Transformation of Speech Sequences in Human Sensorimotor Circuits

Kathrin Müsch[1*], Kevin Himberger[1], Kean Ming Tan[2], Taufik A. Valiante[3], Christopher J. Honey[1]

[1] Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

[2] School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

[3] Toronto Western Hospital, Krembil Research Institute, Toronto, ON M5T 2S8, Canada

*Corresponding author: Kathrin Müsch, kmuesch1@jhu.edu

Author contributions (CRediT taxonomy): Conceptualization, K.M. and C.J.H.; Methodology, K.M. and C.J.H.; Formal Analysis, K.M., K.M.T., and C.J.H.; Investigation, K.M. and K.H.; Resources, K.M., K.M.T., T.A.V., and C.J.H.; Writing – Original Draft, K.M., K.H., K.M.T., T.A.V., and C.J.H.; Visualization, K.M., K.M.T., and C.J.H.; Supervision, T.A.V. and C.J.H.; Funding Acquisition, K.M., K.M.T., T.A.V. and C.J.H.

Conflict of interest: The authors declare no competing financial interests.

1 Summary

2   After we listen to a series of words, we can silently replay them in our mind. Does this mental

3   replay involve a re-activation of our original perceptual representations? We recorded

4   electrocorticographic (ECoG) activity across the lateral cerebral cortex as people heard and then

5   mentally rehearsed spoken sentences. For each region, we tested whether silent rehearsal of

6   sentences involved reactivation of sentence-specific representations established during

7   perception or transformation to a distinct representation. In sensorimotor and premotor

8   cortex, we observed reliable and temporally precise responses to speech; these patterns

9   transformed to distinct sentence-specific representations during mental rehearsal. In contrast,

10 we observed slower and less reliable responses in prefrontal and temporoparietal cortex; these

11 higher-order representations, which were sensitive to sentence semantics, were shared across

12 perception and rehearsal. The mental rehearsal of natural speech involves the transformation

13 of time-resolved speech representations in sensorimotor and premotor cortex, combined with

14 diffuse reactivation of higher-order semantic representations.

15 Keywords: ECoG, sentence repetition, verbal short-term memory, subvocal rehearsal

1 Introduction

2 Immediately after hearing a series of words, we can silently replay them in our minds. What

3 neural processes support this mental replay? Speech comprehension involves phonological,

4 syntactic and semantic processing across widespread circuits in temporal, frontal and parietal

5 cortex (Binder et al., 2009; Hickok and Poeppel, 2007; Pallier et al., 2011; Patterson et al., 2007;

6 Pei et al., 2011), but experiments manipulating the load and rate of verbal rehearsal have

7 implicated a smaller core of regions in verbal short-term memory (Fegen et al., 2015). These

8 core areas, involved in both speech perception and production, include the posterior temporal

9 cortex, motor and premotor areas, and the inferior frontal gyrus.

10 Within the regions implicated in verbal short-term memory, what kind of neural process

11 supports the replay of recent speech? A natural hypothesis is that mental replay arises from

12 neural replay: when we replay a series of words in our minds, the same neural populations may

13 be activated as during the original auditory perception. This "shared representation" hypothesis

14 is consistent with the common observation that activity patterns from perception may remain

15 continuously active during a memory delay period or may be re-activated following periods of

16 inactivity (Lewis-Peacock and Postle, 2008; Mongillo et al., 2008; Stokes, 2015). More generally,

17 "reactivation" of complex sequences of perceptual input is observed during vivid imagery of

18 those sequences (Buchsbaum et al., 2012). A shared representation for hearing and rehearsing

19 speech would also be consistent with "mirror" models, in which the imitation of speech actions

20 is supported by a common set of neurons across perception and production (D'Ausilio et al.,

21 2009; Rizzolatti and Craighero, 2004).

22 An alternative hypothesis is that, when we silently rehearse a series of words, we employ

23 representations that are distinct from those involved in the original auditory perception. In

24 distinction with early observations of shared representations (Pulvermüller et al., 2006), recent

25 intracranial and imaging studies have found that ventral sensorimotor circuits respond with

1   distinct activity patterns during the perception and production of the same syllables (e.g., "ba",

2   Cheung et al., 2016; Arsenault and Buchsbaum, 2016). Moreover, widespread bilateral cortical

3   circuits appear to "transform" between sensory and motor representations when pseudowords

4   (e.g., "pob") are held in mind and spoken aloud (Cogan et al., 2014). Thus, the process of

5   mentally rehearsing an entire sentence may be supported by circuits that transform between

6   sensory and motor representations.

7   We set out to determine which representations were shared and which were transformed

8   during the perception and silent rehearsal of many seconds of natural speech. Functional

9   magnetic resonance imaging (fMRI) studies of sentence perception and production lack the

10   spatiotemporal resolution to map word-by-word brain dynamics at natural speech rates. Prior

11   electrocorticography (ECoG) studies have focused on rehearsal of individual items (e.g., single

12   syllables), lacking syntactic or semantic content and posing little demand on verbal short-term

13   memory. Here, we used ECoG to measure time-resolved neural activity across the lateral

14   surface of the human brain during the perception and silent rehearsal of natural spoken

15   sentences of 5-11 words.

16   The processes supporting verbal short-term memory (reactivation vs. transformation) may vary

17   according to which kind of information is being rehearsed. People are better at recalling

18   coherent strings of words than incoherent strings of words, and this could be explained by the

19   fact that surface features of sentences (e.g., their phonology) may be "regenerated" from

20   abstract features (e.g., semantic and syntactics) that are most readily extracted from coherent

21   sentences (Lombardi and Potter, 1992; Potter and Lombardi, 1990). This leads to the prediction

22   that semantically sensitive brain regions would exhibit a "reactivation" pattern across

23   perception and rehearsal (Bonhage et al., 2014). To test this prediction, we manipulated the

24   internal coherence and contextual meaning of the sentences that were rehearsed.

25   We observed the strongest joint activation across sentence perception and silent rehearsal

26   within the ventral sensorimotor cortex (vSMC), dorsal sensorimotor cortex (dSMC) and dorsal

4

1   premotor cortex (dPMC) of the left hemisphere. Furthermore, increased activation in these

2   areas during silent rehearsal predicted more accurate behavioral recall of the sentence content.

3   Consistent with prior literature (e.g., Cheung et al., 2016; Glanz et al., 2018) the SMC and dPMC

4   responded rapidly during sentence perception, encoding sub-second properties of the input.

5   The fidelity of sensory responses in SMC and dPMC was exceeded only by the superior

6   temporal gyrus (STG) and middle temporal gyrus (MTG). When sentences were silently

7   rehearsed, SMC and dPMC again exhibited sentence-specific activity patterns, but the activity

8   patterns were distinct from those observed during perception of the same sentences.

9   Altogether, the data support a model in which "motor" circuitry (SMC and PMC) supports

10   verbal short-term memory via a sensorimotor transformation (Cogan et al., 2014).

11   We also observed sentence-specific activity in anterior prefrontal cortex (aPFC) and

12   temporoparietal cortex (TPJ). Sentence-specific activity in these areas was less temporally

13   precise and less reliable than in sensory or motor areas. However, patterns in prefrontal areas

14   were sensitive to the contextual meaning of the sentence being rehearsed. Moreover, the

15   representations in these high-level areas were not transformed, but were instead shared across

16   the perception and rehearsal of specific sentences. Activation in these higher order areas is

17   therefore consistent with a continuous activation or reactivation process supporting verbal

18   short-term memory.

19   Together, the data suggest that the sensorimotor cortex of the left hemisphere possesses both

20   the sensory and the motor representations required to act as an audio-motor interface

21   supporting short-term memory for natural speech sequences. "Core" speech rehearsal areas

22   may implement a sensorimotor transformation in support of verbal short-term memory, while

23   more distributed networks, sensitive to semantics, expressed a shared pattern of activity,

24   bridging perception and rehearsal.

25   # Methods

5

Participants

2 16 patients (11 female; 19-50 years old) were recruited from the Surgical Epilepsy Program at

3 Toronto Western Hospital (Ontario, Canada) from the pool of all patients evaluated for

4 neurosurgical treatment of medically refractory epilepsy. The clinical and demographic

5 information of participants is summarized in Table S1. Prior to any experimentation, all

6 participants provided informed consent, which was approved by the local Research Ethics

7 Board of the University Health Network Research Ethics Board and the Internal Review Board of

8 Johns Hopkins University. All procedures followed the Good Clinical Practice of the University

9 Health Network, Toronto.

Stimuli & Experimental Procedure

Experimental Task

12 Participants were asked to memorize and repeat sentences (Figure 1A-D). Each trial contained

13 three phases: perception, silent rehearsal, and production (Figure 1A). In the perception phase,

14 participants listened to a pair of sentences (sentence 1, S1, and sentence 2, S2). Then, in the

15 silent rehearsal phase, participants were asked to silently rehearse S2 once verbatim in their

16 mind, without mouthing it. Finally, in the production phase, participants were asked to vocalize

17 the passage verbatim, at the same pace that they had heard it. Visual symbols on the screen

18 cued participants to each phase.

