Supplemental Information

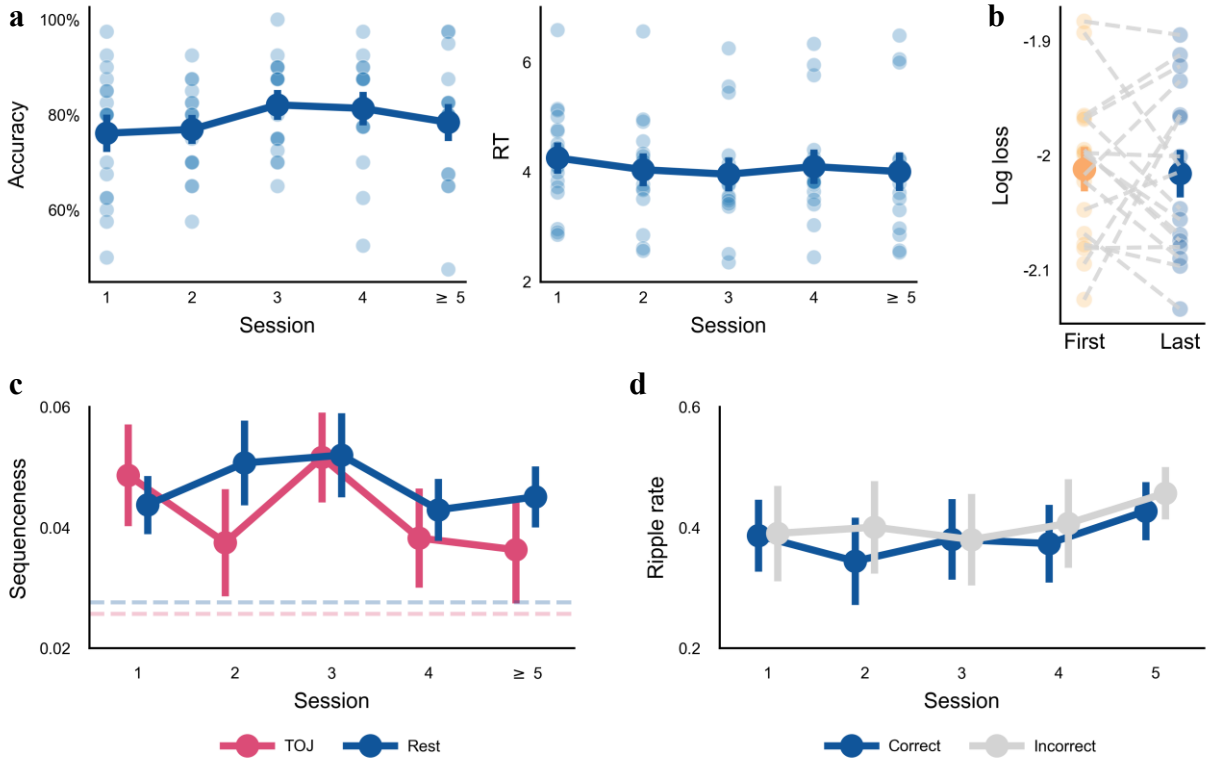

Figure S1. Behavioral and neural measures across sessions, Related to Figure 1

a. Behavioral performance across sessions. Participants demonstrated a clear understanding of task requirements after training, with no observed learning effects throughout the main experimental sessions (left, Accuracy: $p = 0.12,\ \chi^2(1) = 2.46$; right, Reaction Times (RT): $p = 0.70,\ \chi^2 = 0.15$; Likelihood ratio test (LRT), n = 17 participants). Shaded dots denote single observations. Trials that failed due to exceeding the time limit (10 s) of temporal order judgment (TOJ, Figure 1a) were excluded from analysis.

