salience theta, alpha, beta, oscillations

Abbreviations: ASTM, auditory short-term memory; DLPFC, dorsolateral prefrontal cortex; DSS, denoising source separation; EEG, electroencephalography; ERF, event-related field; fMRI, functional magnetic resonance imaging; FN, false negatives; FP, false positives; MEG, magnetoencephalography; PSD, power-spectral density; RMS, root mean square; RT, response time; STM, short-term memory; TN, true negatives; TP, true positives.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *European Journal of Neuroscience* published by Federation of European Neuroscience Societies and John Wiley & Sons Ltd.

1 | INTRODUCTION

Auditory short-term memory (ASTM) refers to the ability to maintain auditory information over durations of a few seconds and is an important requirement for the analysis of acoustic scenes (Wilsch & Obleser, 2016). As such, it is involved in basic, yet vital tasks such as detecting potentially threatening changes in our acoustic environment, as well as highly complex functions like verbal communication or musical performance.

Recent years have shown an increasing interest in human ASTM (Arnott et al., 2005; Buchsbaum et al., 2005; Cappotto et al., 2021; Kumar et al., 2016; Lim et al., 2022; Noyce et al., 2022; Papagno et al., 2017), particularly in the context of electrophysiological oscillatory correlates (Kaiser, 2015; Kumar et al., 2021; Weisz et al., 2020; Wilsch & Obleser, 2016). Although previous work has predominantly focused on contrasting low- and high-load ASTM tasks, here, we ask what patterns of brain activity distinguish successful versus unsuccessful ASTM performance while the task, and stimuli presented to participants, remain fixed.

Accumulating work has identified core components of ASTM, primarily comprising auditory cortex as well as regions in the prefrontal (Albouy et al., 2017; Huang et al., 2013; Kumar et al., 2021, 2016) and parietal cortices (Koelsch et al., 2009; Leung & Alain, 2011). It is suggested that while the contents of ASTM are represented in sensory areas, prefrontal regions are performing executive control functions coordinating the encoding, maintenance and retrieval of stored information (Deutsch et al., 2023; Spitzer & Blankenburg, 2012; Sreenivasan et al., 2014).

Functionally, a common observation is of increases in event-related EEG and MEG activity during high

compared with low ASTM load maintenance (Grimault et al., 2014; Huang et al., 2016; Nolden et al., 2013), which is suggested to originate from elevated firing of neurons in memory relevant areas (Huang et al., 2016).

Similar to observations in the context of visual short-term memory (Ratcliffe et al., 2022; Roux & Uhlhaas, 2014), oscillatory neural activity (as measured with EEG and MEG) is suggested to play an important role in encoding and maintaining information in ASTM (Kaiser, 2015; Wilsch & Obleser, 2016; Weisz et al., 2020; Kumar et al., 2021; Figure 1). In particular, increased theta power (4–7 Hz) has been observed in sensory areas during encoding of ASTM, suggested to reflect chunking of incoming information (Teng et al., 2017), as well as in parietal cortex during maintenance, indicative of ASTM manipulation (Albouy et al., 2017, 2022). Likewise, increase in alpha-band activity (8–13 Hz) over posterior areas has been repeatedly associated with ASTM tasks, interpreted to reflect the suppression of task-irrelevant, potentially distracting information (Obleser et al., 2012; Wisniewski et al., 2017; Wöstmann et al., 2017). Further, enhancements of beta-band power (14–30 Hz) in temporal regions have been observed with increasing levels of ASTM load (Leiberg et al., 2006) and associated with binding of auditory short-term memory contents in sensory areas (Weiss & Mueller, 2012).

Overall, these findings suggest a rich spatio-spectral activity pattern associated with ASTM, in which different functional aspects of the task are carried out by localized activity in specific frequency bands.

However, although recent studies have investigated how these oscillatory signatures are affected by task requirements (e.g. Albouy et al., 2017; Dimakopoulos et al., 2022; Wöstmann et al., 2017) or STM load (Leiberg

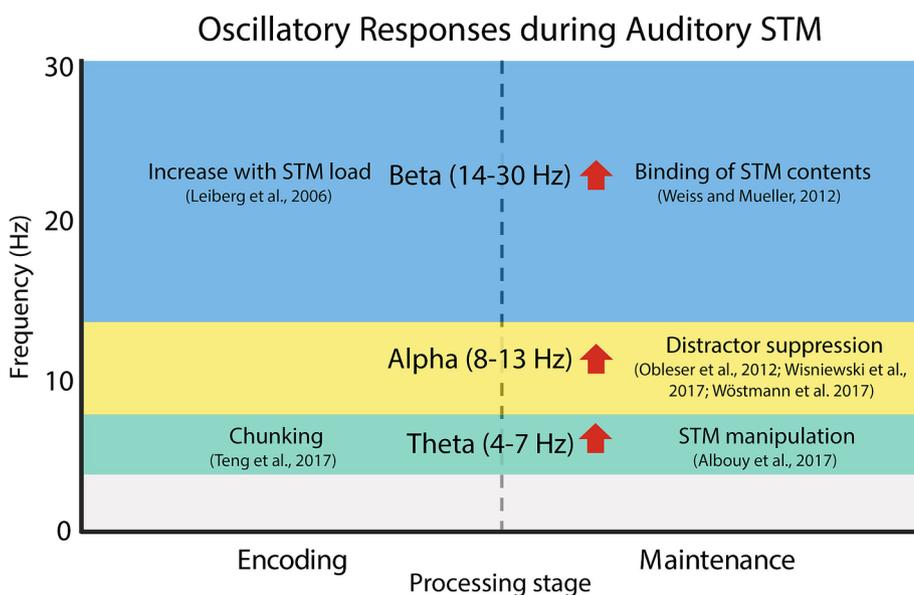


FIGURE 1 Overview of common oscillatory responses during auditory STM tasks reported in previous studies.

et al., 2006; Obleser et al., 2012; Kosachenko et al., 2023), it is currently unknown how they might relate to successful versus unsuccessful ASTM performance where task load per se is constant across trials (for examples in the visual domain, see Lenartowicz et al., 2021; Mapelli & Özkurt, 2019; Proskovec et al., 2018).

Here, using MEG, we examined brain activity associated with (ultimately) correct versus incorrect responses in a challenging short-term memory task. Stimuli (Figure 2) were artificial ‘acoustic scenes’ consisting of three concurrent auditory ‘objects’, akin to a mix of auditory sources in everyday listening environments. Listeners were required to memorize the sounds and, following a silent retention interval, determine whether a probe sound was present in the just-heard scene. We explored the evoked and oscillatory activity during encoding and maintenance, identifying a distributed pattern of neural dynamics, which differentiates subsequently successful versus failed memory performance.

2 | METHODS

2.1 | Participants

Twenty-eight right-handed (Edinburgh Handedness Inventory; Oldfield, 1971) paid participants took part in the study. All reported no neurological illness or hearing deficits. One participant was excluded because of not

being able to properly perform the task, despite repeated training. Four other participants were excluded because of exceptionally poor behavioural performance (mean response times of >2000 ms compared with 1007 ms for the rest of the participants). Therefore, data from 23 participants (16 females, mean age 23.6 years) are presented. The study was conducted in accordance with the research ethics committee of the University College London, and all participants provided written informed consent.

2.2 | MEG recordings

MEG was recorded continuously (600-Hz sampling rate; 100-Hz hardware low-pass filter) using a CTF-275 MEG system, with 274 functioning axial gradiometers arranged in a helmet-shaped array. Electrical coils were attached to three anatomical fiducial points (nasion and left and right pre-auricular), to continuously monitor the position of each participant’s head with respect to the MEG sensors. During the recording, participants were seated in an upright position in a magnetically shielded room. Sounds were presented binaurally via insert earphones (E-ARTONE 3A 10, Etymotic Research), with sound intensity adjusted individually to a comfortable level. To minimize eye movements, a black fixation cross on a grey background was presented throughout the experiment via a screen located about 60 cm in front of the participant’s eyes.



FIGURE 2 Schematic trial structure. Each trial started with an ‘encoding’ stage lasting 2500 ms, during which an auditory scene consisting of three concurrent streams was presented. Stream rates were chosen from a set of three values: 3, 7 and 19 Hz. Carrier frequencies (indicated by the red, green and blue colours) were randomly assigned to each stream on a trial-by-trial basis (taken from a pool of 11 values). The encoding stage was followed by a 2000-ms silent maintenance stage. After this, a probe stimulus was presented for 2000 ms. This probe was either identical to one of the streams presented during encoding (*same* trial; top panel), or a combination of one of the assigned carrier frequencies with one of the two stimulus rates with which it was not originally paired (*different* trial; bottom panel). For instance, in the example plotted here, the green carrier frequency, originally presented (in the encoding stage) at 3 Hz, is now paired with the 7-Hz rate. Participants were asked to provide a speeded response, indicating whether the probe stimulus was *same* or *different*, using one of two buttons.

2.3 | Task and stimuli

At the beginning of each trial, participants were presented with a 2500 ms long auditory 'scene' consisting of three simultaneous streams of repetitive pure tones (Figure 2; 'encoding' stage). Each stream had a unique frequency and stimulation rate. Rates were 3, 7 and 19 Hz, defined by the duration of a single tone and the following inter-tone interval (each 166.5, 71.45 and 26.15 ms, respectively, i.e. SOAs of 333, 142.9 and 52.3 ms, respectively). Each tone had a rise and fall time of 5 ms. Frequencies were randomly chosen from a pool of 11 logarithmically spaced values (200, 292, 405.1, 542, 708.4, 910.7, 1157, 1456, 1820, 2262 and 2799 Hz). As a consequence, the rate content of the scene was fixed but, on each trial, these were paired with different carrier frequencies. The 'encoding' stage was followed by a silent interval of 2000-ms duration, during which participants had to maintain the auditory scene in memory ('maintenance' stage). After the gap, a single auditory stream was presented as a probe for 2000 ms. In 50% of trials, this probe was identical to one of the three initially presented streams (*same* trial), and in 50% of trials, it consisted of one of the three presented carrier frequencies combined with one of the presented rates, that it was not previously paired with (*different* trial). Each participant completed 336 trials in total (56 per probe rate for both the *same* and *different* conditions).

Participants were instructed to provide a speeded response via one of two buttons with their right hand, to indicate whether the probe belonged to the *same* or *different* category. After every button press, visual feedback (lasting 300 ms) was given on whether the answer was correct or incorrect. Following the offset of the feedback was an inter-trial interval varying uniformly between 700 and 2000 ms (in steps of 100 ms).

The fact that the probe consisted of a previously presented frequency and rate rendered it always partially 'familiar', which increased task difficulty. To perform correctly, participants had to accurately bind the appropriate frequencies and rates during the 'encoding' stage and maintain this information in memory until probed. Pilot experiments revealed that performance accuracy is on average at around 70%, suggesting it is a sufficiently challenging task to probe short-term memory and failures thereof (see also Figure 3b).

2.4 | Behavioural analysis

Prior to the statistical analysis, outlier trials with RTs deviating more than 2.5 SDs from the mean were excluded per participant and condition. Mean percentage of true positives (TP), true negatives (TN), false positives

(FP) and false negatives (FN), as well as response times (RTs), were computed separately for each condition. For statistical analysis, performance and RT data were each subjected to a repeated-measures analysis of variance (RMANOVA), including the within-participant variables outcome (TP, TN, FP, FN) and rate (3, 7, 19 Hz).