1
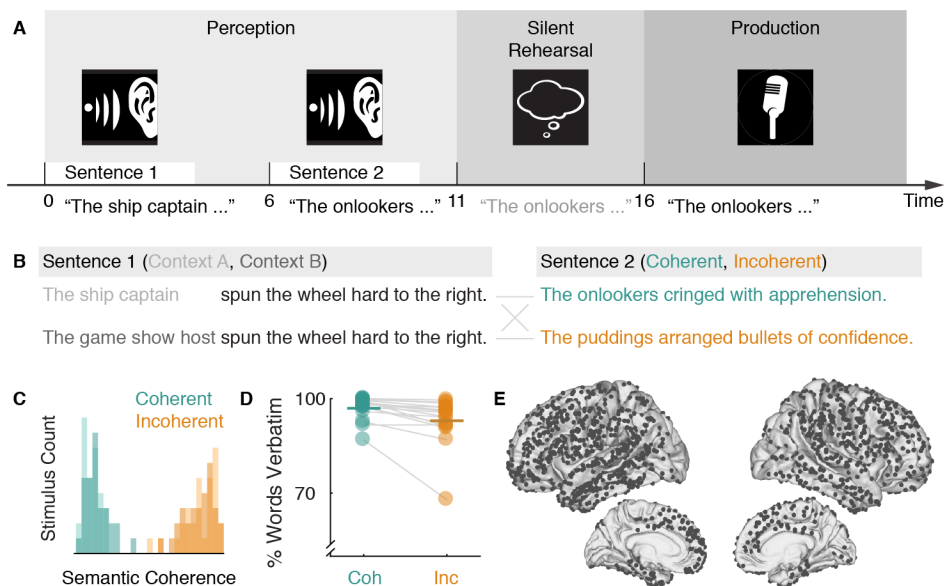
Figure 1. **Experimental Design and Behavioral Performance. A,** Participants first listened to two spoken sentences (sentence 1, sentence 2). They then silently rehearsed sentence 2 verbatim exactly once. Finally, they repeated sentence 2 aloud. **B,** A "stimulus group" consisted of four possible sentences: two versions of sentence 1 and two versions of sentence 2. Over the course of the entire experiment, participants would hear all four sentence 1/sentence 2 pairings. One of the sentences serving as sentence 2 was semantically coherent (teal), while the other was semantically incoherent (orange; also see Table S2). The two sentences serving as sentence 1 were both coherent but their first half differed in such a way that they would provide a very different semantic context for interpreting the coherent sentence 2. See also Table S2. **C,** Multidimensional Scaling of sentence norming data shows a clear separation of coherent and incoherent sentence combinations along the first dimension of the sentence norms. See also Table S3. **D,** Average number of words recalled verbatim for each participant when repeating coherent (Coh) and incoherent (Inc) sentences. **E,** Combined electrode placement for all 16 subjects on the lateral and medial surfaces of the Freesurfer average brain. See also Figure S1.

Stimulus Generation

We created 30 unique "stimulus groups" consisting of four sentences. Examples for the first sentence (S1) and second sentence (S2) are illustrated in Figure 1B, and all stimulus groups are listed in Table S2. The four sentences in a stimulus group were divided into two sentences that could serve as S1 and two sentences that could serve as S2. On a single trial, participants would hear a single pairing of S1 and S2, but across the experiment they would hear all four possible

1   pairings. Not all participants were able to complete all 30 stimulus groups; some participants

2   completed only 12 of the stimulus groups (set 1) while others completed an additional 18 of the

3   groups (set 2; see Table S2).

4   On a given trial, the semantic context and coherence were varied, depending on which

5   combination of S1 and S2 was presented. In half of the trials ("coherent"), S2 was semantically

6   coherent and was a natural semantic extension of S1 (e.g., Figure 1B, top row). In the other half

7   of trials ("incoherent"), S2 was semantically incoherent and did not have any obvious semantic

8   relationship with S1 (Figure 1B, bottom row). For both coherent and incoherent trials,

9   participants were asked to memorize and repeat S2 verbatim.

10   On coherent trials, the precise meaning of S2 depended on the contextual information

11   presented in S1. In particular, by changing only the initial words within S1, we varied the

12   interpretation of S2. Even the final words of S1 were shared across the two contexts. For

13   example, the subject of the sentence in Figure 1B ("ship captain" vs. "game show host")

14   determines whether the "apprehension" in S2 is understood as "suspense about a prize" or

15   "concern about danger.

16   For each stimulus set (set 1, set 2), we used a two-step procedure to match basic linguistic

17   properties of the coherent and incoherent S2 sentences. First, candidate versions of

18   semantically incoherent sentences were created using a bigram generator

19   (http://johno.jsmf.net/knowhow/ngrams/). Second, closed-class words were replaced so that

20   both sentences were matched for mean letter length (coherent: mean, M ± standard deviation,

21   SD = 4.5±1.1; incoherent: $M{\pm}SD$ 4.5±1.0; p = 0.795), log frequency (coherent: $M{\pm}SD$ = 12.5±0.9;

22   incoherent: $M{\pm}SD$ 12.6±0.8; p = 0.788), number of words (coherent: $M{\pm}SD$ = 8.1±1.5;

23   incoherent: $M{\pm}SD$ 8.2±1.3; p = 0.380) and number of syllables (coherent: $M{\pm}SD$ = 11.0±1.3;

24   incoherent: $M{\pm}SD$ 11.2±1.4; p = 0.184). Lexical parameters were derived from the English

25   Lexicon Project (http://elexicon.wustl.edu/default.asp). This procedure ensured that the syntax

26   of the incoherent sentences was largely preserved.

1   A male speaker recorded the sentences in mono in a soundproof chamber with an MXL USB

2   microphone (Torrance, CA). For the S1 sentences in the same stimulus group, the identical

3   audio waveform was pasted into the shared part of each S1 sentence (e.g., "spun the wheel

4   hard to the right" in Figure 1B). The moment when this identical acoustic waveform began in

5   each sentence slightly differed across the two S1 sentences. Furthermore, offline processing

6   included noise removal, cropping, volume normalization and conversion to stereo. S1 and S2

7   were each between 3 and 4 s in duration.

8   **Norming of Sentence Pairs**

9   An independent sample of 76 participants (40 females, 18-55+ years old) rated the 120

10   sentence pairs (30 stimulus groups with 4 combinations each) on eight dimensions using the

11   online platform Amazon's Mechanical Turk via psiTurk (https://psiturk.org/) for monetary

12   compensation. The eight dimensions were: understandability, complexity, surprise, empathy,

13   valence, arousal, visual vividness and auditory vividness (see Table S3 for the norming

14   questions). Each participant rated a subset of 15 of the 120 sentence pairs on all eight

15   dimensions (15 x 8 = 120 judgments). 2 participants were excluded because their rating data

16   were incomplete. Following informed consent, participants were instructed to listen carefully to

17   the audio playback of each sentence pair and make judgments using a slider on the screen.

18   They rated each dimension on a five-point scale and were encouraged to use the full scale for

19   their judgment. In four independent catch trials, participants had to type in the final word of

20   the audio clip to ensure that they were paying attention to the sentence pairs. Two-factor

21   analyses of variance were conducted on each dimension (Table S3). Coherent sentence pairs

22   were designed to elicit a more meaningful and concrete scenario in the listeners mind.

23   Consistent with this goal, they were easier to understand, less complex, more vivid in both

24   visual and auditory modality, more arousing, and more empathetic than incoherent sentence

25   pairs. Incoherent sentence pairs were more surprising than coherent sentence pairs. To

26   visualize the similarity among the different dimensions, we applied Multidimensional Scaling on

27   the 8-dimensional behavioral ratings of each of the 120 unique S1 and S2 combinations. The

1   first dimension clearly separated coherent and incoherent sentence combinations (Figure 1C),

2   confirming our manipulation of coherent and incoherent combinations.

3   **Experimental Procedure**

4   Stimuli were presented at bedside with a M-Audio M-Track Plus sound card (InMusic,

5   Cumberland, RI) on a Dell laptop (Round Rock, TX) using the Psychophysics Toolbox 3.0.12

6   (http://psychtoolbox.org/) and MATLAB R2014b (Mathworks, Natick, MA) about 60 cm from

7   their eyes. Set 1 was always presented first, while set 2 was presented if the participant was

8   willing and able to complete the second set. All sentence pairs were repeated once, resulting in

9   a total of 96 trials for set 1 and 144 trials for set 2. This allowed us to assess how reliable the

10   brain responded to a given sentence. The sets were divided into eight blocks of 12 trials (set 1)

11   or 18 trials (set 2). Each block contained all stimuli for a fixed S1 context (e.g., context 1). Half of

12   the S2 sentences in a block were coherent and half were incoherent. The order of sentence

13   pairs within each block was pseudorandomized such that no more than two coherent or

14   incoherent sentences appeared in a row and that the last three stimuli were not immediately

15   repeated in the first three trials of the next block. With two exceptions, all participants

16   completed set 1. Participant 2 (P2) only completed one repeat for each sentence pair of set 1

17   (48/96). P5 completed 36/48 sentence pairs for repeat 1 and 24/48 sentence pairs for repeat 2

18   of set 1. P6, P7, P8, P10, P11, P13, P14, P15 and P16 completed set 2. The recordings of P14 for

19   set 2 were discarded because of poor data quality (slow drift resulting in clipped recordings),

20   resulting in a total of eight patients completing both sets.