b. Time decodability across sessions. By comparing the loss of time decoders (8 temporal states) trained on the first session and the last session for each participant, we attempted to capture the changes in representational neural patterns over the course of the experiments. Results showed no significant differences ($p = 0.84,\ t(16) = 0.20$, Bayes Factor (BF) $= 0.25$, paired-sample t-test), suggesting stable temporal representations across sessions.

c. Sequenceness across sessions. Sequenceness measures of both on-task (TOJ) and off-task (Rest) replay across sessions showed no significant session effects ($p = 0.42,\ \chi^2(1) = 0.65$; LRT, n = 17 participants). Similar to the main text, this analysis focused on correct TOJ trials, using session-collapsed data for classifier training. Dashed lines indicate the permutation-derived thresholds (see Method).

d. Ripple rates across sessions. Ripple rates in the hippocampus remained stable across sessions, averaging $\sim$ 0.4 Hz, with no significant effect of TOJ responses or sessions on ripple rate (Sessions: $p = 0.91,\ \chi^2(1) = 0.01$; TOJ Responses: $p = 0.066,\ \chi^2(1) = 3.38$; LRT, n = 5 participants). Ripple rate in this analysis was calculated across the entire session, including TOJ, resting, and other trial phases.

Note that only three participants completed more than five sessions in our dataset, so data for session 5–7 was combined in (a) and (c). Error bars represent the between-subjects standard error of the mean (SEM).