2.5 | MEG analysis

2.5.1 | Preprocessing

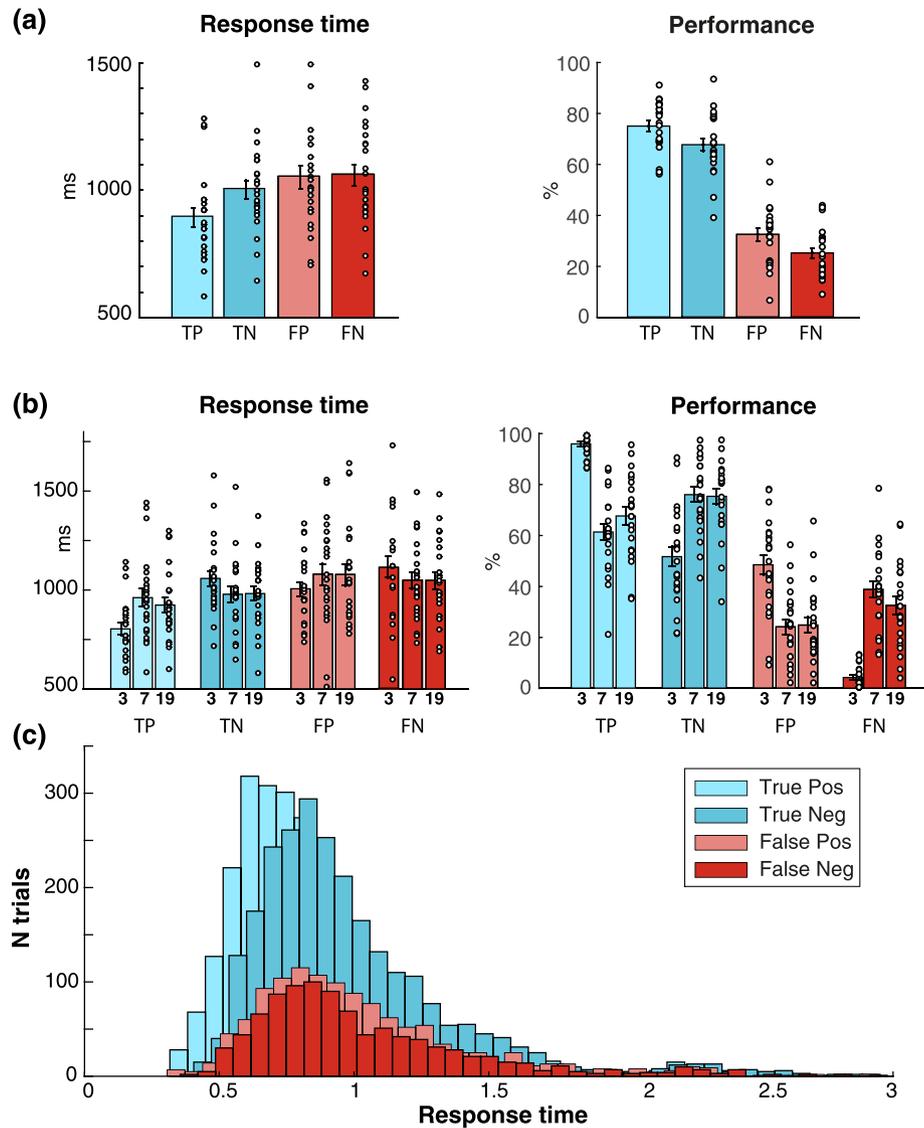
Off-line, MEG was processed using Matlab (The MathWorks Inc., Natick, MA, USA) and FieldTrip (Oostenveld et al., 2011). First, data were epoched from -1000 to 6000 ms around the onset of auditory scenes. Thus, each epoch consists of a 1000-ms 'prestimulus' stage, a 2500-ms 'encoding' stage, a 2000-ms 'maintenance' stage and a 2000-ms 'probe' stage. Trials with power that deviated from the mean by more than twice the standard deviation (across all sensors and time points) were deemed outliers and discarded from further analyses. Artifacts such as slow drifts and 50 Hz line noise were removed using DSS (denoising source separation; de Cheveigné & Parra, 2014), and epochs were baseline corrected relative to a 500-ms prestimulus interval.

All data were then separated into *correct* (TP, TN) and *incorrect* (FP, FN) conditions, based on the behavioural response. Within each participant, the number of *correct* trials was larger than that of *incorrect* trials. As the signal-to-noise ratio increases with the number of trials used in averaging procedures such as the presently investigated event-related fields and event-related oscillations, this difference in the number of trials between conditions potentially confounds their comparison. To ensure similar signal-to-noise ratio in both conditions, we selected a subset of *correct* trials for each participant, equal in number to the *incorrect* trials. Because outlier trials (± 2.5 SD away from the mean) at both tails of the RT distribution were already discarded beforehand, RTs can be used as a marker for rating confidence, as well as to disentangle true *correct* responses from guesses (Ratcliff, 1993; 'true' *correct* responses are expected to be faster than 'guess' responses). Thus, we equated the numbers of trials by ordering all *correct* trials according to response time and discarding the surplus trials from the slow tail of the distribution.

2.5.2 | Time domain analysis

First, we investigated potential differences between event-related fields (ERFs) during *correct* versus *incorrect* trials. As a measure of global instantaneous power, for

FIGURE 3 Behavioural results. (a) Data pooled across probe rates, separately for trials with true positive (TP), true negative (TN), false positive (FP) and false negative (FN) outcomes. Left: average response times. Right: performance (% of *same* [TP + FN] and *different* [TN + FP] trials). Circles indicate individual participants; error bars represent SEM. (b) Same as in a, with separate bars for trials with probes of 3, 7 and 19 Hz stimulus rates. (c) Probability density function of RTs separately for TP, TN, FP and FN responses.



each participant, the root mean square (RMS) amplitude across all channels was calculated separately for the *correct* and *incorrect* trials. To identify time windows where *correct* trials differed from *incorrect* trials, a bootstrapping procedure (10,000 iterations) was performed on the difference in RMS time course between the two conditions. On each iteration, the data were resampled by choosing random sets of 23 participants (with replacement). For each time point, we computed the mean response across participants, and any time point yielding above 95% of iterations in the same direction (positive or negative) was deemed significant (Figure 4a). This analysis yielded widespread significant differences throughout the encoding and maintenance stages (Figure 4a). To identify the channels that contribute to this global effect, a cluster-based permutation procedure (Maris & Oostenveld, 2007; 1000 permutations, min. = 3 channels) was used. The ERF cluster correction was performed on the spatial (i.e. channels) dimension only, separately for the

encoding and maintenance stages. Finally, cortical sources for any observed differences were estimated via the multiple-sparse priors (MSP) method implemented in SPM12. Participant-specific forward models were computed using a Single Shell model and sensor positions projected onto an MNI space template brain by minimizing the sum of squared differences between the digitized fiducials and the MNI template. Data for the encoding and maintenance stages (i.e. from 0 to 4500 ms following stimulus onset) were inverted conjointly and pooled across *correct* and *incorrect* trials. Grand average source solutions were constrained by the restriction that included sources have to be present in each participant, which has been shown to improve group-level statistical power (Litvak & Friston, 2008). After inversion, source estimates were averaged over intervals of significant differences between *correct* and *incorrect* trials observed via the RMS analysis, projected to a 3D source space (MNI space, constrained to gray matter) and smoothed (8-mm

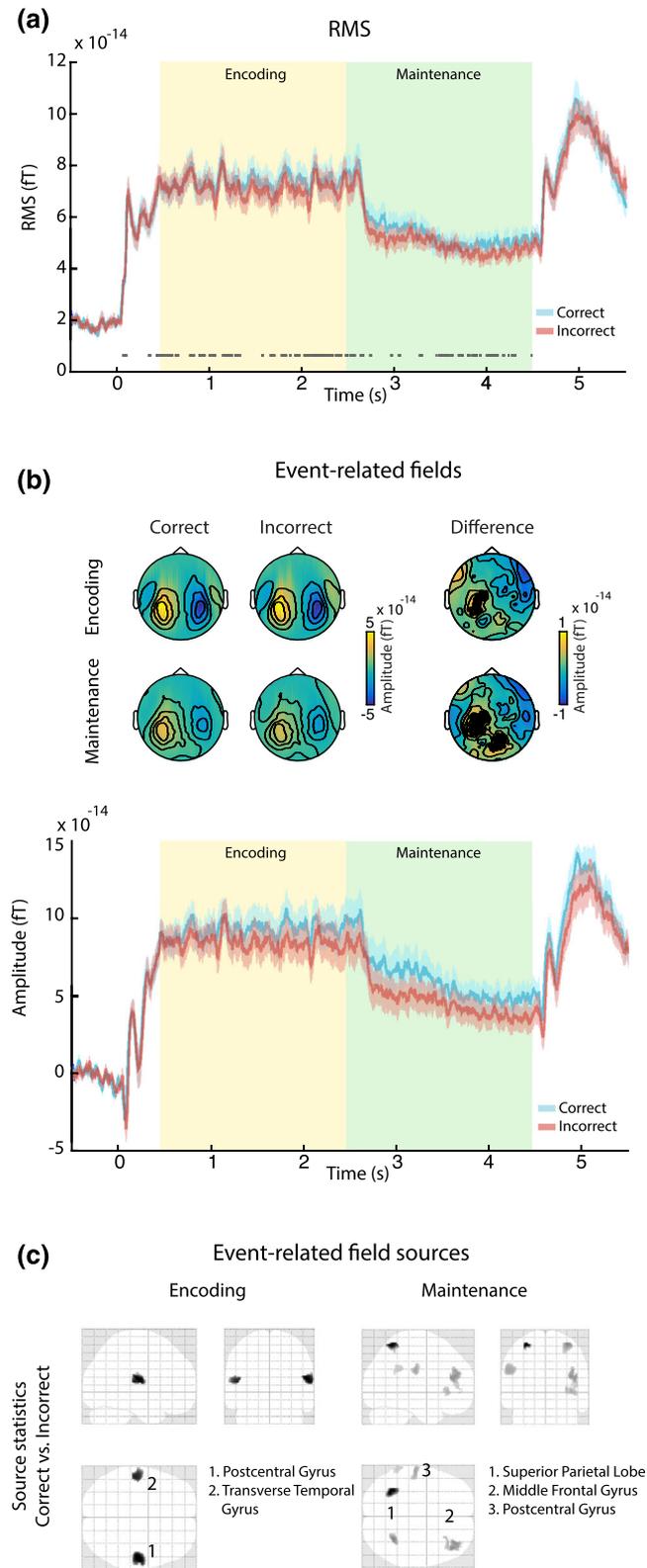


FIGURE 4 Evoked responses across the entire experimental trial. (a) Global instantaneous power (root-mean-square activity across all channels), for *correct* (blue) and *incorrect* (red) trials. The horizontal grey line indicates stages of significant differences as observed via a bootstrap permutation analysis. Yellow and green shaded areas indicate encoding and maintenance stages, respectively. (b) Top: scalp topographies for *correct* (left) and *incorrect* (middle) trials, as well as their difference (right). Black dots indicate channels exhibiting significant differences. Bottom: event-related field traces for *correct* (blue) and *incorrect* (red) trials (mean over channels on the left). Yellow and green shaded areas indicate encoding and maintenance stages, respectively. (c) Source-level difference between *correct* and *incorrect* trials, separately for the encoding (left) and maintenance stage (right). Numbers indicate significant clusters ($p < 0.05$), sorted by p -values from low to high. Insets indicate anatomical areas containing the peak voxel of each cluster.

full width at half maximum [FWHM] Gaussian smoothing kernel) to create images of source activity for each participant. Given that the goal of source reconstruction was to localize the neural generators of sensor-space effects previously identified as significant, statistical maps of source activity are displayed with uncorrected voxel-wise thresholds (Gross et al., 2013).