21   To ensure that participants attended to both sentences, we presented 15 catch trials. After a

22   completed trial the participants had to judge whether a presented picture matched with the

23   situation described by the entire sentence pair (the combination of S1 and S2) that they had

24   heard. Catch trials were only presented after coherent trials. On average, participants correctly

25   responded to 80.2±3.7% (*M* ± standard error of the mean, *SEM*) of the catch trials.

Data Acquisition & Electrode Localization

Signals were recorded from combinations of 64 contact grids, strips of 4-12 subdural platinum-iridium electrodes with 3-mm diameter and 10-mm inter-electrode spacing (PMT, Chanhassen, MN) and 4-contact Spencer depth electrodes (Ad-Tech Medical Instrument Corporation, Racine, WI). Electrode placement was solely based on clinical criteria (Figures 1E, S1). A 4-contact subgaleal strip electrode (PMT, Chanhassen, MN) electrode over the parietal midline and facing away from the brain was used for ground and reference. For data acquisition patient connectors were transferred to a separate research amplifier system (NeuroScan SynAmps2; Compumedics, Charlotte, NC). Data was recorded at 5 kHz and 0.05 Hz hardware band-pass filtering. Event markers from the stimulus presentation were sent to the SynAmps2 through a parallel port. The SynAmps2 also recorded the participant's overt spoken response and the surrounding acoustics through a Tube MP Studio V3 (ART ProAudio, Niagara Falls, NY). Electrodes were localized with the freely available iELVIS toolbox for MATLAB (http://ielvis.pbworks.com; Groppe et al., 2017). The co-registration procedure mapped the individual structural T1 image acquired before electrode implantation on the postimplant CT image and corrected for postimplant brain shift. All electrode maps are displayed on the Freesurfer average brain.

Data Analysis

Behavioral Data

Voice recordings from the task were used to manually transcribe the participant's verbal responses exactly as uttered. For each response, we compared the transcription to the target sentence. To compute a verbatim score, we counted the words that were identical to those in the original sentence (Meltzer et al., 2016). Words did not need to be in the exactly correct order for this criterion. Words were not included if they were not part of the original sentence.

1   Few words were recalled out of order (0.6%) or added to the sentence (2%, including

2   metacognitive statements such as "I think, I got that wrong"). To compute an accuracy score,

3   for each participant we computed the ratio of words recalled verbatim to the number of total

4   words within a sentence. We computed a paired t test (two-sided, $\alpha < 0.05$) to compare

5   performance in coherent and incoherent trials.

6   **Data Preprocessing**

7   Electrodes exhibiting artifacts, epileptiform activity, excessive noise or no signal were excluded

8   from the analysis by offline visual inspection. As the raw voltage signals of some recording

9   sessions in a given patient exhibited different means and variance, preprocessing was done for

10 each session separately (about 6-12 minutes of recording). Raw voltage signals were

11 downsampled to 500 Hz. To minimize filter artifacts, all filtering was performed on continuous

12 data. Data was high-pass filtered at 0.1 Hz (Butterworth, order 4) and line noise was removed

13 with a bandstop filter (Butterworth) from 59.5-60.5 Hz, 119.5-120.5 Hz and 179.5-180.5. Similar

14 to an average-rereferencing approach, the mean voltage signal across all remaining channels of

15 a given subject was projected out from each individual channel by using linear regression. The

16 voltage trace of each channel was z-scored, including only the trial periods for calculating the

17 mean and standard deviation. To avoid edge artifacts, filtering was always performed on longer

18 segments of continuous data.

19 Trial-level segmentation and baseline correction were performed after filtering. The filtered

20 signal was segmented into epochs of 1 s prestimulus and 23 s poststimulus intervals, except for

21 P1 who had a prestimulus interval of 0.5 s. We used a common trial baseline approach, in which

22 we excluded the highest and lowest 5% of trials to minimize the effects of outlier baseline

23 shifts: for each channel, the mean signal of the prestimulus interval was subtracted from each

24 sample in the poststimulus interval of each trial. Bad trials were excluded from baseline

25 correction (P6 and P10 had one bad trial each). To obtain the envelope of the acoustic data

1 channel, we filtered the signal between 200 and 2000 Hz, applied the Hilbert transform and

2 resampled the rectified and log-transformed signal to 500 Hz.

Computation of Power Time Courses

3

4 We estimated time courses of signal power using Morlet wavelets. In the frequency range 70-

5 200 Hz, power time courses were computed separately for each frequency in steps of 5 Hz. 120

6 and 180 Hz were excluded as harmonics of line noise. We then took the logarithm of each

7 power time course estimate. These estimates were then z-scored, again including only the trial

8 periods for calculating the mean and standard deviation. Finally, we computed the mean across

9 each of the frequency-specific z-scored time courses to yield a single "broadband" 70-200 Hz

10 power time course.

11 Because of the long duration of our trials, we observed transient motor-related artifacts due to

12 swallowing that contaminated all channels in the same frequency range as broadband power.

13 We removed this artifact in two steps. First, we generated a temporally sparse noise regressor

14 by identifying deviants from the median broadband power time course across all channels in a

15 recording session. At times when the median broadband power across channels was less than

16 three times the interquartile range from its median value, the noise regressor was set to zero;

17 at all other times (i.e. during bursts), the noise regressor was set equal to the mean broadband

18 power across channels. We then projected out this noise regressor from the continuous

19 broadband power time course by taking the residual from a linear regression of the noise

20 regressor on the broadband power signal in each channel. Before segmentation and baseline

21 correction, broadband power time courses were smoothed with a 100 ms Hamming window.

Comparison of Activation in Trial vs. Baseline

22

23 Activation within trials was compared to that during baseline within each participant with a

24 paired permutation test in sliding windows of 500 ms. The baseline window extended from -

13

1 750 to -250 ms (exception: P1 had a shorter baseline period, and their window extended from -

2 500 to 0 ms). For each window, we computed the "activation" as the mean of the broadband

3 power samples in that window. For each electrode, we then measured the difference between

4 activation during the trial and the baseline period. We assessed the significance of this

5 difference by comparing against a null distribution generated from randomly shuffling the

6 labels ("baseline" or "trial") 10,000 times. This procedure was computed separately for

7 coherent and incoherent sentences. The null hypothesis of no differences between trial and

8 baseline was rejected when the false discovery rate (FDR) across electrodes within a participant

9 $q < 0.01$ (Benjamini and Hochberg, 1995).

10 **Comparison of Correct vs. Incorrect Trials**

11 Based on the overt memory recall in the production phase, trials were grouped into "correct"

12 and "incorrect" bins: correct production of the sentence required an exact match to the original

13 sentence, while all other trials were marked as incorrect. Broadband power for correct and

14 incorrect trials were then compared with an independent-sample permutation test within each

15 participant based on Welch's t-statistic. All samples within a task phase were averaged (S1: 0-

16 4s; gap 1: 4-6s; S2: 6-10s; gap 2: 10-11s; silent rehearsal: 11-16s; production: 16-21s). For each

17 electrode, Welch's t-statistic for the observed difference between correct and incorrect

18 sentences was compared against a null distribution of Welch's t-statistic generated from

19 randomly shuffling the sentence labels ("correct" or "incorrect") 10,000 times. The null

20 hypothesis of no differences between conditions was rejected when $q < 0.05$ (FDR across

21 electrodes within a participant, as above).

22 **Comparison of Coherent vs. Incoherent Trials**

23 Coherent and incoherent trials were compared with a paired permutation test within each

24 participant. To rule out the possibility that activation differences could be driven by behavioral

25 differences, i.e., incorrect recall, we only included trials on which S2 was recalled verbatim

14

1 correctly or with a single small error (e.g., "a" vs. "the"). All samples within a task phase were

2 averaged (S1: 0-4s; gap 1: 4-6s; S2: 6-10s; gap 2: 10-11s; silent rehearsal: 11-16s; production:

3 16-21s). For each electrode, the observed difference between coherent and incoherent

4 sentences was compared against a null distribution generated from randomly shuffling the

5 sentence labels ("coherent" or "incoherent") 10,000 times. The null hypothesis of no

6 differences between conditions was rejected when $q < 0.05$ (FDR across electrodes within a

7 participant as above (Benjamini and Hochberg, 1995)).

8 **Reliability Analysis: Correlated Component Analysis**

9 To identify brain regions that encoded stimulus-specific information, we employed Correlated

10 Component Analysis (Dmochowski et al., 2012). This linear decomposition technique identifies

11 a set of weights across channels that will produce a weighted-average signal whose Pearson

12 correlation coefficient across repeats is maximal. Intuitively, this can be thought of as assigning

13 greater weighting to electrodes whose activity time courses are reliable across repeats of a

14 stimulus. This technique can also be understood as similar to principal component analysis;

15 instead of weighting electrodes such that the variance of a single repeat is maximized, this

16 technique weights electrodes such that the correlation between two repeats is maximized,

17 relative to the within-repeat correlation.

18 More formally, we define X_1 and X_2 as the matrices that contain the data in a given region of

19 interest (ROI) for the first and second repeats of all stimulus groups, respectively. Each matrix

20 has dimensions [C x D], where C is the number of channels in the ROI and D (= T x S) is the total

21 number of data samples obtained from concatenating T trials with S samples per trial.