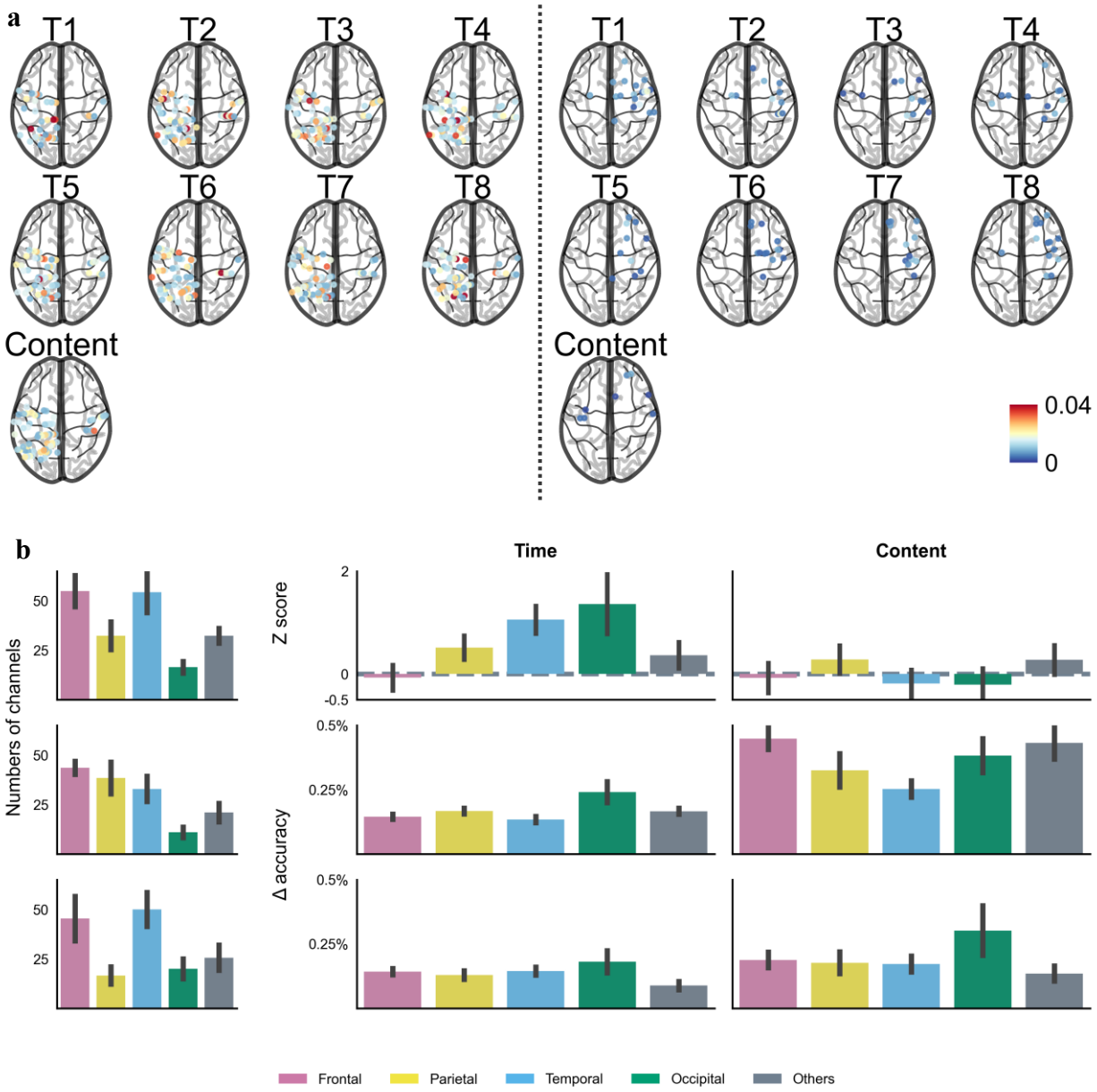

Figure S2. Exemplar feature maps and regional decoding contributions, Related to Figure 2

a. Exemplar feature map. Feature maps for time decoders (T1–T8) and content decoder from one high-performing (left) and one low-performing (right) representative participant. Colors represent the absolute values of decoder coefficients.

(b) Regional contributions to decoding tasks. Panels show the number of recording sites (left) and feature importance for decoding time (middle) and content (right), analyzed for the whole brain (top), left hemisphere (middle), and right hemisphere (bottom). Feature ablation analysis shows that temporal lobe activity uniquely contributes to time decoding accuracy ($p < 0.01,\ t_{13} = 3.97$, one-sample t-test, FWE-corrected for multiple comparisons), with no significant effects observed in other brain regions. Feature importance was assessed as the z-scored decrease in decoding accuracy following feature ablation, compared to a null distribution derived from randomly dropped out channels of equal numbers. Error bars represent SEM across participants.