2.5.3 | Analysis of oscillatory power

To study differences in oscillatory neural activity between *correct* and *incorrect* trials, all data were transformed into time–frequency domain by applying a sliding window Fourier transform with a single Hanning taper. Power at frequencies from 1 to 30 Hz was computed in 0.5-Hz steps, using a fixed time window (400 ms) and a fixed frequency smoothing (1 Hz). For baseline correction, oscillatory power in each trial was normalized relative to the interval from -400 to -200 ms (indicating the centre of the sliding time window, thus including data from -600 to 0 ms prior to the beginning of the trial) before the onset of the auditory stimulus.

For statistical comparison between *correct* and *incorrect* trials, cluster-based permutation procedures were employed (Maris & Oostenveld, 2007; 1000 permutations, $\min. = 3$ channels), with cluster correction in the temporal and spatial (i.e. channel) domain. Statistics were calculated separately for the ‘prestimulus’, ‘encoding’ and ‘maintenance’ stages, within the delta (1–4 Hz), theta (4–7 Hz), alpha (8–13 Hz) and low beta (14–20 Hz) ranges.

Similar to the analysis of ERF responses, we estimated the sources underlying the effects in oscillatory power and phase observed at the sensor level. Here, we used the minimum-norm method (Hämäläinen &

Ilmoniemi, 1994) implemented in SPM 12, which has been shown to be suitable for the localization of induced oscillatory activity (Halder et al., 2019). Data for the encoding stage (from 500 to 2500 ms, i.e. excluding the evoked response from stimulus onset) and maintenance stage (from 2500 to 4500 ms) were first inverted separately and independently for the theta (4–7 Hz), alpha (8–14 Hz) and lower beta (15–20 Hz) frequency ranges. Source solutions were pooled across *correct* and *incorrect* trials and constrained to be consistent across participants. Significant effects from sensor-space were localized within the brain (in MNI space, constrained to grey matter) after summarizing source power in the specific time–frequency range of interest for each participant using a Morlet wavelet projector (Friston et al., 2006).

2.5.4 | Power-spectral density analysis

In addition to effects on ongoing endogenous neural oscillations, the presentation of regular auditory inputs has been shown to cause entrainment of neural activity (Falk et al., 2017; Kösem et al., 2018; van Bree et al., 2020). As an indication of successful versus unsuccessful ASTM performance, we speculated that the strength of entrainment during encoding reflects the quality of sensory processing, and its continuation during active maintenance might correlate with the fidelity of ASTM representations and subsequent performance. As a measure of the fidelity of encoding and maintenance of individual sound sources, we computed power-spectral density (PSD) values separately for each participant, condition (*correct* vs. *incorrect*) and the encoding and maintenance stage. To maximize reproducibility of the evoked response across trials, we applied DSS (de Cheveigné & Parra, 2014; de Cheveigné & Simon, 2008). For each participant, the first two DSS components (i.e., the two ‘most reproducible’ components; determined from the encoding stage, i.e. from 0 to 2500 ms relative to scene onset) were retained and used to project the data from the entire trial back into sensor-space. Similar analyses, in which only the first, the first three, or the first twelve components were retained, yielded the same pattern of results and are not reported here. We additionally analysed the data from the maintenance stage without prior application of DSS, because it can be argued that entrained activity might show increased temporal variability across trials and participants after stimulus presentation has ended. Thus, only retaining evoked components might be detrimental to observing entrainment effects during the maintenance stage. However, the results were highly similar to those following DSS and are thus not reported.

To increase the spectral resolution of the PSD analysis, we next concatenated all trials into a single time series, separately for each participant, condition and stage. This procedure results in a maximum frequency resolution of 0.02 Hz. PSD values were computed from 1 to 21 Hz by applying a fast Fourier transform with a single Hanning taper. The power at each frequency bin was normalized by dividing it by the mean power across the 3rd to 20th neighbouring bin above and below. For each participant and frequency, we then selected the subset of 10 channels that showed the largest normalized PSD across both the *correct* and *incorrect* conditions. Because no stimulus-related PSD peaks were observed in the maintenance stage, subsequent analyses focused on the encoding interval. For statistical comparison, we calculated an ANOVA using the variables Condition (*correct* vs. *incorrect*) and rate (3, 7, 19 Hz). Finally, we estimated the sources underlying the phase-locked responses at the stimulation frequencies of 3, 7 and 19 Hz, using the minimum-norm method. Data for the encoding stage (from 500 to 2500 ms, i.e. excluding the evoked response from stimulus onset) were inverted independently for 3, 7 and 19 Hz but pooled across *correct* and *incorrect* trials. After summarizing source power in each time–frequency range of interest for each participant using a Morlet wavelet projector, source statistics maps were computed by testing the activity in each frequency band against 0.

3 | RESULTS

3.1 | Behavioural responses suggest that search of STM is affected by confidence

Figure 3 shows response times (RTs) and performance for each behavioural outcome, both pooled across and separately for the three probe rates (Figure 3a,b, respectively).

Statistical analysis was performed via a two-way repeated measures ANOVA using the variables outcome (TP, TN, FP, FN) and rate (3, 7, 19 Hz).

For performance (Figure 3a,b, left), we observed a main effect of outcome ($F_{3,66} = 89.65$, $p < .001$), due to participants producing more *correct* (TP and TN) than *incorrect* (FP, FN) responses.

For RTs, we observed a main effect of outcome ($F_{3,66} = 20.26$, $p < .001$). As evident from Figure 3a, participants were slower in TN than in TP trials ($t_{22} = -6.18$, $p < .001$), consistent with the notion that TN responses are likely to be preceded by an exhaustive search of ASTM contents causing longer RTs, whereas TP responses are not. Participants were also slower in FP than in TP trials ($t_{22} = -10.36$, $p < .001$; despite responding ‘same’ in both cases) and faster in TN than in FN

trials ($t_{22} = -2.34$, $p = .029$; despite responding ‘different’ in both cases). This suggests that some aspect of the cognitive processes underlying the ASTM comparison (e.g. the STM representation, the sensory representation of the probe stimulus and the decision process) was prolonged in *incorrect* trials, on average leading to delayed responses. This is also supported by the different distribution of RTs for TP and TN compared with FP and FN trials in the probability density function in Figure 3c).

In summary, the behavioural results indicate that the present task is well suited to tap into auditory short-term memory. Participants performed above chance and gave significantly more *correct* (i.e. TP and TN) than *incorrect* (i.e. FP and FN) responses in all conditions. Further, the fact that RTs were faster for *correct* compared with *incorrect* responses already points towards a potentially facilitated STM comparison- and decision-making process during *correct* trials.

3.2 | Behavioural responses reveal perceptual dominance of the 3-Hz rate

Examining performance whilst including rate as a factor (Figure 3b), we further observed a significant interaction between the variables *outcome* and *rate* ($F_{6,132} = 41.07$, $p < .001$). This interaction was driven by the 3-Hz condition (Figure 3b, left), with follow-up *t*-tests demonstrating more TP and FP responses for 3 Hz compared with 7 Hz ($t_{22} = 10.32$, $p < .001$; $t_{22} = 5.78$, $p < .001$) and 19 Hz trials ($t_{22} = 8.41$, $p < .001$; $t_{22} = 5.17$, $p < .001$), and fewer TN and FN responses for 3 Hz compared with 7 Hz ($t_{22} = -5.78$, $p < .001$; $t_{22} = -10.32$, $p < .001$) and 19 Hz trials ($t_{22} = -5.17$, $p < .001$; $t_{22} = -8.41$, $p < .001$). This suggests a tendency to respond with ‘same’ in trials, in which a 3-Hz probe was presented, in both ‘same’ and ‘different’ trials (see Figure 2).

For reaction times (Figure 3b), an ANOVA similarly revealed a significant interaction for RTs between the variables *outcome* and *rate* ($F_{6,132} = 7.36$, $p < .001$). Looking at Figure 3b, this interaction was clearly driven by the 3-Hz probe condition. Follow-up *t*-tests confirmed that RTs to 3-Hz probes were significantly faster than those to 7- and 19-Hz probes for TP ($t_{22} = -7.73$, $p < .001$; $t_{22} = -6.67$, $p < .001$) and FP outcomes ($t_{22} = 3.20$, $p = .004$; $t_{22} = 3.78$, $p = .001$), and slower than those to 7- and 19-Hz probes for TN ($t_{22} = -2.2$, $p = .039$; $t_{22} = -2.45$, $p = .023$) but not for FN outcomes ($ps > .20$). In other words, 3-Hz probes led to faster RTs whenever participants deemed the probe the same as the memorized stimulus, but to slower RTs when they correctly identified it as different.

Overall, this pattern of results suggests that the 3-Hz rate is perceptually dominant and likely unbound from its carrier frequency such that participants are biased to respond that the probe was present even if it is presented with the incorrect carrier frequency (i.e. it is a ‘different’ trial).

3.3 | MEG data reveal sustained increases in power during correct compared with incorrect trials

Figure 4 displays the group-level event-related field (ERF) results. The presentation of the ‘scene’ stimuli evoked increased sustained activity that persisted for the duration of the encoding stage. This was followed by a drop in sustained power, during the maintenance stage. Activity associated with the onset of the probe, and subsequent button press, is seen during the retrieval stage. In the following, we focus on understanding how activity during the encoding and maintenance stages (i.e. before the probe has been presented) differs between *correct* and *incorrect* trials.

We observed patches of significantly larger instantaneous power (RMS across all channels) for *correct* compared with *incorrect* trials, as determined via a bootstrapping procedure, throughout most of the encoding and maintenance stages (Figure 4a). This difference is also evident in the fieldmaps (Figure 4b), whose dipolar topography suggests an underlying auditory source. A cluster-based permutation analysis identified a left-hemispheric channel cluster during the encoding stage ($p = .037$; 21 channels) and a left-hemispheric and centrally located cluster during the maintenance stage ($p = .011$; 33 channels), indicated in Figure 4b. Consistent with the RMS results, event-related field data from those channels (Figure 4b) show higher activity during trials that were ultimately *correct*. Strikingly, increased activity for correct trials persists in the maintenance stage, even when sound is no longer presented.

3.4 | Differences between correct and incorrect trials are restricted to auditory areas during the encoding stage, shifting to non-auditory areas during the maintenance stage

Source inversion based on multiple-sparse priors (Figure 4c) suggests that the above-observed differences arise predominantly in bilateral temporal areas during the encoding stage, in line with stronger auditory cortex activation during *correct* compared with *incorrect* trials.

During the maintenance stage, we found bilateral differences in parietal areas as well as prefrontal and temporal areas. Together, these results indicate that neural activity differentiating *correct* from *incorrect* trials during the encoding stage is predominantly restricted to auditory sensory areas. Subsequently, during short-term maintenance, this shifts to other, non-auditory areas.

3.5 | No differences in power-spectral density between *correct* and *incorrect* trials

Figure 5 illustrates the results of the PSD analysis. During the encoding stage (Figure 5a), distinct peaks are visible at the stimulation frequencies of 3, 7 and 19 Hz and their harmonics. However, no representation of the stimulus frequencies was visible during the maintenance stage (not shown in Figure 5). Focusing on the encoding stage,

Figure 5B displays the peak PSD values for each rate and outcome conditions. The ANOVA using the variables outcome (*correct* vs. *incorrect*) and stimulus rate (3, 7, 19 Hz) yielded a significant main effect of stimulus rate ($F_{2,44} = 12.35$, $p < .001$). Previous work (Elhilali et al., 2009; Huang & Elhilali, 2020) has revealed large-scale effects of attention on PSD measures of auditory tracking such that attended sounds are associated with increased sustained (steady-state) neural representation. The absence of a difference between *correct* and *incorrect* trials here suggests that success or failure in the present task is not driven by variable attentive tracking of the scene components. Source level estimation of the PSD at 3 Hz (tested against 0; all displayed voxels are $p < 0.05$) shows extensive activation with a bilateral peak in temporal regions, in line with auditory sensory activation (Figure 5c). Source estimations of the 7- and 19-Hz components yielded highly similar statistical maps (not shown).