22 Intuitively, the response time courses are more reliable across repeats when the diagonal

23 values of the [C x C] covariance matrix $X_1^T X_2$ are larger. In order to find a linear combination of

24 weights for each channel that maximizes the correlation across repeats, we solve the

25 generalized eigenvalue problem:

1 $$(1) \ \underset{w}{\text{maximize}} \frac{w^T R_B w}{w^T R_W w}$$

2   where $R_B = \frac{(X_1^T X_2 + X_2^T X_1)}{D-1}$ and $R_W = \frac{X_1^T X_1 + X_2^T X_2}{D-1}$ are the between-repeat covariance and the

3   within-repeat covariance, respectively. The matrices R_W and R_B were subsequently

4   symmetrized by taking the mean of each matrix with its transpose. To solve the generalized

5   eigenvalue problem in a manner that is more stable, we used a shrinkage procedure to

6   regularize the pooled within-repeat covariance matrix R_W (Blankertz et al., 2011). We define

7     $R_{W\_REG} = (1 - \gamma) R_W + \gamma \lambda I$, where λ is the mean eigenvalue of R_W, the matrix I is the identity,

8   and $\gamma = 0.1$ is a shrinkage parameter. Thus, to estimate the weights w we solve the generalized

9   eigenvalue problem with R_{W_REG} in the denominator and R_B in the numerator, thereby

10   maximizing the across-repeat correlation relative to the within-repeat correlation. The [*C* X *1*]

11   generalized eigenvector with maximum generalized eigenvalue then provides the optimal

12   weighting across the *C* channels which maximizes ρ.

13   *Cross-Validated Correlation Across Repeats and Within Task Phases*

14   To test reliability of responses, we employed a cross-validation procedure: weights that

15   maximize correlation were identified in a subset of the stimulus groups, and then the reliability

16   of the weighted signal was estimated in held-out stimulus groups. To maximize the power and

17   stability of this procedure, we only considered data from the subjects (N=8) who completed

18   both repeats of all 30 stimulus groups totaling 240 trials. Both correct and incorrect trials were

19   included, because the analysis requires a pair (first and second repeat) for each trial. The

20   electrodes from these eight subjects were pooled and grouped into nine ROIs based on the

21   Freesurfer parcellation of the Destrieux atlas (Destrieux et al., 2010). These ROIs included STG,

22   MTG, anterior temporal lobe (ATL), TPJ, vSMC (z<40), dSMC (z≥40), inferior frontal gyrus (IFG),

23   aPFC (y≥23) and dPMC (y<23). Figure 2C shows the grouping of electrodes into ROIs and their

24   average time courses for all 16 subjects. Table S4 lists the areas of the Destrieux atlas for each

25   ROI. To be included in the pool of electrodes for this analysis, electrodes needed to show

16

1 activity significantly different from baseline within at least one of the task phases (including

2 gaps) in at least one of the conditions (coherent, incoherent).

3 Using a sub-sampling and cross-validation approach, weights were estimated for two thirds of

4 the stimulus groups within a given task phase (S1, S2, silent rehearsal) and then tested on the

5 remaining stimulus groups (Figure 5A-B). Weights were estimated separately for coherent and

6 incoherent sentences. Thus, weight estimation sampled 40 sentences and their repeats (in-

7 sample data: random selection of 20 stimulus groups x 2 S1 contexts x 2 repeats). Weights were

8 then tested on the 20 held-out sentences and their repeats (out-of-sample data: the remaining

9 10 stimulus groups x 2 S2 contexts x 2 repeats). This procedure was repeated 1,000 times with

10 different samples of stimulus groups to obtain a more robust estimate of repeat reliability. To

11 avoid stimulus-specific reliability being driven by the mean response to each sentence, we

12 normalized each trial by subtracting the mean and dividing by the standard deviation of its

13 particular condition (sentence type, context, repeat). The normalization was performed

14 separately for in-sample and out-of-sample data. Similar to subsampling across time, we

15 averaged data across 200 ms bins. To feed a single time series for each channel into the

16 Correlated Component Analysis, we concatenated all trials for a given channel.

17 After estimating w for a given ROI and task phase within in-sample data (resulting in maximum

18 ρ for in-sample data), we assessed the out-of-sample reliability by measuring the correlation ρ

19 in held-out data from the same task phase. In addition to measuring true repeat reliability,

20 where each sentence is paired with its exact repetition in the held-out data, we also computed

21 the correlation using the same weights but with trial-ordering shuffled in the matrix X_2. This

22 shuffled correlation provides a null distribution of ρ, against which we can compare the

23 empirically observed, sentence-specific ρ values.

24 To assess whether the contextual manipulation (the two contexts of S1 within each stimulus

25 group) had any effect on brain responses later in each trial, we also computed the out-of-

17

1   sample correlation using the same weights, but with the responses in X_2 swapped across the

2   two versions of S1 within each stimulus group.

3   Overall, for each ROI we obtained 1,000 estimates of electrode weights w and 1,000

4   corresponding estimates of repeat-correlation values ρ for the true repeats, the null (shuffled)

5   data, and for the repeats in which we swapped the S1 context.

6   *Cross-Validated Correlation Across Repeats and Across Task Phases*

7   In some analyses (Figures 7, 8, S4B-C), the estimated weights were also applied to held-out data

8   from a different task phase. For those cases, we averaged all 1,000 estimates of weights w

9   before we applied them to the held-out data. Since the sign of the weight is arbitrary, we first

10   used principal component analysis to find the principal vector of weights and then multiplied all

11   weight vectors that were opposite to that one with -1 before averaging across all weight

12   vectors. Since the test data for this analysis is from a different task phase, it is completely

13   separate from the data used to estimate the weights; thus, we were able to estimate the ρ

14   values across task phases by subsampling from all 30 stimulus groups.

15   *Statistical Assessment of Correlated Component Analysis*

16   We assessed whether the distribution of the maximum correlation values from the 1,000 draws

17   was stimulus-specific and significantly different from zero. To this end, we compared the true

18   repeats of the held-out data for a given ROI against the shuffled repeats of the held-out data in

19   a two-sided paired test for each ROI. Specifically, we computed the difference between the

20   maximum correlation values of each fold and tested whether the 2.5[th] and 97.5[th] quantile

21   contained 0. To assess an effect of the context of S1, we also compared the maximum

22   correlation values of the true repeats with those in which the context was swapped. This test

23   was performed one-sided because we hypothesized that the identical context elicited more

24   reliable responses than a mismatched context. Finally, we compared coherent and incoherent

18

1   sentences in a two-sample test by randomly sampling the maximum correlation values for each

2   of the conditions 2,000 times, computing the difference, and testing whether the 2.5th and

3   97.5th quantile contained 0.

4   **Cross-Temporal Analysis**

5   To identify whether neural responses were specific to a particular moment within each

6   stimulus, we implemented a cross-temporal analysis. This approach was inspired by cross-

7   temporal decoding methods developed in the working memory literature (King and Dehaene,

8   2014). Intuitively, this analysis asks: if a particular ROI has an elevated weighted response at

9   time t_1 on the first repeat, will the response also be elevated at time t_2 within the second

10   repeat? If this is true, only when $t_1 = t_2$, then the neural response is specific not only to the

11   stimulus but also specific to a particular moment within the stimulus. On the other hand, if the

12   responses are correlated across repeats for all pairs of time-points, then the response is

13   stimulus-specific but not timepoint-specific.

14   For each ROI we generated a "response time course" weighted according to the weights w from

15   the Correlated Components Analysis. The response time course was generated using held-out

16   data in each ROI both for the perception phase (listening to S2) and for the silent rehearsal

17   phase, with the weights w estimated from the in-sample data for each phase. More formally,

18   for each time point t within a trial and for each repeat i we have a matrix $Y_i(t)$ of the trial x

19   channel data at that moment for that repeat. We compute a matrix of weighted time-series

20   $Z_{i(t)} = Y_i(t)w$, which contains the weighted neural response in that ROI for each trial. We then

21   correlate the weighted responses of repeat 1 and repeat 2 across trials. This correlation across

22   trials can be computed for any pair of time-points t_1 and t_2 leading to a time-time matrix of

23   Pearson correlation values. If the neural response is time-point specific, then the correlations

24   will be strong only along the diagonal of the time-time matrix, where $t_1 = t_2$. This time-time

25   correlation procedure was repeated for the held-out data of each fold of subsampling

26   procedure. The resulting time-time matrices were averaged across folds and sentence type

1   (coherent, incoherent). Finally, the correlation matrices were symmetrized by taking the mean

2   of the correlation of repeat1-to-repeat2 and the correlation of repeat2-to-repeat1.

3   *De-Noising of Time-Time Correlation Maps*

4   For visualization purposes, we performed matrix de-noising on the time-time maps using the 2-

5   d fused lasso procedure (Tibshirani and Taylor, 2011). Briefly, this procedure estimates the

6   original time-time maps with the constraint that many adjacent elements in the estimated

7   time-time maps should be similar to each other. Suppose that C is the original time-time

8   correlation matrix, of dimension T by T, where T is the number of timepoints. The 2-d fused

9   lasso procedure solves the following optimization problem:

10
$$(2) \; \underset{U}{\text{minimize}} \sum_{t=1}^{T} \sum_{s=1}^{T} (C_{ts} - U_{ts})^2 + \lambda \left[\sum_{t=2}^{T} \sum_{s=1}^{T} |U_{ts} - U_{(t-s)s}| + \sum_{t=1}^{T} \sum_{s=2}^{T} |U_{ts} - U_{t(s-1)}| \right]$$

11   where λ is a tuning parameter that controls the number of adjacent elements that are similar to

12   each other, and U is the estimated time-time map. We fit the model with several different

13   values of λ, and the resulting matrices are plotted as heatmaps with $\lambda = 0.03$ as the best

14   balance of smoothing and structure preservation (Figures 6, S6).