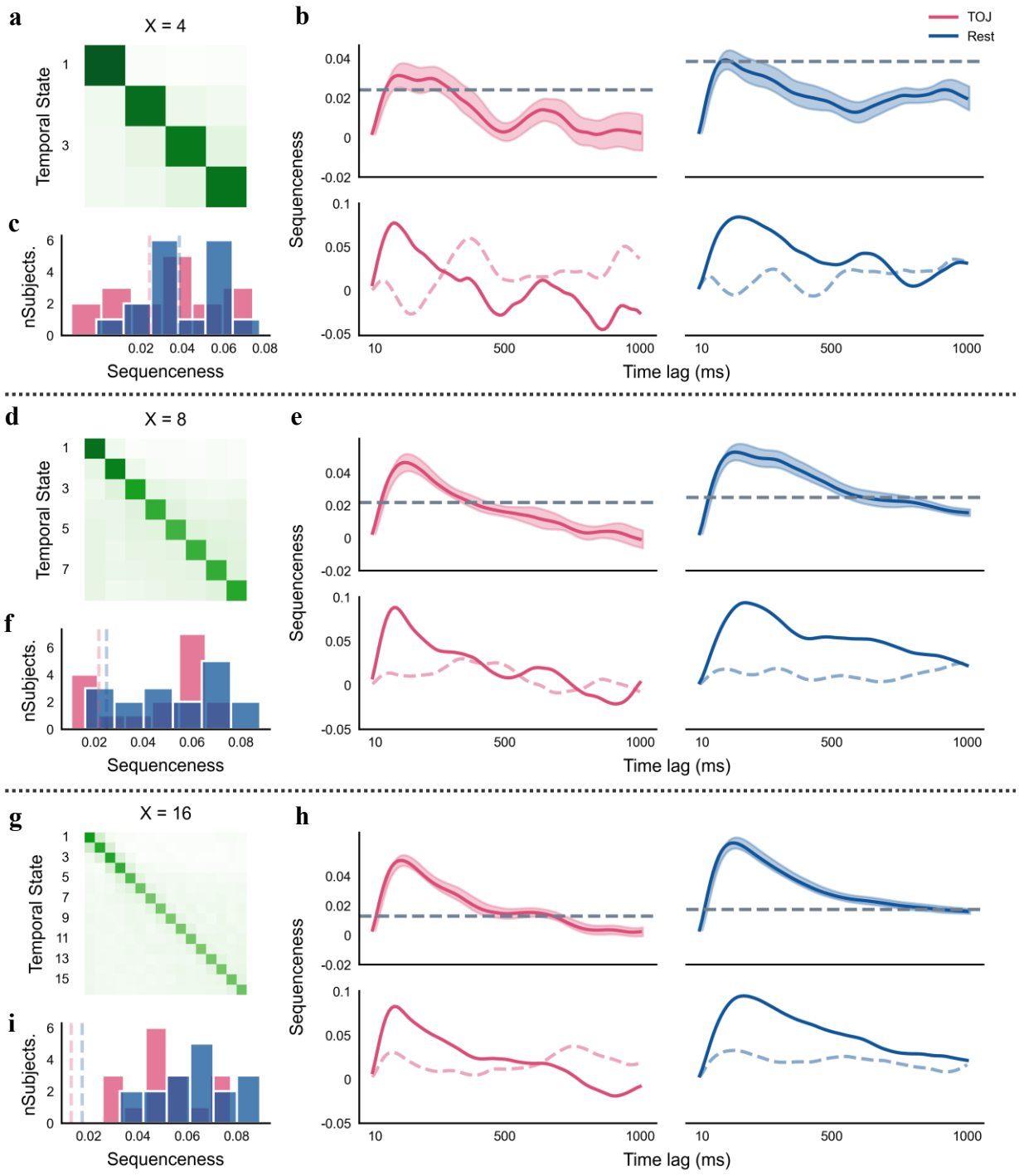

Figure S3. Sequenceness results with different temporal state bin widths, Related to Figures 1, 3

Sequenceness analysis was repeated with temporal states defined by bin widths of 1000 ms, 500 ms and 250 ms, which correspond to different temporal state numbers $X = 4, 8, and 16$, respectively. Evidence supporting temporal replay increased with finer temporal resolution.

(a, d, g) Decoding accuracy. Time classifiers significantly exceeded chance levels, achieving accuracies of $87.58 \pm 1.94\%$ (a, vs. chance: 25%), $65.20 \pm 3.01\%$ (d, vs. chance: 12.5%) and $45.50 \pm 2.35\%$ (g, vs. chance: 6.25%). Confusion matrices illustrate classifiers' performance, with darker colors indicating higher probabilities.

(b, e, h) Sequenceness plot. Temporal replay was observed across all conditions during TOJ and rest periods. Group-level results (upper panels) showed consistent peak lag across different bin widths, ranging between 110-140 ms (on task replay: 110, 120, 120 ms; off-task replay: 110, 130, 130 ms). Gray dashed lines represent permutation-derived thresholds and shaded areas indicate SEM across participants. Lower panels show examples

from one ”good” (solid line) and one ”bad” (dashed line) participant in each condition.
(c, f, i) Distribution of sequenceness across participants. At $X = 4$, 10 participants showed significant on-task replay, and 8 showed off-task replay (c). At $X = 8$, 13 participants showed significant on-task replay, and 15 showed off-task replay (f). All participants showed significant sequenceness for both on-task and off-task replay at $X = 16$ (i). 17 participants included. Dashed lines indicate group-level significance thresholds, corresponding to gray dashed lines in (b,e,h).