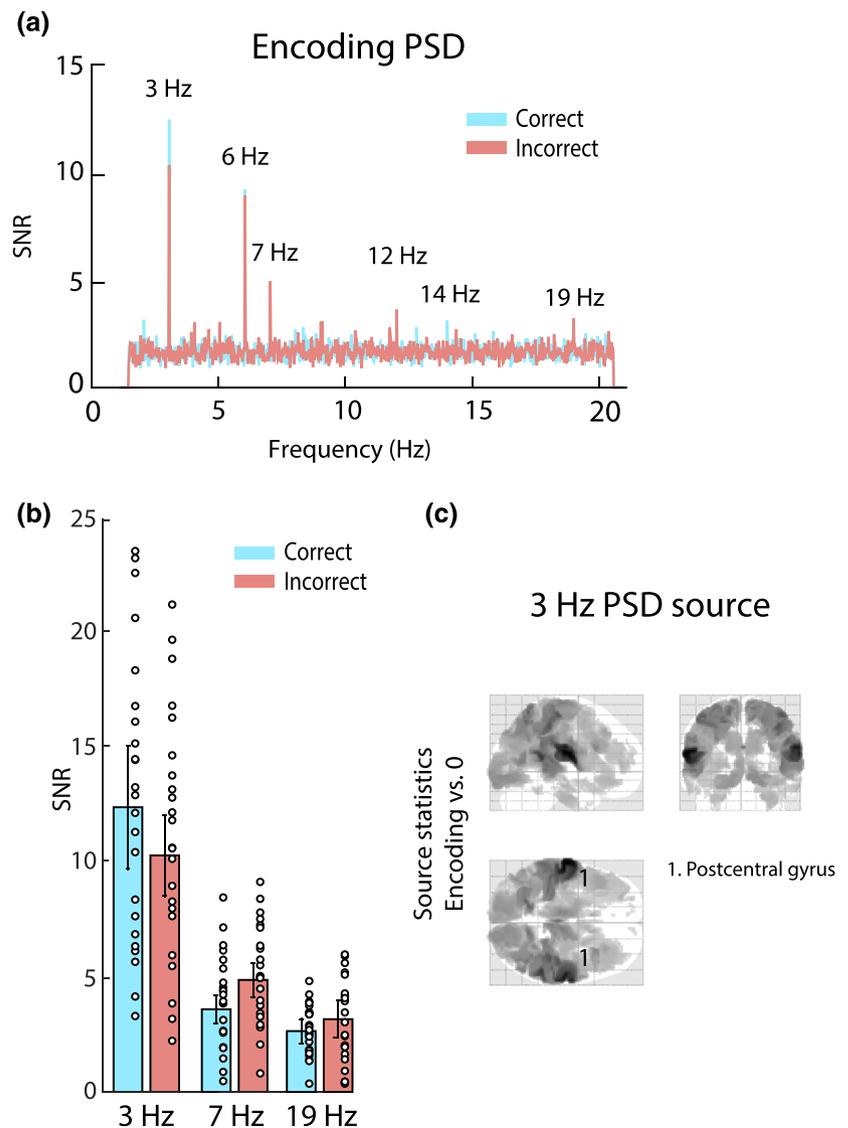


FIGURE 5 Power-spectral density (PSD). (a) Normalized PSD during the encoding phase (no noticeable peaks were found during the maintenance stage). Blue represents data from *correct* trials, and red data from *incorrect* trials. Numbers highlight peaks at stimulation frequencies as well as harmonics. (b) PSD values for the three rates (3, 7, 19 Hz) for *correct* (blue) and *incorrect* (red) trials. Circles indicate individual participants, error bars represent SEM. (c) Source-level difference in oscillatory power at 3 Hz, between *correct* and *incorrect* trials during the encoding time window. The number indicates the significant cluster ($p < 0.05$), and the inset indicates the anatomical area containing the cluster's peak voxel.

3.6 | Widespread increases in oscillatory alpha, beta and theta power during *correct* trials

Next, we looked into differences between *correct* and *incorrect* trials, separately within the alpha, beta and theta frequency bands. Figure 6 displays time-frequency planes and source inversions of the observed effects in the alpha band. At scalp level (Figure 6a), distinct increases in power relative to the baseline can be seen, which appear to be stronger in the *correct* compared with the *incorrect* condition. Using cluster-corrected permutation tests in time and space, we observed significantly larger power for *correct* versus *incorrect* trials, during both the encoding ($p = .014$, 1500–2500 ms, 137 channels) and maintenance stages ($p = .004$, 2500–3650 ms, 117 channels). Source inversion of the respective time-frequency windows (Figure 6b) revealed increased activity for *correct* versus *incorrect* trials in left parietal areas during the encoding stage and in a more widespread parieto-occipital area as well as a frontal region during the maintenance stage.

Figure 7 shows time-frequency results and source inversions of the observed effects in the theta and beta bands. In the theta band, we found significantly higher power in *correct* compared with *incorrect* trials, during the second half of the encoding stage ($p = .031$, 1560–2410 ms, 93 channels). Source inversion of this time-frequency window yielded increased activity in left temporal regions for the *correct* versus the *incorrect* condition.

In the beta band, we again found higher power in *correct* versus *incorrect* trials, however during an earlier time window ($p = .02$, 646–1281 ms, 80 channels) in the encoding stage, as well as during two extended time windows in the maintenance stage ($p = .005$, 2632–3340 ms, 152 channels and $p = .007$, 3766–4510 ms, 140 channels). For the effect during the encoding time window, source inversion revealed a likely source in left temporal areas, similar to the theta band. For the source inversion of the effect during maintenance, we inverted a single time stage spanning the time windows of both clusters. Here, we found increased activity for *correct* versus *incorrect* trials in widespread left-temporal and occipital areas,

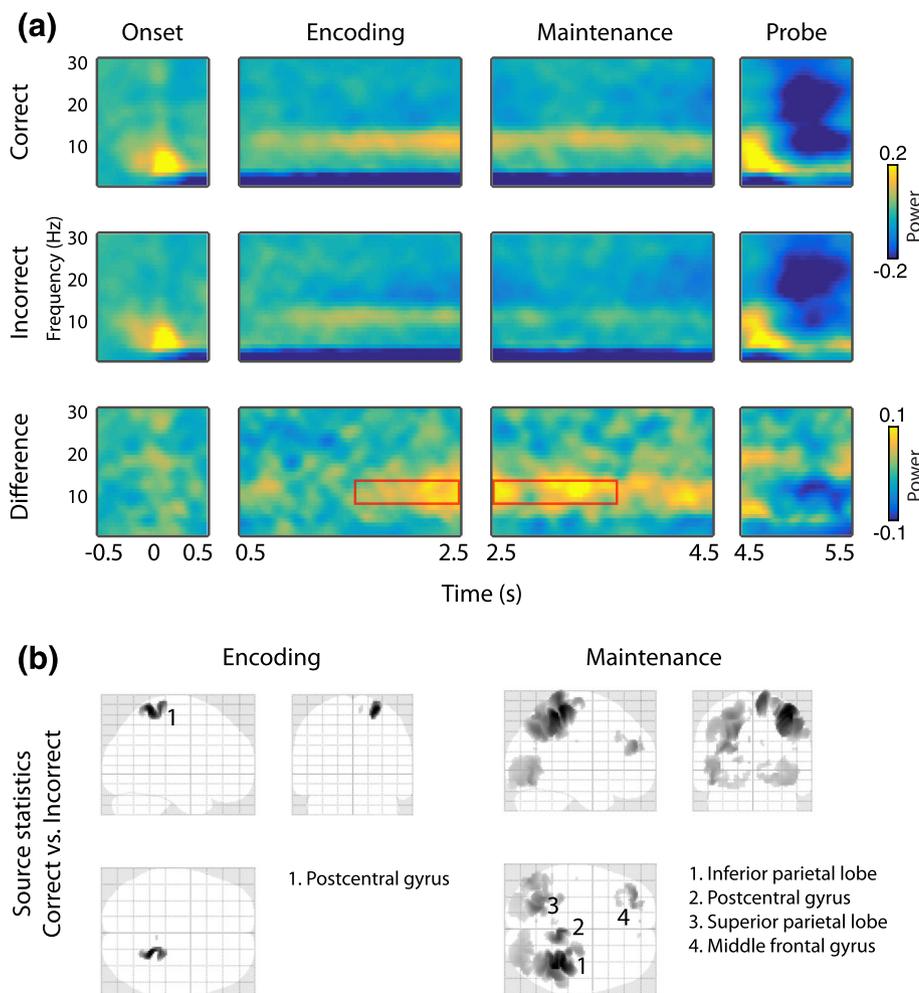


FIGURE 6 Results of alpha-band (8–14 Hz) time-frequency analysis. (a) Time-frequency representations of oscillatory power during the stimulus onset, encoding, maintenance and probe phase of the experiment. Data are shown separately for *correct* (top) and *incorrect* (middle) trials, as well as their difference (bottom). Red boxes indicate time windows exhibiting significant differences, as tested via a cluster-based permutation procedure. The plotted data represent the mean across all sensors belonging to the significant cluster. (b) Source-level difference between *correct* and *incorrect* trials during the time windows determined via scalp analysis, separately for the encoding (left) and maintenance stage (right). Numbers indicate significant clusters ($p < 0.05$), sorted by p -values from low to high. Insets indicate anatomical areas containing the peak voxel of each cluster.