15   # Results

16   **Recall Behavior**

17   We measured neural responses while 16 participants performed a sentence repetition task. The

18   participants listened to a pair of sentences and were asked to mentally rehearse the second

19   sentence and then reproduce it aloud (Figure 1A). Half of the sentences to be remembered

20   were semantically coherent, while the other half consisted of nonsense sentences that were

1   semantically incoherent (Figure 1B, Table S2). Participants were able to accurately reproduce

2   both coherent and incoherent sentences in their overt recall (Figure 1D), with coherent

3   sentences slightly better recalled than incoherent sentences (coherent: 96.9% verbatim, *SEM*:

4   0.9; incoherent: 93.0%, *SEM* 1.9; $t(15) = 3.4$, $p = 0.004$).

5   **Broadband ECoG Activity During Sentence Perception and Rehearsal**

6   To characterize neural activity during sentence perception and rehearsal, we focused on

7   changes in broadband power (70-200 Hz) of the local electrical field measured by subdural EcoG

8   electrodes. Amplitude changes in this frequency band are a robust estimate of population firing

9   (Manning et al., 2009; Miller et al., 2010; Ray and Maunsell, 2011), provide reliable responses

10   to audiovisual stimuli across much of the lateral cortex (Honey et al., 2012) and are sensitive to

11   speech perception and production (Cogan et al., 2014; Crone et al., 2001; Flinker et al., 2015).

12   Aggregating data across participants resulted in dense coverage of the cortical surface in left

13   and right hemispheres, excepting only the occipital cortex (Figure 1E).

14   **Patterns of Mean Activation During Perception, Silent Rehearsal and Production**

15   Which regions of the brain were active during perception, which were active during rehearsal,

16   and which were active during both perception and rehearsal? To quantify the overall activation

17   patterns, we contrasted broadband power responses during each task phase against a pre-trial

18   baseline (lateral view, Figure 2B, medial view, Figure S2A; see Methods). For S1, broadband

19   power increased above baseline over left and right superior temporal gyrus, middle temporal

20   gyrus, dorsal motor and premotor cortex, as well as in the left inferior frontal gyrus and

21   anterior frontal cortex. The observation of both sensory and motor activity during passive

22   sentence perception is consistent with prior fMRI and EcoG studies (Fedorenko et al., 2016;

23   Honey et al., 2012; Humphries et al., 2007; Lerner et al., 2011). When participants listened to

24   S2, which they would need to subsequently rehearse, the pattern of activation was similar, with

25   one clear difference: a "ring" of below-baseline activity across widespread electrodes in inferior

1   temporal, parietal and dorsal frontal regions. This ring of below-baseline activity was most

2   pronounced in the left hemisphere, and was not observed during S1, suggesting that it was

3   related to participants' active attention as they prepared for the subsequent rehearsal of S2.

4   Silent rehearsal produced an activation pattern that was different from the perception phases:

5   broadband activity remained above baseline only in the vicinity of the left motor and premotor

6   cortex as well as in a very small number of lateral and inferior frontal electrodes bilaterally, and

7   in one left posterior temporal site. Widespread electrodes across frontal, temporal and parietal

8   cortex reduced their activity below baseline. To the extent that increases in broadband power,

9   an estimate of population firing, index functional activation (Miller et al., 2012), these data

10   most strongly implicate the left motor and premotor cortices in silent rehearsal.

11   The same patterns observed in the trial-averaged activity were present in time-resolved single-

12   trial responses (Figure 2A). For example, a typical electrode in the superior temporal gyrus

13   exhibited broadband power modulations that reflected auditory responses: they increased

14   during the presentation of each sentence, showed little modulation during silent rehearsal and

15   increased again during production, when the participant heard their own voice (Flinker et al.,

16   2010). An example electrode in ventral motor cortex expressed slightly elevated activity during

17   both perception and rehearsal, with the strongest modulations during production. An electrode

18   in the premotor cortex, the area active during both perception and rehearsal, exhibited

19   consistent single trial responses during all phases. Consistent with the observations of Cheung

20   et al. (2016), broadband power responses in the superior temporal gyrus and premotor cortex

21   tracked the stimulus acoustics. For example, the sample electrodes shown in Figure 2A

22   exhibited single-trial correlations with the auditory envelope of the sentence of r=0.47

23   (superior temporal gyrus) and r=0.57 (premotor cortex).

24   To visualize the time-resolved activation patterns at a regional level, we grouped the electrodes

25   into nine anatomically-defined ROIs (Figure 2C; see Methods). Electrodes exhibiting above or

26   below baseline activity in at least one task phase in any stimulus condition (coherent or

1   incoherent) were included in the regional summary. The average time courses of the ROIs

2   reflect the overall activation pattern in Figure 2B. Although we observed above-baseline

3   responses during sentence perception, broadband power trended towards below-baseline just

4   until the onset of production in all ROIs (Figure 2C).
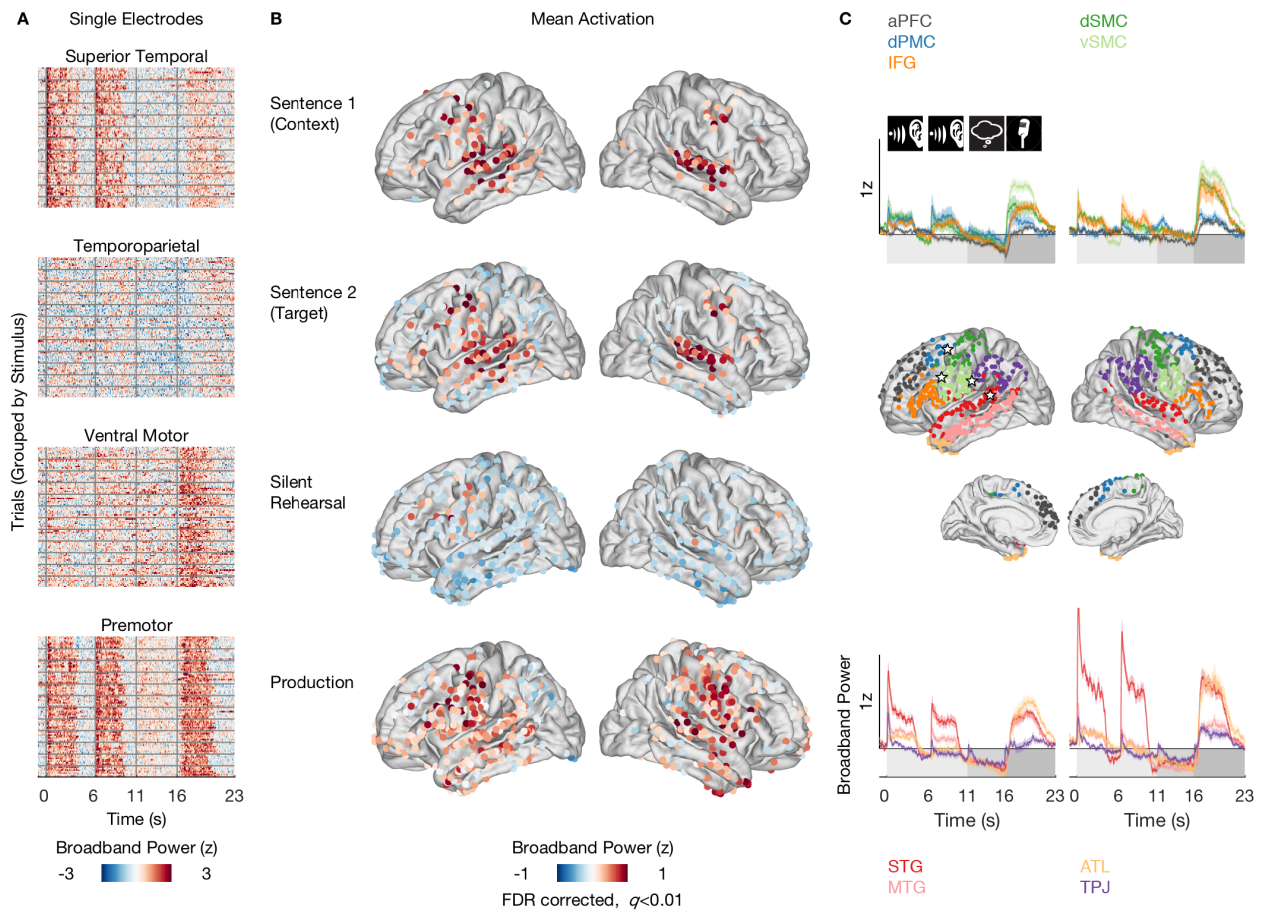
6   Figure 2. **Patterns of Mean Activation During Perception, Silent Rehearsal and Production. A**,
7   Stacked single trial responses for example electrodes. Responses are grouped by the twelve
8   stimulus groups shown. Broadband power changes were consistent across trials within an
9   electrode but differed across brain regions. Electrode locations are indicated by stars (2C,
10   middle). Onset of sentence 1, sentence 2, silent rehearsal and production are indicated by
11   vertical gray bars. **B**, Spatial distribution of broadband power responses aggregated across
12   participants and conditions for each task phase in the left and right hemispheres. Only
13   electrodes activated at the level $q < 0.01$ after FDR correction are shown. See also Figure S2. **C**,
14   Average time courses and electrode groupings within regions of interest (ROI), split by left and
15   right hemisphere. The top panel shows the average time courses for anterior prefrontal cortex

1 (aPFC, gray), dorsal premotor cortex (dPMC, dark blue), ventral sensorimotor cortex (vSMC,
2 light green), dorsal sensorimotor cortex (dSMC, dark green), inferior frontal gyrus (IFG, orange).
3 The bottom panel shows the average time courses for superior temporal gyrus (STG, red),
4 middle temporal gyrus (MTG, pink), anterior temporal lobe (ATL, yellow) and temporoparietal
5 junction (TPJ, purple). Onset of perception (sentence 1, sentence 2), silent rehearsal and
6 production are indicated by icons and shading. See also Table S4.