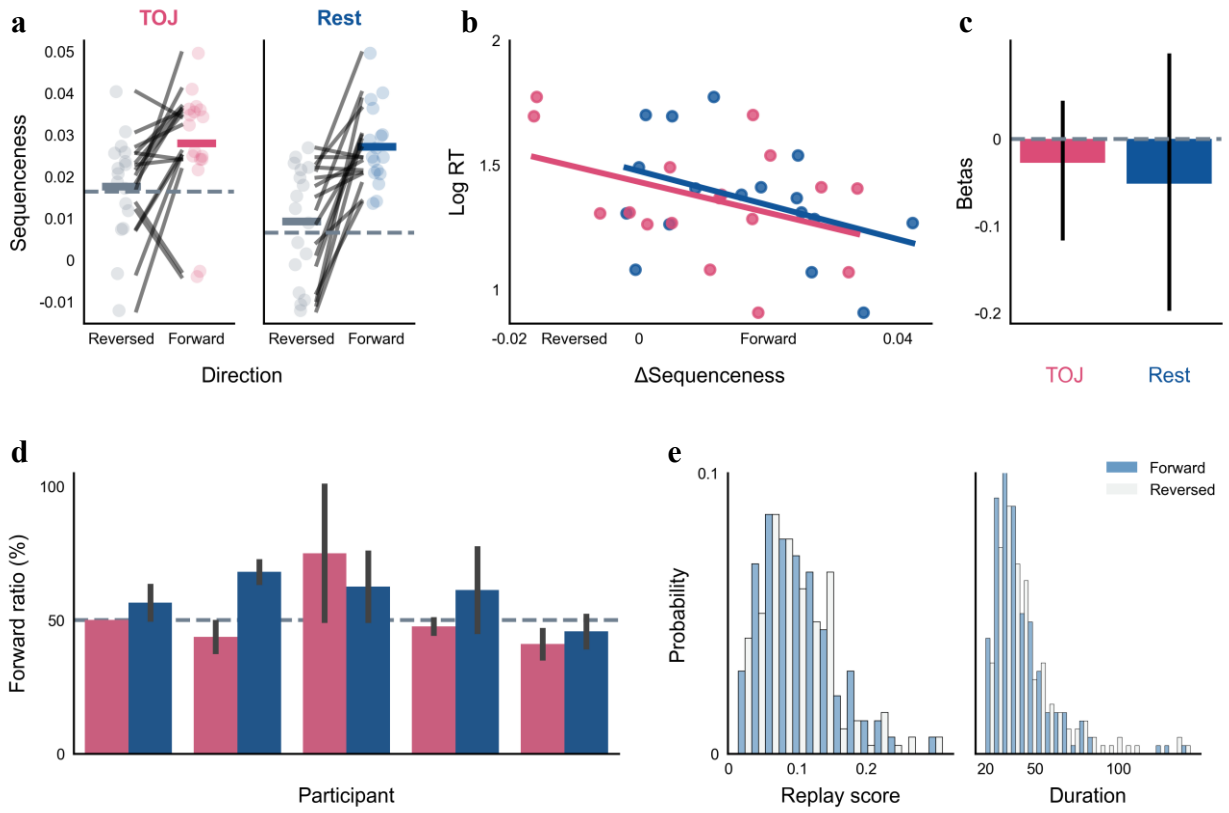

Figure S4. Replay events in forward and reversed directions, Related to Figures 3, 4