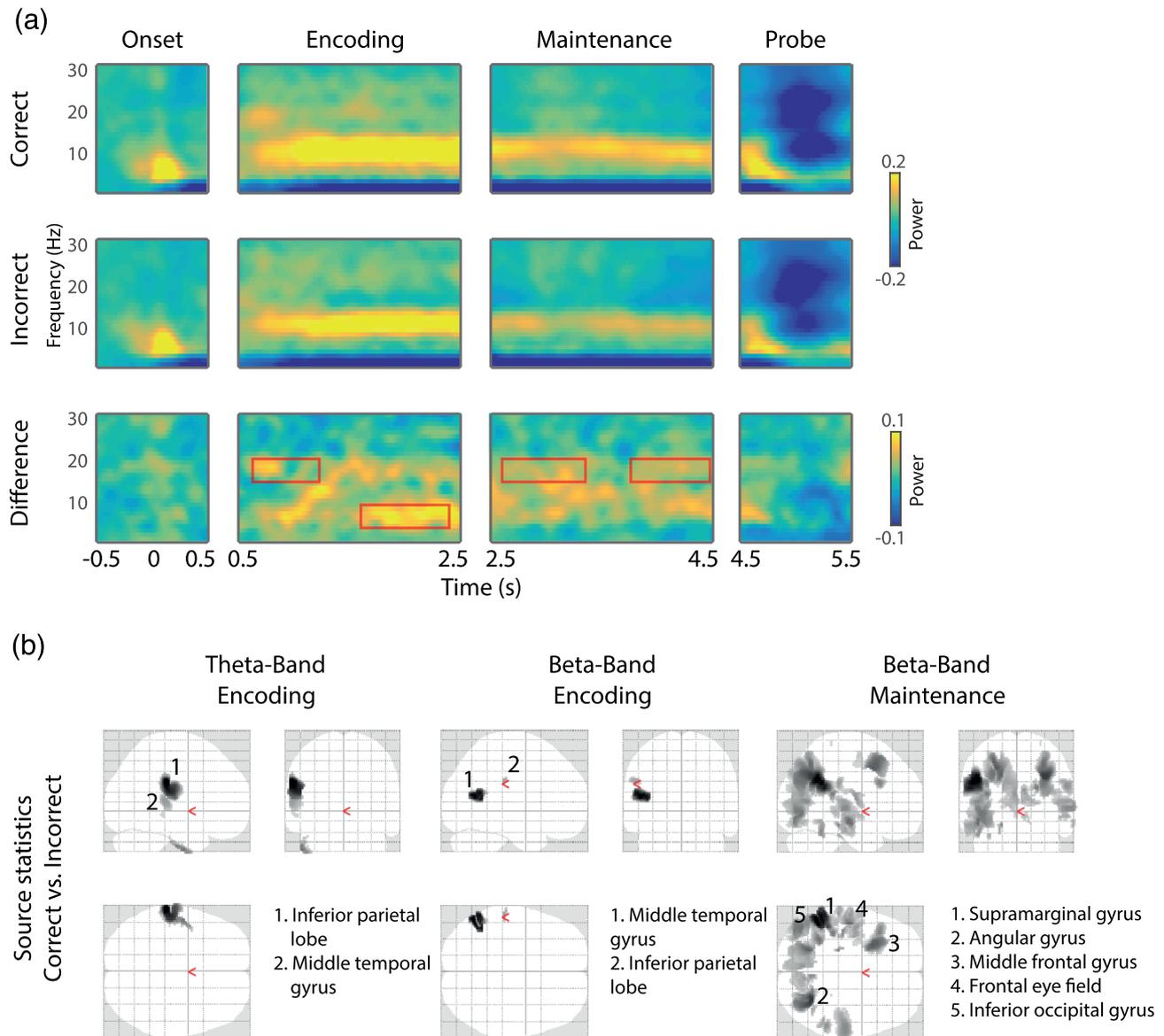


FIGURE 7 Results of theta- (4–7 Hz) and beta-band (15–20 Hz) time–frequency analysis. (a) Time–frequency representations of oscillatory power during the stimulus onset, encoding, maintenance and probe phase of the experiment. Data are shown separately for *correct* (top) and *incorrect* (middle) trials, as well as their difference (bottom). Red boxes indicate time windows exhibiting significant differences, as tested via a cluster-based permutation procedure. The plotted data represent the mean across all sensors belonging to the significant cluster. (b) Source-level difference between *correct* and *incorrect* trials during the time windows determined via scalp analysis, separately for the theta-band encoding stage (left), and the beta-band encoding (middle) and maintenance stages (right). Numbers indicate significant clusters ($p < 0.05$), sorted by p -values from low to high. Insets indicate anatomical areas containing the peak voxel of each cluster. [Correction added on 13 September 2023, after first online publication: The red boxes in Figure 7a were incorrectly filled instead of outlined to highlight the areas of an effect. This has been corrected in this version]

overlapping with the areas observed for the other theta- and beta-band effects.

Taken together, oscillatory power in both the alpha and beta bands revealed more widespread differences during maintenance compared with encoding, yet in different cortical areas. Interestingly, these differences were largely outside of auditory sensory areas. This is in contrast to the effects observed in the encoding time window for beta and theta bands, whose estimated

sources were in or close to left-hemispheric auditory areas. Given the fact that the time windows and/or estimated source of the significant effects were different between the three investigated frequency bands, it is unlikely that they are due to a broad-band increase in power during *correct* trials. Rather, we suggest that each frequency band is part of a distinct functional mechanism, which jointly contributes to successful task performance.

4 | DISCUSSION

Auditory short-term memory plays a critical role in our daily life, from communication, to scene analysis and music appreciation. Understanding the brain mechanisms that support auditory memory is important for elucidating fundamental brain functions and understanding individual differences and impairments, for example such as those arising in the course of aging (Bopp & Verhaeghen, 2005; Cooper et al., 2006; Rimmele et al., 2012; Ruzzoli et al., 2012; Sur & Golob, 2020). Previous work into auditory short-term memory predominantly focused on contrasting tasks of different difficulty (i.e. high vs. low memory load). Here, we ask a somewhat different question—what are the neural correlates that distinguish ultimately successful versus unsuccessful short-term retention of a complex auditory scene?

We employed a novel ASTM task (Figure 2), which required listeners to encode and retain three simultaneously presented artificial auditory objects, defined by frequency and rate. Thus, the proper execution of the STM task necessitates the *correct* and sustained binding of the frequency and rate of each of the three auditory streams, and failure to do so potentially underlies *incorrect* task performance (Bizley & Cohen, 2013; Darwin, 1997; Joseph et al., 2016, 2015). We analysed activity during the encoding and maintenance stages of the task (i.e. before the probe stimulus was presented) to determine what patterns of activation might be associated with ultimately *correct* or *incorrect* responses to the probe. Our results reveal a distributed pattern of enhanced ERFs and oscillatory neural activity in several frequency bands specific to encoding and maintenance stages and indicative of successful short-term memory performance.

4.1 | STM search and confidence effects revealed in behavioural response patterns

Overall, listeners demonstrated an average hit rate of 71%, confirming a challenging behavioural task. We observed faster RTs for TP (true positive; i.e. when the probe was present in the scene and identified as such) versus TN responses (true negative; when the probe was absent from the previously heard scene and identified as such), which suggests a non-exhaustive scan of STM storage in cases participants correctly determined the probe to be present in the just heard scene. Participants were also faster during *correct* compared with *incorrect* trials, indicative of higher levels of response confidence in the former (Kahana & Loftus, 1999; Murdock, 1985; Weidemann & Kahana, 2016).

4.2 | Behavioural responses suggest perceptual salience of the 3-Hz rate

Interestingly, performance for trials containing 3-Hz probes stood out in two aspects. First, mean RTs were faster for 3 Hz compared with 7 and 9 Hz probe trials when participants gave a *same* response (i.e. TP and FP), regardless of whether the outcome was *correct* or *incorrect*. The opposite was true for TN trials, in which they gave a *different* response. Second, participants gave significantly more *same* than *different* responses for 3 Hz compared with 7 or 19 Hz probe trials. Again, this was the case regardless of whether the outcome was *correct* or *incorrect* and despite the fact that participants received training and feedback after every trial. This performance pattern indicates that slow repetition rates (3 Hz here) might be perceptually more salient than faster rates. That is, they might pop out perceptually and attract attention, but, as discussed below, depending on the task context might ultimately impede performance.

Previous works have highlighted differences in the natural occurrence of and processing mechanisms for auditory stimuli with slow compared with fast amplitude modulation (e.g. Giraud & Poeppel, 2012; Poeppel, 2003). One hypothesis is that this reflects the tuning of the auditory system to the syllabic rate of spoken language (~4 Hz; Chandrasekaran et al., 2009; Chi et al., 1999; Elliott & Theunissen, 2009). Notably, here, 3-Hz probes were also more often incorrectly identified as being present in the scene. This suggests that rate and frequency information were not strictly bound in these sounds and that the 3-Hz rate information dominated the perceptual representation even though the task required participants to remember the rate in association with the carrier. These incidental observations speak to a longstanding question (e.g. see reviewed in Sohoglu et al., 2020) regarding whether spectral and temporal acoustic features are represented independently, or in an integrated manner.

4.3 | No differences in power-spectral density between correct and incorrect trials during the encoding period

The analysis of brain responses focused on the encoding and maintenance stages—that is before the presentation of the probe (and the ensuing behavioural response). This allowed us to pinpoint the aspects of brain activity that were most different between subsequently successful or failed trials.

As expected, we found spectral peaks at the three stimulus rate frequencies (3, 7, 19 Hz and harmonics)

during the encoding stage. In line with the observed ERFs (see below), these were localized in or close to primary auditory cortex. The spectral amplitude at 3 Hz was significantly larger compared with 7 and 19 Hz. This matches the behavioural findings of overall faster responses as well as more *same* than *different* responses for trials with 3-Hz probes and suggests an increased saliency of 3 Hz compared with 7 and 19 Hz stimuli.

Contrary to our expectations, however, PSD did not differ between subsequently *correct* and *incorrect* trials. In the visual domain, such a differentiating neural pattern has been observed during the STM maintenance of several stimuli (Killebrew et al., 2018; Peterson et al., 2014), and one study found a comparable result using only a single auditory stimulus stream (Falk et al., 2017). As stated above, and given the many demonstrations that focused attention is associated with increased SNR at the stimulus rate (Ding & Simon, 2013; Elhilali et al., 2009; Mesgarani & Chang, 2012; Riecke et al., 2014; Xiang et al., 2010; Zion Golumbic et al., 2013), the absence of a difference between *correct* and *incorrect* trials here suggests that stimulus-directed attention per se is not what affected success or failure in the present task. Namely, *correct* trials are not simply those trials during which participants tracked the unfolding auditory streams more intently during the encoding stage.

Further, we observed no persistence of stimulus rate representation into the maintenance stage. In principle, such a continuation of rhythmic neural activity beyond the offset of the entraining auditory stimulus is possible, as two recent studies have demonstrated (Köseme et al., 2018; van Bree et al., 2020). A likely reason for the current lack of observable PSD effects lies in the more complex stimulus characteristics (i.e. a mix of three simultaneous sound streams instead of a single stream used in previous work) and consequent potential variability in rehearsal strategies across trials and participants.

4.4 | **Correct trials are associated with increased event-related activity relative to incorrect trials**

Starting early on during the encoding stage and continuing throughout the maintenance stage, we observed increased ERF activity for *correct* compared with *incorrect* trials, particularly at temporo-parietal sensors. Previous studies have reported analogous increases in event-related EEG and MEG activity during the maintenance period of high compared with low-load ASTM tasks (Grimault et al., 2014; Huang et al., 2016; Nolden et al., 2013). Similarly, Cano and Knight (2016) observed stronger ERP responses in a think/no-think paradigm,

where participants tried to remember an auditorily presented word pair compared with when they tried to suppress it. In a cross-species comparison between humans and monkeys, Huang et al. (2016) provided evidence that such increased ERF activity likely originates from elevated firing of neurons in auditory cortex. This is in line with our observed source activity particularly during the encoding stage, in which *correct* compared with *incorrect* trials showed stronger activity in or very close to primary auditory cortex (see also Albouy et al., 2013; Kumar et al., 2016, 2021).

During maintenance, source-level activity showed a more widely distributed pattern. In particular, we observed stronger ERFs for *correct* compared with *incorrect* trials in the dorsolateral prefrontal cortex (DLPFC), a key structure in visual and auditory short-term memory operations (Barbey et al., 2013; Bodner et al., 1996; Goldman-Rakic, 1995) and sustained attentional control (Kane & Engle, 2002). Load-related increases in DLPFC activity were previously found in an MEG study during a non-verbal auditory STM task (Grimault et al., 2014) as well as in a fMRI study specifically dissociating attention- and load-related activations during auditory STM (Huang et al., 2013).

Rather than the actual storage of information, recent models have proposed that PFC activity during short-term memory operations likely reflects executive top-down signals to sensory areas concerned with maintenance (Lara & Wallis, 2015; Postle, 2006; Sreenivasan et al., 2014) as well as rehearsal strategies (Awh et al., 1996; D'Esposito et al., 2000). Similarly, our further observed parietal ERF sources (superior parietal lobe, postcentral gyrus) have previously been associated with motor sequencing during speech preparation and/or production (Dhanjal et al., 2008; Heim et al., 2012). Although the present stimuli are unlikely maintained in the same way as speech, this hints at a potential involvement of silent rehearsal processes, especially during successful maintenance (see also Nolden et al., 2013).