Sensorimotor and Premotor Sites Coactive During Perception and Silent Rehearsal

8 Neural circuits that exhibit increased activity during both perception and silent rehearsal of a

9 sentence are candidates for supporting the short-term memory of that sentence. Across

10 participants, motor and premotor sites appear to be the areas most consistently involved in

11 both perception and silent rehearsal phases of the task (Figure 2B). To spatially refine this

12 finding, and to rule out the possibility that perceptual and rehearsal activations occurred in

13 different subsets of participants, we measured the conjunction of sites within patients that

14 were statistically active above or below baseline in both task phases (149 sites in total). This

15 analysis confirmed that almost all sites active above baseline in both task phases (17 of the 20

16 sites) were clustered around sensorimotor and premotor cortex of the left hemisphere (Figure

17 3A, red electrodes). Three isolated exceptions were observed in the left posterior superior

18 temporal cortex, the right middle superior temporal gyrus and the right inferior frontal gyrus.

19 Sites that activated during S2 and deactivated during rehearsal (36 sites, green electrodes)

20 clustered around superior and middle temporal cortex. Sites that were deactivated in both task

21 phases (92 sites, blue electrodes) were mostly distributed around temporal and motor cortex

22 (see deactivated sites during S2 in Figure 2B). Finally, there was one site in left frontal cortex

23 that was activated during rehearsal but deactivated during perception (gray electrode). The

24 overall results of this analysis are (i) joint perception and rehearsal activation in sensorimotor

25 and premotor areas (see also, Glanz et al., 2018; Towle et al., 2008), and (ii) widespread

26 suppression of rehearsal activity in a ring of surrounding temporal and parietal sites.

Motor and Premotor Activation Levels Predict Behavioral Accuracy

24

1   If a region encodes and maintains sentence-related information, its activity levels should be

2   correlated with the accuracy of subsequent memory-dependent behavior. We therefore

3   assessed the accuracy of memory-dependent behavior using the overt speech that participants

4   produced in the production phase at the end of every trial. Trials were divided into "correct"

5   and "incorrect" bins: correct production of the sentence required an exact match to the original

6   sentence, while all other trials were marked as incorrect. For each electrode, we computed the

7   difference in activation associated with correct and incorrect trials (Figure 3B).

8   Greater activation in the superior temporal cortex during the encoding phase (perception of S2)

9   was predictive of less successful sentence memory (Figure 3B). The same effect was also

10   observed in a small number of parietal and motor sites. This pattern is consistent with the role

11   of all these areas in the initial encoding of auditory and linguistic information. To identify the

12   regions supporting working memory, we examined activity during silent rehearsal that was

13   predictive of accurate sentence memory (Figure 3C). When motor and premotor regions

14   exhibited increased activation, this predicted less accurate later sentence production.

15   Because the behavioral associations identified in Figure 3B may be mediated by stimulus

16   properties (e.g., sentences that are easier to remember because of their content may produce

17   less activations) they should be interpreted with caution. Nonetheless, the results indicate that

18   motor and premotor sites are important sites for memory-dependent behavior.

25
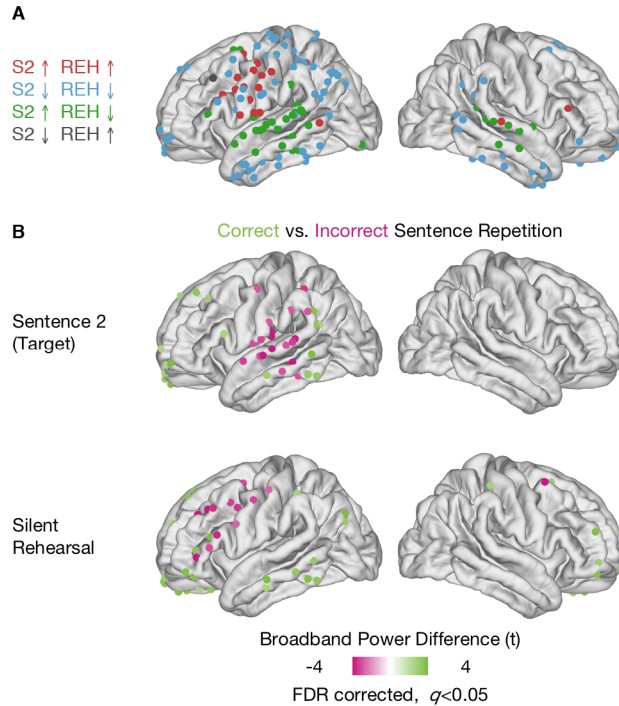
Figure 3. **Coactivation and Memory-Dependent Behavior in Sensorimotor and Premotor Sites.**
A, Motor and premotor sites coactive during perception and silent rehearsal. Electrodes are color coded based on their combined level of activation/deactivation in the perception phase (sentence 2, S2) and silent rehearsal phase (REH). Sites that are activated above baseline in both task phases (red) cluster around sensorimotor and premotor cortex. **B**, Spatial distribution of broadband power differences for correct (light green) and incorrect sentences (pink) during sentence 2 (top) and silent rehearsal (bottom). Trials were grouped based on the overt sentence repetition behavior in the production phase. Only electrodes activated at the level $q < 0.05$ after FDR correction are shown.

Coherent Sentences Elicit Greater Broadband Power in Semantic Network During Perception and Rehearsal

To probe semantic selectivity, we assessed activation differences during perception and silent rehearsal when S2 was semantically coherent or incoherent (lateral view, Figure 4, medial view Figure 4). No electrodes exhibited differences between coherent and incoherent trials during S1 or gap 1. This was expected, since the nature of the trial (coherent S2, incoherent S2) is not known to the participant before the onset of S2.

1   During perception, silent rehearsal and production of S2, coherent sentences were associated

2   with greater broadband power in the middle and inferior temporal gyrus, temporoparietal

3   cortex and in dorsal frontal areas (Figure 4A). However, overall activity in these areas was

4   below baseline, suggesting that incoherent sentences were the cause of the reduction, while

5   activity for coherent sentences was closer to baseline levels (Figure 4B). The pattern of greater

6   broadband power for coherent sentences in posterior parietal, middle temporal and dorsal

7   prefrontal cortex strongly resembles the "semantic network" identified by a meta-analysis of

8   120 fMRI studies (Binder et al., 2009).

9   During S2, perception of incoherent sentences elicited increased broadband power in left and

10 right superior temporal gyri and left inferior frontal gyrus. The activation difference in the

11 superior temporal gyrus disappeared during silent rehearsal and production but remained for

12 the inferior frontal gyrus in those phases. Inspection of the time courses (Figure 4B) revealed

13 that processing incoherent sentences resulted in stronger broadband power in all areas. This

14 pattern of activation difference between coherent and incoherent sentences was stronger in