a. Replay evidence in both forward and reversed directions. On-task replay occurred at a time lag of reverse: 100 ms and forward: 120 ms; off-task replay occurred at reverse: 100 ms and forward: 140 ms. Forward replay showed significantly greater sequenceness than reversed replay ($p < 0.01,\ \chi^2(1) = 24.74$, LRT). Thresholds were computed separately for forward and reversed replay, and the gray dashed lines indicate the greater ones.
b. Relationship between replay directionality and reaction times. Replay directionality was defined as the difference in sequenceness between forward and reversed direction. Positive $\Delta Sequenceness$ indicates forward dominance, while the negative $\Delta Sequenceness$ indicates reversed dominance in replay events. Despite a numerical trend, no significant correlation was found between replay directionality and participants' mean reaction times. (On-task replay: $p = 0.11,\ r = -0.40$; off-task replay: $p = 0.078,\ r = -0.44$, Pearson's correlation).
c. Trial-by-trial effects of replay directionality on reaction time. Multilevel regression analysis revealed no significant relationship between replay directionality and response times at the single-trial level (On-task: $\beta = -0.028$, $[-0.12,\ 0.044]$; Off-task: $\beta = -0.051,\ [-0.20,\ 0.099]$). Error bars represent 95% confidence intervals derived from 1000 bootstrapped samples
d. Ratio of forward ripple-associated replay. Bars above the gray dashed line (50%) indicate a higher proportion of forward replay events than reversed ones. The overall ratio of forward replay (52.35%) was numerically higher than reversed replay (47.65%), but no significant effect of replay direction was detected ($p = 0.35,\ \chi^2(1) = 0.89$, LRT). Error bar represents SEM computed across sessions for each participant.
e. Distribution of ripple-associated replay scores and durations. Forward and reversed replay during hippocampal SWRs showed similar distributions for replay scores ($p = 0.84,\ D = 0.065$) and duration ($p = 0.62,\ D = 0.079$, two-sample Kolmogorov-Smirnov test). Replay score was calculated as the absolute value of weighted linear correlation. Data were pooled across on-task and off-task conditions.