Together, the ERF results are consistent with previous demonstrations of persistent activity during encoding and maintenance of auditory short-term memories (Fuster & Jervey, 1982; Huang et al., 2016; Constantinidis et al., 2018; but see also Masse et al., 2020). Critically, unlike the previous investigations reviewed above where the stimuli/task load were manipulated to affect ASTM performance, here, task difficulty was kept fixed throughout the experiment. The activations observed suggest that parietal activity is also modulated on a trial-by-trial basis. This activity likely either directly represents the memorized information or pertains to supporting mechanisms for rehearsal or suppression of distractors during the maintenance period that ultimately results in a successful versus unsuccessful performance.

4.5 | Successful ASTM performance relies on distinct sources of alpha-band activity in posterior and fronto-parietal regions

As hypothesized, we observed increased ongoing alpha-band activity in *correct* compared with *incorrect* trials, during both encoding and maintenance stages. Particularly, the latter exhibited power differences in a large area around right posterior parietal regions. This extends previous studies, which reported general increases in posterior alpha-band activity during ASTM operations compared with rest (Kawasaki et al., 2010; Luo et al., 2005), as well as parametrically increasing with STM load (Obleser et al., 2012; Wisniewski et al., 2017). For example, Obleser et al. (2012) had participants memorize different numbers of spoken digits and found that alpha-band power during the maintenance interval increases with memory load. This memory-related increase in posterior alpha-power has also consistently been shown for other sensory modalities (Jensen et al., 2002; Spitzer & Blankenburg, 2012; Van Dijk et al., 2010) and is widely suggested to reflect functional inhibition of task-irrelevant cortical areas (Jensen & Mazaheri, 2010; Klimesch et al., 2007; Wilsch & Obleser, 2016). That is, alpha-band mediated inhibition ensures that sufficient processing resources are allocated for attending to and maintaining auditory stimulus information. Our data further develops this line of research by demonstrating that a certain level of functional inhibition via alpha-band power increase seems necessary for successful ASTM performance. Because no additional salient external sensory inputs were present in our experiment, the observed posterior alpha activity likely reflects protection from internal rather than external noise.

Interestingly, in addition to the posterior increases of alpha power associated with functional inhibition, our data also show a frontal as well as a temporo-parietal source. The latter, which comprised the strongest source of alpha-band power differences in our data, has repeatedly been found to be activated during ASTM (Gaab et al., 2003; Obleser et al., 2012; Paulesu et al., 1993; Wilsch et al., 2015). For instance, in an fMRI study, Gaab et al. (2003) observed increased activity in the supramarginal gyrus, overlapping with our temporo-parietal source, during a pitch retention task, which correlated with task performance. More generally, an increasing number of studies have reported corresponding increases in alpha activity outside of sensory areas during STM operations (Leiberg et al., 2006; Palva et al., 2010; Wilsch et al., 2018). These activations hint at an important role of alpha in more executive functions and reflect aspects of top-down modulation of sensory areas during short-term memory operation, in order to temporally maintain sensory

representations (Wilsch & Obleser, 2016). In this light, our results extend existing knowledge by demonstrating that successful ASTM operations rely, at least partly, on two distinct types of alpha-band activity: One originating in posterior regions, concerned with functional inhibition, and one originating in fronto-parietal regions, involved in executive control of sensory maintenance.

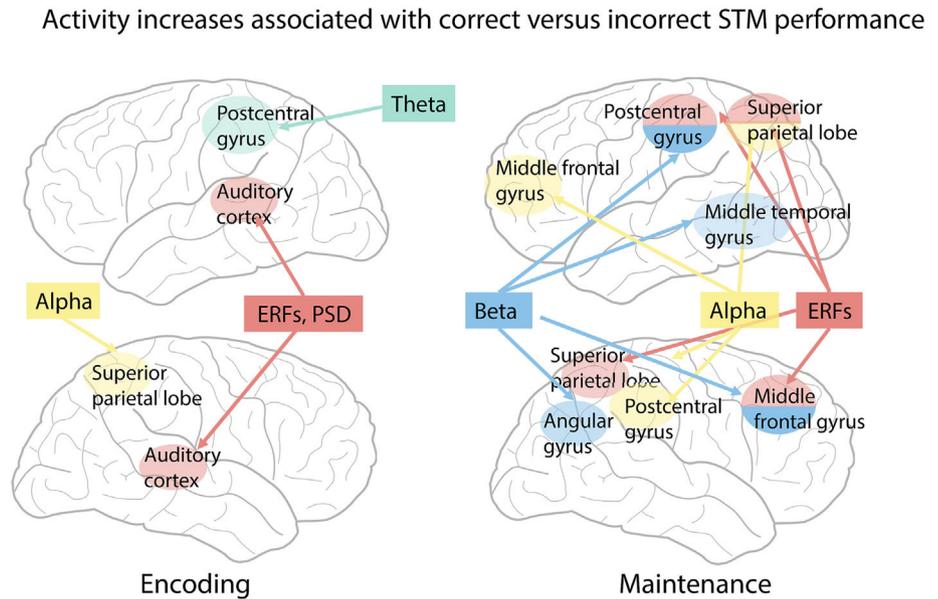
4.6 | Increased left parietal theta-band activity during the encoding stage differentiates *correct* from *incorrect* trials

In contrast to the alpha and beta bands, we observed effects in theta-band activity only during the encoding stage where activity was enhanced during subsequently *correct* compared with *incorrect* trials. These effects arose partway through the trial and therefore may be related to fatigue or random failure of the underlying STM encoding mechanisms during the *incorrect* trials.

In visual STM tasks, theta activity has been prominently observed in a large number of studies, both in posterior sensory and association areas (Costers et al., 2020; Raghavachari et al., 2006, 2001) as well as in frontal-midline regions (Jensen & Tesche, 2002; Kawasaki et al., 2010; Roux & Uhlhaas, 2014; Meltzer et al., 2017; Proskovec et al., 2018; Van Ede et al., 2017). Although there is less known regarding the role of theta in STM outside the visual domain (but see, e.g. Cano & Knight, 2016), recent studies have demonstrated an increase in ASTM performance by enhancing ongoing cortical theta activity during maintenance using transcranial magnetic stimulation (Albouy et al., 2017) and rhythmic visual stimulation (Albouy et al., 2022). Based on an a priori measurement of STM-related neural activity via EEG and MEG, their site of stimulation was located over the left intraparietal sulcus. Notably, our presently observed source for the contrast of theta power between *correct* and *incorrect* trials was in the same region, around the left supramarginal gyrus, as well as the left middle temporal gyrus.

Unlike the present results, the theta-band effect from Albouy et al. (2017) was present in the maintenance stage. Although likely tapping into the same neural mechanisms, a possible reason for this difference is that the task used by Albouy et al. (2017) required active manipulation of STM contents compared with a simple maintenance in the present study. Using fMRI, other studies have also found evidence for auditory STM-related operations in this region, especially during highly demanding tasks (Foster et al., 2013; Foster & Zatorre, 2010; Zatorre et al., 2010). Additionally, the present effects in the alpha and beta band power were also located around the supramarginal gyrus, further supporting its importance in auditory STM operations.

FIGURE 8 Summary of findings. Illustration of brain areas showing increased neural activity during *correct* versus *incorrect* ASTM performance, separately for encoding (left) and maintenance stages (right). ERF, event-related fields; PSD, power-spectral density.



4.7 | Increased beta-band power during maintenance may relate to binding of ASTM contents in sensory areas

Corresponding to our results in the alpha and theta bands, beta-band power was elevated during *correct* versus *incorrect* trials.

Compared with other frequency bands, the role of beta-band activity during STM, and in particular for auditory information, has been considerably less studied. In line with our data, Leiberg et al. (2006) found that beta-band activity in temporal areas increases parametrically with auditory STM load. In addition, several previous studies have reported analogously increased levels of beta-band activity during visual STM tasks, predominantly in visual sensory areas (Chen & Huang, 2016; Daume et al., 2017; Deiber et al., 2007; Fodor et al., 2018; Killebrew et al., 2018). For instance, in a recent visual short-term memory study, Fodor et al. (2018) observed reduced beta-band activity in participants with mild cognitive impairment compared with healthy controls, along with reduced STM performance. Further, linking structural MRI to the EEG measures revealed that the strength of beta-band activity correlated significantly with the size of the hippocampus, entorhinal cortex and parahippocampal cortex, all of which are suggested to be involved in STM operations (Axmacher et al., 2008; Kumar et al., 2016). The fact that we observed more pronounced and widespread differences between *correct* and *incorrect* trials during the maintenance stage might suggest that beta band is particularly involved in the preservation of sensory information. In line with this interpretation, Weiss and Mueller (2012) speculated that beta-band activity reflects the binding of auditory short-term memory contents in sensory areas.

Spatially, the strongest difference between *correct* and *incorrect* trials in the beta band was localized around the supramarginal gyrus. This mirrors our findings in the alpha- and theta-band results and adds a further indication of this region's important role in auditory STM operations (Celsis et al., 1999; Gaab et al., 2003; Paulesu et al., 1993; Salmon et al., 1996). Further significant effects encompassed the left angular gyrus and the left middle temporal gyrus. A joint increase in these two regions during auditory processing compared with a non-auditory control condition has been found by both Goycoolea et al. (2005) and Ahmad et al. (2003) using fMRI.

Finally, two additionally observed sources constitute the secondary visual cortex and the frontal eye field. We speculate that these reflect increased suppression of potentially distracting visual information in subsequently *correct* trials.

5 | CONCLUSIONS

This study goes beyond previous findings by providing a thorough account of oscillatory and evoked neural activity patterns differentiating (ultimately) successful from unsuccessful ASTM performance (see Figure 8 for a summary). Notably, we show that successful compared with unsuccessful ASTM performance was associated with a significant overall enhancement of event-related fields and oscillatory activity in the theta, alpha and beta frequency ranges in line with previous reports of persistent neural activity changes during STM. Effects in each frequency band were linked with different underlying sources, suggesting that they represent distinct functional mechanisms that jointly enable successful memory

performance. Spatially, these effects were confined to posterior sensory areas during encoding and spread to parietal and frontal areas during maintenance, presumably reflecting STM contents and top-down control processes, respectively.

AUTHOR CONTRIBUTIONS

Ulrich Pomper: conceptualization, investigation, formal analysis, writing—original draft; **Lorenza Zaira Curetti:** investigation, formal analysis, writing—review editing; **Maria Chait:** funding acquisition, supervision, conceptualization, formal analysis, writing—original draft.

ACKNOWLEDGEMENTS

This research was supported by an EC Horizon 2020 grant and a BBSRC international partnering award to MC and an Austrian Science Fund Young Independent Researchers Group grant to UP (ZK66).

CONFLICT OF INTEREST STATEMENT

None.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16140>.

DATA AVAILABILITY STATEMENT

Data and relevant code are accessible via https://osf.io/wafby/?view_only=136797a93c1c429492045185e60f0640.