15 the left hemisphere, though the overall pattern was similar in both hemispheres.
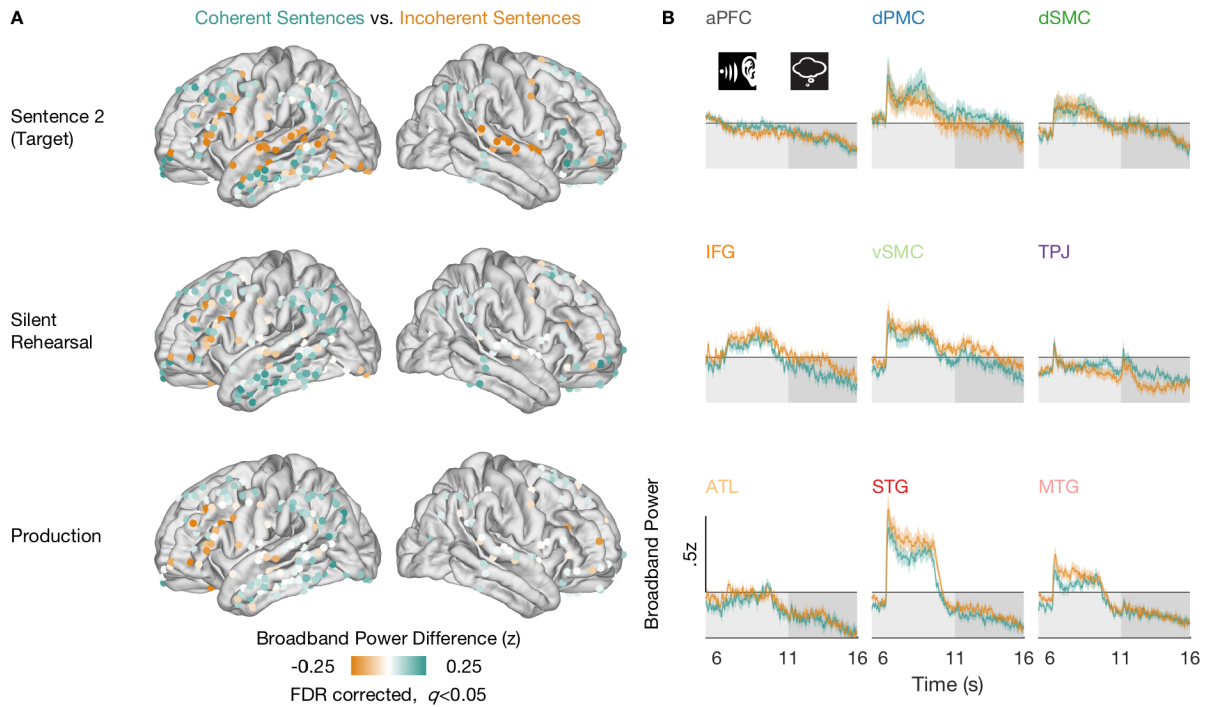
Figure 4. **Perception and Silent Rehearsal of Coherent Sentences Produces Increased Broadband Power in Semantic Network. A**, Spatial distribution of broadband power differences (coherent – incoherent) for coherent (teal) and incoherent sentences (orange) across task phases. Depicted are all sites showing a difference during either sentence 2, silent rehearsal or production. Only electrodes activated at the level $q < 0.05$ after FDR correction are shown. See also Figure S2. **B**, Average time courses for coherent and incoherent trials in the nine ROIs for the S2 and Rehearsal phases (indicated by icons and shading).

Activation Time Courses are Sentence-Specific During Perception and Silent Rehearsal

Next, we assessed which of the widespread regions implicated in sentence perception and silent rehearsal were encoding sentence-specific information. A region might increase or decrease its activity due to general task demands (e.g., "listen"), without encoding information about the specific sentence that is being heard. Therefore, we measured whether activity time courses in each region were sentence-specific. Since we were interested in the nature of the representation during verbal short-term memory, we focused on sentence-specific

28

1 representations during perception and silent rehearsal. To this end, we applied a technique for

2 measuring region-level response reliability (Figure 5; see Methods, Reliability Analysis) to

3 quantify sentence-specific responses in nine anatomical ROIs (Figure 2C). In each ROI, and

4 separately for coherent and incoherent sentences, we subsampled stimuli to estimate a set of

5 electrode weights (i.e., a spatial filter) that maximizes the reliability of responses across

6 stimulus repeats (Figure 5A). Then, to obtain an unbiased estimate of how much sentence-

7 specific information was contained in the time courses of each ROI, we measured the reliability

8 of the weighted responses during the same task phase, but in different (out-of-sample)

9 sentences and their repeats (Figure 5B).

10 During perception, we observed that auditory, linguistic, and motor cortices of the left

11 hemisphere exhibited the most reliable and sentence-specific responses to both coherent and

12 incoherent sentences (Figures 5C, S5A; see Figure S3 for electrode-level reliability). During the

13 auditory presentation of S1 (which did not have to be rehearsed), sentence-specific response

14 time courses were observed in the left STG and MTG as well as in left vSMC. During the

15 presentation of S2 (the target sentence for rehearsal), we additionally observed sentence-

16 specific information in left dSMC and in left dPMC. These effects were observed for both

17 coherent and incoherent sentences. In the right hemisphere (Figures S4A, S5B), sentence-

18 specific responses were observed in the STG both during S1 and S2 and for coherent and

19 incoherent sentences. Sentence-specific information was also present in the right vSMC, but

20 this effect achieved statistical significance only for the coherent sentences.

21 During silent rehearsal, we observed reliable sentence-specific responses only in the left dSMC.

22 This effect only achieved statistical significance for the coherent sentences but was similar for

23 incoherent sentences. Looking more broadly across the surrounding areas, however, we did

24 observe many weakly reliable responses across many temporal and peri-Rolandic regions

25 during rehearsal (Figure S5). The within-phase reliability analysis could not determine whether

26 these weakly reliable responses were truly sentence-specific, but we return to these areas in

27 more powerful cross-phase analyses that have access to a larger sample of held-out sentences.
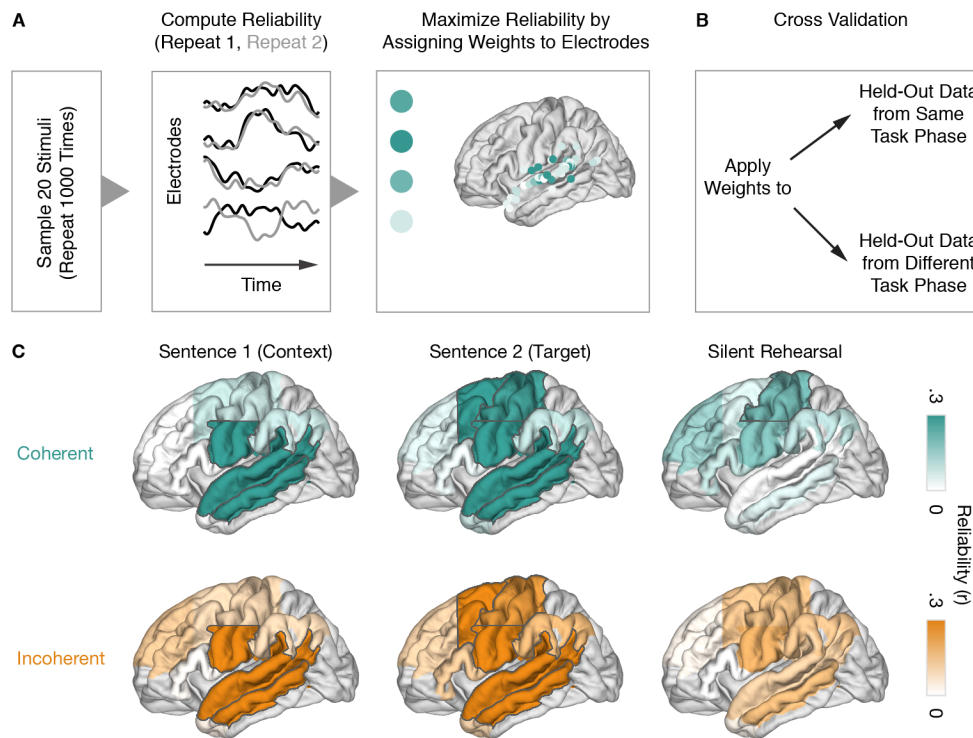
1

2 Figure 5. **Sentence-Specific Activity in Sensory and Motor Cortices.** Weighted reliability

3 analysis workflow and stimulus-specific encoding during perception and silent rehearsal. **A,**

4 Feature selection based on a subset of data. Reliability is maximized in a ROI by assigning

5 different weights to electrodes. Data from eight participants who completed all trials went into

6 this analysis. **B,** Cross validation using the weights from A on out-of-sample data from the same

7 task phase (Figures 5, 6, S4A, S5, S6) or from a different task phase (Figures 6, 7, 8, S4B, S4C,

8 S6). **C,** Sentence-specific encoding for coherent (teal, top) and incoherent sentences (orange,

9 bottom) during perception (sentence 1, sentence 2) and silent rehearsal by ROI. The highlighted

10 areas exhibit a significant difference from the null distribution ($p < 0.05$, two-sided,

11 uncorrected). See also Figure S3 for electrode-level reliability. See also Figures S4A and S5A.

12 **Temporal Precision of Sentence-Specific Neural Activity**

13 The previous analysis revealed sentence-specific neural activity that was most robust in

14 temporal, sensorimotor and premotor cortices – but is this sentence-specific activity also

15 specific to individual moments within the stimulus? To answer this question, we compared the

16 sentence-specific activity across all pairs of individual time points during perception and during

17 rehearsal. Using the weights optimized for each ROI in each fold of our cross-validation

30

1   procedure, we computed the correlation of a specific moment's activity level in the first

2   presentation of a stimulus, comparing it against all possible timepoints in the second

3   presentation of that stimulus.

4   The time-specific correlation analysis produces a two-dimensional time-time correlation matrix

5   (left hemisphere, Figure 6; right hemisphere, Figure S6). The diagonal of the matrix reveals

6   correlation at matching timepoints (time t in repeat 1 matched against the same point in repeat

7   2). If sentence-specific activity in a region varies from moment to moment, then the time-

8   correlation values along the diagonal of the matrix will be higher than the off-diagonal values.

9   Conversely, if the sentence-specific activity pattern evolves more slowly and is not specific to

10   individual moments in the stimulus, then the temporal pattern will be more stationary, and

11   non-matching timepoints will also be correlated across repeats; in this case, off-diagonal values

12   will be as high as the on-diagonal.