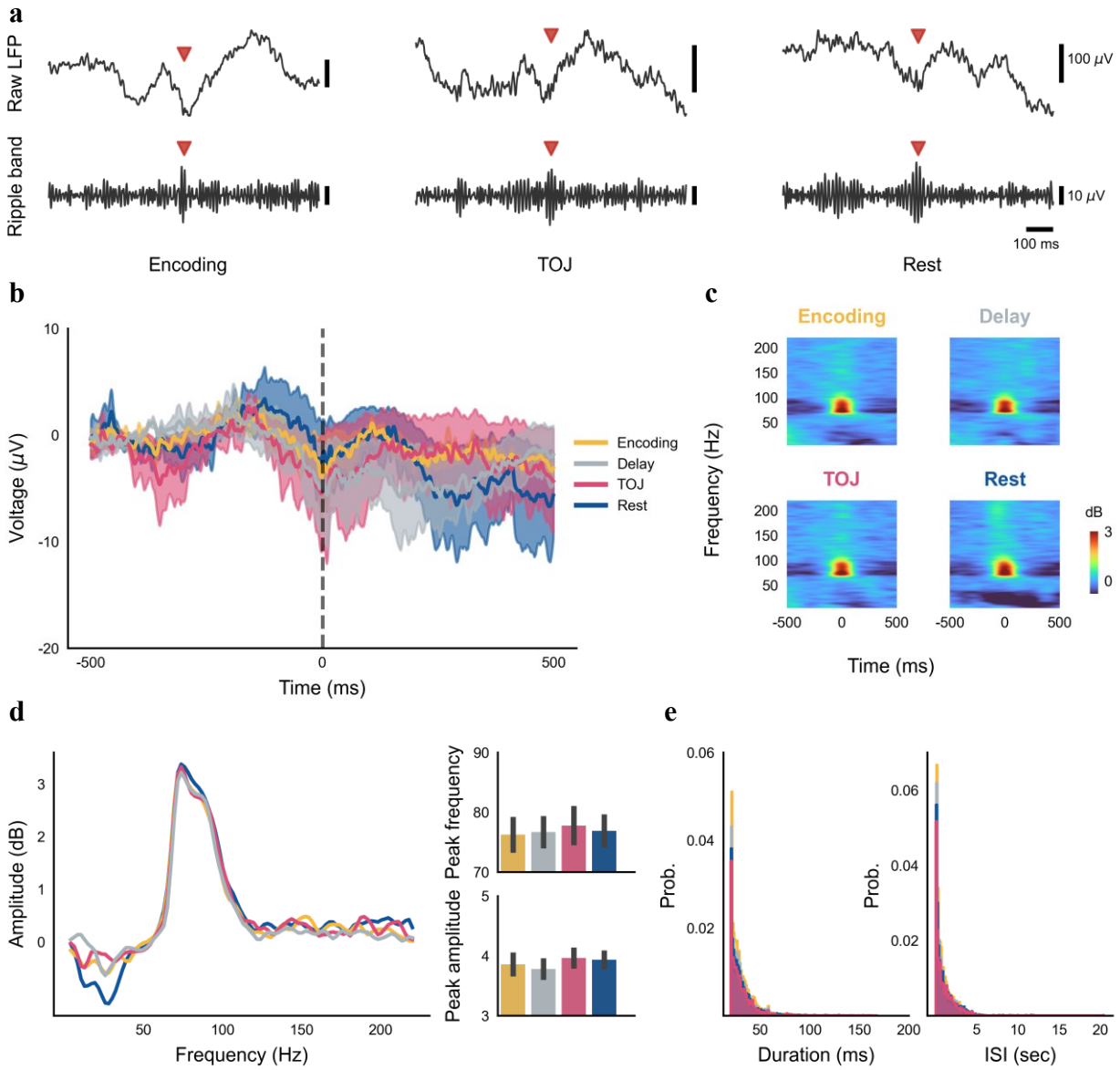

Figure S5. Spectral properties and temporal distribution of SWRs across task phases, Related to Figure 4

a. Examples of individual SWR events. Ripples recorded during video viewing (left), memory retrieval (middle), and resting-state (right) phases from a representative participant. SWR peaks are marked by red triangles.
b. Peri-ripple field potential. Grand-averaged LFPs aligned to ripple peaks for four conditions: Encoding (video viewing), Delay, TOJ (memory retrieval), and Rest (resting-state). Shaded areas represent SEM across participants (n = 5).
c. Peri-ripple spectrogram. Wavelet spectrograms centered on ripple peaks for each condition, normalized by the mean power in each frequency across epochs within the same condition.
d. Spectral characteristics at ripple peaks. Normalized spectra at ripple peaks reveal consistent spectral signatures across conditions. No significant differences were observed in peak frequency ($p = 0.72,\ \chi^2(3) = 1.35$) or amplitude ($p = 0.22,\ \chi^2(3) = 4.44$, Friedman test) .
e. Distribution of ripple durations and inter-ripple intervals. Duration and inter-ripple intervals (ISI) were presented separately for each condition, revealing no significant differences across conditions (duration: $p = 0.13,\ \chi^2(3) = 8.57$; inter-ripple intervals: $p = 0.28,\ \chi^2(3) = 6.29$, Friedman test). Inter-ripple intervals were computed within trials for each condition.

Patient #	Sex	Age	Total sessions	Total sites	Hippocampus sites
1	M	20	4	112	Y
2	F	14	3	134	N
3	F	26	4	132	N
4	M	28	2	148	N
5*	M	25	1	140	N
6	F	19	5	120	N
7	M	28	5	126	N
8	M	37	7	48	N
9	M	23	3	116	N
10	F	12	4	136	N
11	M	25	5	106	N
12	M	34	5	202	Y
13	M	39	5	172	N
14	M	27	5	146	Y
15	M	24	4	184	Y
16	M	12	5	144	Y
17	F	35	5	126	N
18	M	16	4	118	N

Table S1. Patient demographics

* Patient #5 was excluded from the analysis due to poor time decoding accuracy.