ORCID

Ulrich Pomper  <https://orcid.org/0000-0002-6534-968X>

Maria Chait  <https://orcid.org/0000-0002-7808-3593>

REFERENCES

- Ahmad, Z., Balsamo, L. M., Sachs, B. C., Xu, B., & Gaillard, W. D. (2003). Auditory comprehension of language in young children: Neural networks identified with fMRI. *Neurology*, *60*(10), 1598–1605. <https://doi.org/10.1212/01.WNL.0000059865.32155.86>
- Albouy, P., Martinez-Moreno, Z. E., Hoyer, R. S., Zatorre, R. J., & Baillet, S. (2022). Supramodality of neural entrainment: Rhythmic visual stimulation causally enhances auditory working memory performance. *Science Advances*, *8*(8), eabj9782. <https://doi.org/10.1126/sciadv.abj9782>
- Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain*, *136*(5), 1639–1661. <https://doi.org/10.1093/brain/awt082>
- Albouy, P., Weiss, A. A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron*, *94*(1), 193–206.e5. <https://doi.org/10.1016/j.neuron.2017.03.015>
- Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience*, *17*(5), 819–831. <https://doi.org/10.1162/0898929053747612>
- Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, *7*(1), 25–31. <https://doi.org/10.1111/j.1467-9280.1996.tb00662.x>
- Axmacher, N., Schmitz, D. P., Wagner, T., Elger, C. E., & Fell, J. (2008). Interactions between medial temporal lobe, prefrontal cortex, and inferior temporal regions during visual working memory: A combined intracranial EEG and functional magnetic resonance imaging study. *The Journal of Neuroscience*, *28*(29), 7304–7312. <https://doi.org/10.1523/JNEUROSCI.1778-08.2008>
- Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, *49*(5), 1195–1205. <https://doi.org/10.1016/j.cortex.2012.05.022>
- Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews. Neuroscience*, *14*, 693–707. <https://doi.org/10.1016/j.micinf.2011.07.011>. *Innate*
- Bodner, M., Kroger, J., & Fuster, J. M. (1996). Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport*, *7*(12), 1905–1908. <https://doi.org/10.1097/00001756-199608120-00006>
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *60*(5), 223–233. <https://doi.org/10.1093/geronb/60.5.P223>
- van Bree, S., Sohoglu, E., Davis, M., & Zoefel, B. (2020). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLoS Biology*, *19*(2), e3001142. <https://doi.org/10.1101/2020.06.26.170761>
- Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*(4), 687–697. <https://doi.org/10.1016/j.neuron.2005.09.029>
- Cano, M. E., & Knight, R. T. (2016). Behavioral and EEG evidence for auditory memory suppression. *Frontiers in Human Neuroscience*, *10*, 133. <https://doi.org/10.3389/fnhum.2016.00133>
- Cappotto, D., Auksztulewicz, R., Kang, H., Poeppel, D., Melloni, L., & Schnupp, J. (2021). Decoding the content of auditory sensory memory across species. *Cerebral Cortex*, *31*(7), 3226–3236. <https://doi.org/10.1093/cercor/bhab002>
- Celsis, P., Boulouvar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, *9*(1), 135–144. <https://doi.org/10.1006/nimg.1998.0389>
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436. <https://doi.org/10.1371/journal.pcbi.1000436>
- Chen, Y., & Huang, X. (2016). Modulation of alpha and beta oscillations during an n-back task with varying temporal memory load. *Frontiers in Psychology*, *6*, 2031. <https://doi.org/10.3389/fpsyg.2015.02031>
- de Cheveigné, A., & Parra, L. C. (2014). Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage*, *98*, 487–505. <https://doi.org/10.1016/j.neuroimage.2014.05.068>

- de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods*, 171(2), 331–339. <https://doi.org/10.1016/j.jneumeth.2008.03.015>
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, 106(5), 2719–2732. <https://doi.org/10.1121/1.428100>
- Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X. L., Wang, M., & Arnsten, A. F. T. (2018). Persistent spiking activity underlies working memory. *The Journal of Neuroscience*, 38(32), 7020–7028. <https://doi.org/10.1523/JNEUROSCI.2486-17.2018>
- Cooper, R. J., Todd, J., McGill, K., & Michie, P. T. (2006). Auditory sensory memory and the aging brain: A mismatch negativity study. *Neurobiology of Aging*, 27(5), 752–762. <https://doi.org/10.1016/j.neurobiolaging.2005.03.012>
- Costers, L., Van Schependom, J., Laton, J., Baijot, J., Sjøgård, M., Wens, V., De Tiège, X., Goldman, S., D'Haeseleer, M., D'hooghe, M. B., Woolrich, M., & Nagels, G. (2020). Spatiotemporal and spectral dynamics of multi-item working memory as revealed by the n-back task using MEG. *Human Brain Mapping*, 41(9), 2431–2446. <https://doi.org/10.1002/hbm.24955>
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, 1(9), 327–333. [https://doi.org/10.1016/S1364-6613\(97\)01097-8](https://doi.org/10.1016/S1364-6613(97)01097-8)
- Daume, J., Gruber, T., Engel, A. K., & Frieze, U. (2017). Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory. *The Journal of Neuroscience*, 37(2), 313–322. <https://doi.org/10.1523/JNEUROSCI.2130-16.2016>
- Deiber, M. P., Missonnier, P., Bertrand, O., Gold, G., Fazio-Costa, L., Ibañez, V., & Giannakopoulos, P. (2007). Distinction between perceptual and attentional processing in working memory tasks: A study of phase-locked and induced oscillatory brain dynamics. *Journal of Cognitive Neuroscience*, 19(1), 158–172. <https://doi.org/10.1162/jocn.2007.19.1.158>
- D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: Evidence from event-related fMRI studies. *Experimental Brain Research*, 133(1), 3–11. <https://doi.org/10.1007/s002210000395>
- Deutsch, P., Czoschke, S., Fischer, C., Kaiser, J., & Bledowski, C. (2023). Decoding of working memory contents in auditory cortex is not distractor-resistant. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 43(18), 3284–3293. <https://doi.org/10.1523/JNEUROSCI.1890-22.2023>
- Dhanjal, N. S., Handunnetthi, L., Patel, M. C., & Wise, R. J. S. (2008). Perceptual systems controlling speech production. *The Journal of Neuroscience*, 28(40), 9969–9975. <https://doi.org/10.1523/JNEUROSCI.2607-08.2008>
- Dimakopoulos, V., Mégevand, P., Stieglitz, L. H., Imbach, L., & Sarnthein, J. (2022). Information flows from hippocampus to auditory cortex during replay of verbal working memory items. *eLife*, 11, e78677. <https://doi.org/10.7554/eLife.78677>
- Ding, N., & Simon, J. Z. (2013). Robust cortical encoding of slow temporal modulations of speech. *Adv. Exp. Med. Biol.*, 787, 373–381. https://doi.org/10.1007/978-1-4614-1590-9_41
- Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biology*, 7(6), e1000129. <https://doi.org/10.1371/journal.pbio.1000129>
- Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology*, 5(3), e1000302. <https://doi.org/10.1371/journal.pcbi.1000302>
- Falk, S., Lanzilotti, C., & Schön, D. (2017). Tuning neural phase entrainment to speech. *Journal of Cognitive Neuroscience*, 29(8), 1378–1389. https://doi.org/10.1162/jocn_a_01136
- Fodor, Z., Sirály, E., Horváth, A., Salacz, P., Hidas, Z., Csibri, É., Szabó, Á., & Csukly, G. (2018). Decreased event-related beta synchronization during memory maintenance marks early cognitive decline in mild cognitive impairment. *Journal of Alzheimer's Disease*, 63(2), 489–502. <https://doi.org/10.3233/JAD-171079>
- Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*, 75, 27–35. <https://doi.org/10.1016/j.neuroimage.2013.02.044>
- Foster, N. E. V., & Zatorre, R. J. (2010). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, 53, 26–36. <https://doi.org/10.1016/j.neuroimage.2010.06.042>
- Friston, K., Henson, R., Phillips, C., & Mattout, J. (2006). Bayesian estimation of evoked and induced responses. *Human Brain Mapping*, 27(9), 722–735. <https://doi.org/10.1002/hbm.20214>
- Fuster, J. M., & Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience*, 2(3), 361–375. <https://doi.org/10.1523/jneurosci.02-03-00361.1982>
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory—An fMRI study with sparse temporal sampling. *NeuroImage*, 19(4), 1417–1426. [https://doi.org/10.1016/S1053-8119\(03\)00224-6](https://doi.org/10.1016/S1053-8119(03)00224-6)
- Giraud, A.-L. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511–517. <https://doi.org/10.1038/nn.3063>
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14(3), 477–485. [https://doi.org/10.1016/0896-6273\(95\)90304-6](https://doi.org/10.1016/0896-6273(95)90304-6)
- Goycoolea, M., Mena, I., & Neubauer, S. (2005). Functional studies of the human auditory pathway after monaural stimulation with pure tones. Establishing a normal database. *Acta Oto-Laryngologica*, 125(5), 513–519. <https://doi.org/10.1080/00016480510026250>
- Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., Zatorre, R., Robitaille, N., & Jolicoeur, P. (2014). 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. <https://doi.org/10.1016/j.neuroimage.2014.03.020>
- Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J. R., van Wassenhove, V., Wibrall, M., & Schoffelen, J. M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, 65, 349–363. <https://doi.org/10.1016/j.neuroimage.2012.10.001>

- Halder, T., Talwar, S., Jaiswal, A. K., & Banerjee, A. (2019). Quantitative evaluation in estimating sources underlying brain oscillations using current source density methods and beamformer approaches. *eNeuro*, 6(4). <https://doi.org/10.1523/ENEURO.0170-19.2019>
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1), 35–42. <https://doi.org/10.1007/BF02512476>
- Heim, S., Amunts, K., Hensel, T., Grande, M., Huber, W., Binkofski, F., & Eickhoff, S. B. (2012). The role of human parietal area 7A as a link between sequencing in hand actions and in overt speech production. *Frontiers in Psychology*, 3, 534. <https://doi.org/10.3389/fpsyg.2012.00534>
- Huang, N., & Elhilali, M. (2020). Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. *eLife*, 9, e52984. <https://doi.org/10.7554/eLife.52984>
- Huang, Y., Matysiak, A., Heil, P., König, R., & Brosch, M. (2016). Persistent neural activity in auditory cortex is related to auditory working memory in humans and nonhuman primates. *eLife*, 5, e15441. <https://doi.org/10.7554/eLife.15441>
- Huang, S., Seidman, L. J., Rossi, S., & Ahveninen, J. (2013). Distinct cortical networks activated by auditory attention and working memory load. *NeuroImage*, 83, 1098–1108. <https://doi.org/10.1016/j.neuroimage.2013.07.074>
- Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex*, 12(8), 877–882. <https://doi.org/10.1093/cercor/12.8.877>
- Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4, 186. <https://doi.org/10.3389/fnhum.2010.00186>
- Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *The European Journal of Neuroscience*, 15(8), 1395–1399. <https://doi.org/10.1046/j.1460-9568.2002.01975.x>
- Joseph, S., Kumar, S., Husain, M., & Griffiths, T. D. (2015). Auditory working memory for objects vs features. *Frontiers in Neuroscience*, 9, 1–8, 13. <https://doi.org/10.3389/fnins.2015.00013>
- Joseph, S., Teki, S., Kumar, S., Husain, M., & Griffiths, T. D. (2016). Resource allocation models of auditory working memory. *Brain Research*, 1640, 183–192. <https://doi.org/10.1016/j.brainres.2016.01.044>
- Kahana, M. J., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 323–384). MIT Press.
- Kaiser, J. (2015). Dynamics of auditory working memory. *Frontiers in Psychology*, 6, 1–6, 613. <https://doi.org/10.3389/fpsyg.2015.00613>
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, 9(4), 637–671. <https://doi.org/10.3758/BF03196323>
- Kawasaki, M., Kitajo, K., & Yamaguchi, Y. (2010). Dynamic links between theta executive functions and alpha storage buffers in auditory and visual working memory. *The European Journal of Neuroscience*, 31(9), 1683–1689, no. <https://doi.org/10.1111/j.1460-9568.2010.07217.x>
- Killebrew, K. W., Gurariy, G., Peacock, C. E., Berryhill, M. E., & Caplovitz, G. P. (2018). Electrophysiological correlates of encoding processes in a full-report visual working memory paradigm. *Cognitive, Affective, & Behavioral Neuroscience*, 18(2), 353–365.
- Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, 53(1), 63–88. <https://doi.org/10.1016/j.brainresrev.2006.06.003>
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, 30(3), 859–873. <https://doi.org/10.1002/hbm.20550>
- Kosachenko, A. I., Kasanov, D., Kotyusov, A. I., & Pavlov, Y. G. (2023). EEG and pupillometric signatures of working memory overload. *Psychophysiology*, 60(6), e14275. <https://doi.org/10.1111/psyp.14275>
- Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28(18), 2867–2875.e3. <https://doi.org/10.1016/j.cub.2018.07.023>
- Kumar, S., Gander, P. E., Berger, J. I., Billig, A. J., Nourski, K. V., Oya, H., Kawasaki, H., Howard, M. A., & Griffiths, T. D. (2021). Oscillatory correlates of auditory working memory examined with human electrocorticography. *Neuropsychologia*, 150, 107691. <https://doi.org/10.1016/j.neuropsychologia.2020.107691>
- Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016). A brain system for auditory working memory. *The Journal of Neuroscience*, 36(16), 4492–4505. <https://doi.org/10.1523/JNEUROSCI.4341-14.2016>
- Lara, A. H., & Wallis, J. D. (2015). The role of prefrontal cortex in working memory: A mini review. *Frontiers in Systems Neuroscience*, 9, 173. <https://doi.org/10.3389/fnsys.2015.00173>
- Leiberg, S., Lutzenberger, W., & Kaiser, J. (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Research*, 1120(1), 131–140. <https://doi.org/10.1016/j.brainres.2006.08.066>
- Lenartowicz, A., Truong, H., Enriquez, K. D., Webster, J., Pochon, J. B., Rissman, J., Bearden, C. E., Loo, S. K., & Bilder, R. M. (2021). Neurocognitive subprocesses of working memory performance. *Cognitive, Affective, & Behavioral Neuroscience*, 21(6), 1130–1152. <https://doi.org/10.3758/s13415-021-00924-7>
- Leung, A. W., & Alain, C. (2011). Working memory load modulates the auditory “what” and “where” neural networks. *NeuroImage*, 55(3), 1260–1269. <https://doi.org/10.1016/j.neuroimage.2010.12.055>
- Lim, S. J., Thiel, C., Sehm, B., Deserno, L., Lepsien, J., & Obleser, J. (2022). Distributed networks for auditory memory differentially contribute to recall precision. *NeuroImage*, 256, 119227. <https://doi.org/10.1016/j.neuroimage.2022.119227>
- Litvak, V., & Friston, K. (2008). Electromagnetic source reconstruction for group studies. *NeuroImage*, 42(4), 1490–1498. <https://doi.org/10.1016/j.neuroimage.2008.06.022>
- Luo, H., Husain, F. T., Horwitz, B., & Poeppel, D. (2005). Discrimination and categorization of speech and non-speech sounds in

- an MEG delayed-match-to-sample study. *NeuroImage*, 28(1), 59–71. <https://doi.org/10.1016/j.neuroimage.2005.05.040>
- Mapelli, I., & Özkurt, T. E. (2019). Brain oscillatory correlates of visual short-term memory errors. *Frontiers in Human Neuroscience*, 13, 33. <https://doi.org/10.3389/fnhum.2019.00033>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Masse, N. Y., Rosen, M. C., & Freedman, D. J. (2020). Reevaluating the role of persistent neural activity in short-term memory. *Trends in Cognitive Sciences*, 24(3), 242–258. <https://doi.org/10.1016/j.tics.2019.12.014>
- Meltzer, J. A., Kiehl, A., Panamsky, L., Links, K. A., Deschamps, T., & Leigh, R. C. (2017). Electrophysiological signatures of phonological and semantic maintenance in sentence repetition. *NeuroImage*, 156, 302–314. <https://doi.org/10.1016/j.neuroimage.2017.05.030>
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7393), 233–236. <https://doi.org/10.1038/nature11020>
- Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, 13(6), 511–521. <https://doi.org/10.3758/BF03198322>
- Nolden, S., Grimault, S., Guimond, S., Lefebvre, C., Bermudez, P., & Jolicoeur, P. (2013). The retention of simultaneous tones in auditory short-term memory: A magnetoencephalography study. *NeuroImage*, 82, 384–392. <https://doi.org/10.1016/j.neuroimage.2013.06.002>
- Noyce, A. L., Lefco, R. W., Brissenden, J. A., Tobbyne, S. M., Shinn-Cunningham, B. G., & Somers, D. C. (2022). Extended frontal networks for visual and auditory working memory. *Cerebral Cortex*, 32(4), 855–869. <https://doi.org/10.1093/cercor/bhab249>
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *The Journal of Neuroscience*, 32(36), 12376–12383. <https://doi.org/10.1523/JNEUROSCI.4908-11.2012>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869. <https://doi.org/10.1155/2011/156869>
- Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neural synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), 7580–7585. <https://doi.org/10.1073/pnas.0913113107>
- Papagno, C., Comi, A., Riva, M., Bizzi, A., Vernice, M., Casarotti, A., Fava, E., & Bello, L. (2017). Mapping the brain network of the phonological loop. *Human Brain Mapping*, 38(6), 3011–3024. <https://doi.org/10.1002/hbm.23569>
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362(6418), 342–345. <https://doi.org/10.1038/362342a0>
- Peterson, D. J., Gurariy, G., Dimotsantos, G. G., Arciniega, H., Berryhill, M. E., & Caplovitz, G. P. (2014). Neuropsychologia the steady-state visual evoked potential reveals neural correlates of the items encoded into visual working memory. *Neuropsychologia*, 63, 145–153. <https://doi.org/10.1016/j.neuropsychologia.2014.08.020>
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time”. *Speech Communication*, 41(1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139(1), 23–38. <https://doi.org/10.1016/j.neuroscience.2005.06.005>
- Proskovec, A. L., Wiesman, A. I., Heinrichs-Graham, E., & Wilson, T. W. (2018). Beta oscillatory dynamics in the prefrontal and superior temporal cortices predict spatial working memory performance. *Scientific Reports*, 8(1), 8488. <https://doi.org/10.1038/s41598-018-26863-x>
- Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., Madsen, J. R., & Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *The Journal of Neuroscience*, 21(9), 3175–3183. <https://doi.org/10.1523/jneurosci.21-09-03175.2001>
- Raghavachari, S., Lisman, J. E., Tully, M., Madsen, J. R., Bromfield, E. B., & Kahana, M. J. (2006). Theta oscillations in human cortex during a working-memory task: Evidence for local generators. *Journal of Neurophysiology*, 95(3), 1630–1638. <https://doi.org/10.1152/jn.00409.2005>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- Ratcliffe, O., Shapiro, K., & Staresina, B. P. (2022). Fronto-medial theta coordinates posterior maintenance of working memory content. *Current Biology*, 32(10), 2121–2129.e3. <https://doi.org/10.1016/j.cub.2022.03.045>
- Riecke, L., Scharke, W., Valente, G., & Gutschalk, A. (2014). Sustained selective attention to competing amplitude-modulations in human auditory cortex. *PLoS ONE*, 9(9), e108045. <https://doi.org/10.1371/journal.pone.0108045>
- Rimmele, J., Sussman, E., Keitel, C., Jacobsen, T., & Schröger, E. (2012). Electrophysiological evidence for age effects on sensory memory processing of tonal patterns. *Psychology and Aging*, 27(2), 384–398. <https://doi.org/10.1037/a0024866>
- Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, 18(1), 16–25. <https://doi.org/10.1016/j.tics.2013.10.010>
- Ruzzoli, M., Pirulli, C., Brignani, D., Maioli, C., & Miniussi, C. (2012). Sensory memory during physiological aging indexed by mismatch negativity (MMN). *Neurobiology of Aging*, 33(3), 625.e21–625.e30. <https://doi.org/10.1016/j.neurobiolaging.2011.03.021>
- Salmon, E., Van Der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., & Luxen, A. (1996). Regional brain activity during working memory tasks. *Brain*, 119(5), 1617–1625. <https://doi.org/10.1093/brain/119.5.1617>
- Sohoglu, E., Kumar, S., Chait, M., & Griffiths, T. D. (2020). Multi-voxel codes for representing and integrating acoustic features

- in human cortex. *NeuroImage*, 217, 116661. <https://doi.org/10.1016/j.neuroimage.2020.116661>
- Spitzer, B., & Blankenburg, F. (2012). Supramodal parametric working memory processing in humans. *The Journal of Neuroscience*, 32(10), 3287–3295. <https://doi.org/10.1523/JNEUROSCI.5280-11.2012>
- Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, 18(2), 82–89. <https://doi.org/10.1016/j.tics.2013.12.001>
- Sur, S., & Golob, E. J. (2020). Neural correlates of auditory sensory memory dynamics in the aging brain. *Neurobiology of Aging*, 88, 128–136. <https://doi.org/10.1016/j.neurobiolaging.2019.12.020>
- Teng, X., Tian, X., Rowland, J., & Poeppel, D. (2017). Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scale. *PLoS Biology*, 15(11), e2000812.
- Van Dijk, H., Van Der Werf, J., Mazaheri, A., Medendorp, W. P., & Jensen, O. (2010). Modulations in oscillatory activity with amplitude asymmetry can produce cognitively relevant event-related responses. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), 900–905. <https://doi.org/10.1073/pnas.0908821107>
- Van Ede, F., Jensen, O., Maris, E. (2017). Supramodal theta, gamma, and sustained fields predict modality-specific modulations of alpha and beta oscillations during visual and tactile working memory. *Journal of Cognitive Neuroscience*, 29(8), 1455–1472. <https://doi.org/10.1162/jocn.1455-1472>
- Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*, 3(4), 150670. <https://doi.org/10.1098/rsos.150670>
- Weiss, S., & Mueller, H. M. (2012). “Too many betas do not spoil the broth”: The role of beta brain oscillations in language processing. *Frontiers in Psychology*, 3, 201. <https://doi.org/10.3389/fpsyg.2012.00201>
- Weisz, N., Kraft, N. G., & Demarchi, G. (2020). Auditory cortical alpha/beta desynchronization prioritizes the representation of memory items during a retention period. *eLife*, 9, e55508. <https://doi.org/10.7554/eLife.55508>
- Wilsch, A., Henry, M. J., Herrmann, B., Herrmann, C. S., & Obleser, J. (2018). Temporal expectation modulates the cortical dynamics of short-term memory. *The Journal of Neuroscience*, 38(34), 7428–7439. <https://doi.org/10.1523/JNEUROSCI.2928-17.2018>
- Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., & Obleser, J. (2015). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cerebral Cortex*, 25(7), 1938–1946. <https://doi.org/10.1093/cercor/bhu004>
- Wilsch, A., & Obleser, J. (2016). What works in auditory working memory? A neural oscillations perspective. *Brain Research*, 1640(B), 193–207. <https://doi.org/10.1016/j.brainres.2015.10.054>
- Wisniewski, M. G., Thompson, E. R., & Iyer, N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology*, 54(12), 1916–1928. <https://doi.org/10.1111/psyp.12968>
- Wöstmann, M., Lim, S.-J., & Obleser, J. (2017). The human neural alpha response to speech is a proxy of attentional control. *Cerebral Cortex*, 27(6), 3307–3317. <https://doi.org/10.1093/cercor/bhx074>
- Xiang, J., Simon, J., & Elhilali, M. (2010). Competing streams at the cocktail party: Exploring the mechanisms of attention and temporal integration. *The Journal of Neuroscience*, 30(36), 12084–12093. <https://doi.org/10.1523/JNEUROSCI.0827-10.2010>
- Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, 22(4), 775–789. <https://doi.org/10.1162/jocn.2009.21239>
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, 77(5), 980–991. <https://doi.org/10.1016/j.neuron.2012.12.037>

How to cite this article: Pomper, U., Curetti, L. Z., & Chait, M. (2023). Neural dynamics underlying successful auditory short-term memory performance. *European Journal of Neuroscience*, 1–20. <https://doi.org/10.1111/ejn.16140>