13   In the left hemisphere, we observed three sentence-specific response patterns, with varying

14   degrees of temporal specificity (Figure 6). First, the STG and MTG exhibited sentence-locked

15   activity only along the time-time diagonal and only during sentence perception. Thus, their

16   responses are locked to specific moments within a sentence during perception, in line with

17   their role in tracking low-level perceptual features (Cheung et al., 2016). A second pattern was

18   observed in aPFC and TPJ: here the sentence-specific responses were temporally diffuse (not

19   restricted to the time-time diagonal). Moreover, the reliable activity pattern was shared across

20   perception and silent rehearsal; this pattern is suggestive of a slowly evolving representation

21   that spans perception and rehearsal.

22   Finally, in vSMC, dSMC and dPMC, we observed a temporal encoding pattern that indicated a

23   transformation between task phases. These sensorimotor circuits exhibited robust time-locked

24   sentence-specific activity during sentence perception (i.e., elevated along the diagonal),

25   combined with off-diagonal, slowly varying activity. During rehearsal, sentence-specific activity

26   was more temporally diffuse. Notably, the sentence-specific activity in sensory and motor

1   regions was distinct between perception and rehearsal phases: the electrode weights from

2   perception showed little generalization to rehearsal (Figure 6B, cross-phase).

3   Among areas of the right hemisphere, only the STG exhibited sentence-locked activity along the

4   diagonal during perception, suggesting a close tracking of auditory and phonological

5   information as in the left hemisphere (Figure S6).

6

7   Figure 6. **Temporal Specificity of Sentence-Locked Activity.** See also Figure S6. **A**, Cross-time
8   cross-trial correlation maps during perception and silent rehearsal. Within each ROI, a positive
9   correlation at time (t1, t2) indicates that higher (spatially-weighted) activation values at time t1
10   on the first presentation of a sentence were associated with higher activation values at time t2
11   on the second presentation of the sentence. Thus, a positive correlation along the diagonal
12   (t1=t2) indicates that responses are not only sentence-specific but also locked to individual
13   timepoints within a stimulus. The correlations are computed either using the region-specific
14   weights derived from the sentence 2 phase (S2, top part of each panel) or the weights derived
15   from the rehearsal phase (REH, bottom part of each panel). Correlation values are averaged
16   across coherent and incoherent sentences. The gray lines indicate the onsets of the silent gap
17   following sentence 2 and the silent rehearsal cue, respectively. **B**, Statistical summary of the
18   cross-time cross-trial correlation values shown in Panel A. The distribution of correlation values

1 is shown for the on-diagonal component (t1=t2, black), the off-diagonal component (dark gray)
2 and the cross-phase component (light gray) of the correlation matrix. Results are summarized
3 separately for the sentence 2 and silent rehearsal phases and for each ROI. The fat horizontal
4 line represents the median of the out-of-sample correlation distribution, while the box width
5 represents its interquartile range. The vertical thin lines extend to the minimum/maximum of
6 the out-of-sample correlation distribution, excluding outliers.

7 **Shared Sets of Electrodes in Prefrontal Cortex Sensitive to Contextual Information**

8 **of the Past**

9 The temporal specificity analyses (Figures 6, S6) suggest that, as we hear and then mentally

10 rehearse sentences, some brain regions represent sentence-specific information in common

11 sets of electrodes across the perceptual and rehearsal phases. To more quantitatively test this

12 phenomenon, we applied our cross-validation procedure to the rehearsal phase using weights

13 that were identified during perception of S2. Since the perceptual and rehearsal phases are

14 completely separate data, this also enabled us to increase the power of our cross-validation

15 procedure: we used the average weights estimated on 20 stimuli from the perceptual phase

16 (Figures 5C, S4A) and measured the repeat correlations in the full set of 30 (instead of 10) held-

17 out trials from the rehearsal phase. Using this approach, we identified sentence-specific activity

18 with shared electrode weights in aPFC and TPJ. This cross-phase effect was observed separately

19 for both coherent and incoherent sentences (Figure 7A). Thus, even though the sentence-

20 specific activity in these high-level areas is weak and temporally diffuse (Figure 6), it is

21 expressed in a common set of electrodes across perception of S2 and silent rehearsal.

22 We also observed significant sentence-specific information with shared electrode weights in

23 dSMC (for coherent sentences only) and in dPMC (for incoherent sentences only). No sentence-

24 specific activity with shared electrode weights was observed in the right hemisphere (Figure

25 S4B).

1
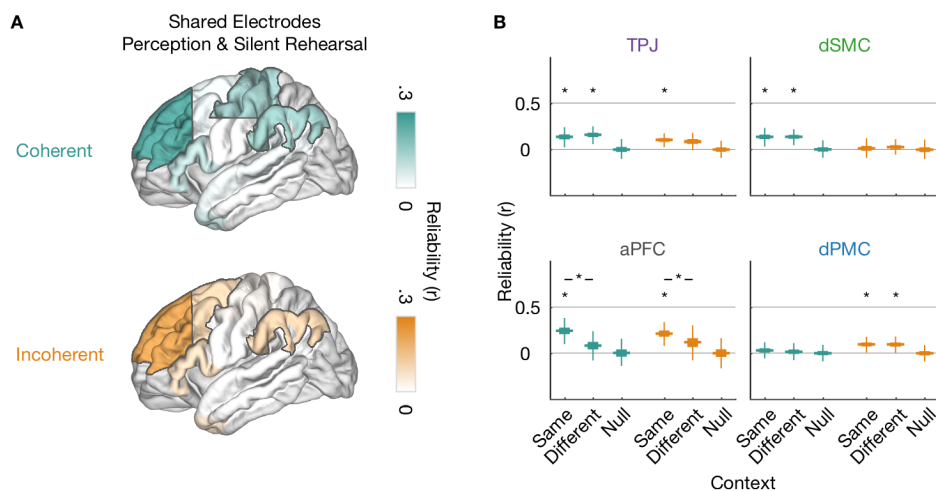
Figure 7. **Shared Sets of Electrodes Encode Sentence-Specific Information. A**, Reliability for shared electrodes in perception (sentence 2) and silent rehearsal that encode sentence-specific information for coherent (teal, top) and incoherent (orange, bottom) sentences. Weights from sentence 2 were applied to time courses from silent rehearsal. The highlighted areas exhibit a significant difference from the null distribution ($p < 0.05$, two-sided, uncorrected). See also Figure S4B. **B**, Reliability scores for the four ROIs sharing sets of electrodes between sentence 2 and silent rehearsal. Reliability is shown for the three cases when the second repeat of sentence 2 was preceded by the same sentence 1 context (Same) or the different sentence 1 context from the same stimulus group (Different). As a control, we also computed reliability for when both sentence 1 and sentence 2 were drawn from a nonmatching stimulus group (Null). Differences from the null distribution are indicated by single asterisks ($p < 0.05$ two-sided, uncorrected), differences between same vs. different contexts are indicated by lines ($p < 0.05$ one-sided, uncorrected).

What kind of information is encoded in the frontal and parietal electrodes that encode

sentence-specific activity in both phases? Some models of verbal short-term memory have

suggested that higher level sentence features (semantics and syntax) are used to "regenerate"

information patterns at the time of sentence recall (Lombardi and Potter, 1992; Potter and

Lombardi, 1990). Therefore, we tested whether the shared sentence-specific representations in

each ROI were sensitive to the high-level semantics of the stimuli. To do so, we measured

whether the sentence-specific activity in each region during silent rehearsal was the same

across the two different semantic contexts that were generated by the non-rehearsed S1. For

example, we tested whether the "spinning the wheel" sentence was represented the same way

in the "ship captain" context and in the "game show" context" (see Figure 1B for both

34

1   contexts). Thus, the phonological and syntactic features of the rehearsed sentence were

2   identical, while its high-level semantics were altered by the context manipulation.

3   Only the aPFC was sensitive to the contextual information of S1 at the time of rehearsal ($p <$

4   0.05, one-sided, Figure 7B). In all other ROIs reliability was not detectibly affected by the

5   contextual information of S1 (Figure 7B). This suggests that the most robustly encoded

6   sentence-specific information during silent rehearsal (i.e., motor and premotor regions) were

7   not representing high-level sentence context. Instead, the high-level information was present in

8   the more temporally diffuse representations of the aPFC.

9   **Sentence-Specific Activity is Shared Between Perception and Rehearsal in**

10   **Prefrontal and Temporoparietal Cortex**

11   Given that common sets of electrodes encoded sentence-specific information across perception

12   of S2 and silent rehearsal, we next assessed whether the sentence-specific neural activity

13   patterns themselves were shared across perception and rehearsal. To this end, we directly

14   compared the weighted and concatenated response time courses from the perception phase

15   (S2) against the silent rehearsal phase. We employed the electrode weights from the S2 phase,

16   and computed correlations against held-out data from a separate repeat: i.e., S2 phase of

17   repeat 1 compared with silent rehearsal of repeat 2, and vice versa. We observed that

18   sentence-specific time courses in aPFC, dPMC and TPJ of the left hemisphere were shared

19   between perception and silent rehearsal. This effect was observed independently for coherent

20   and incoherent sentences (Figure 8). For coherent sentences, sentence-specific time courses

21   were also shared across task phases in the vSMC. In the right hemisphere, sentence-specific

22   time courses were shared only in the MTG and only for incoherent sentences (Figure S4C).
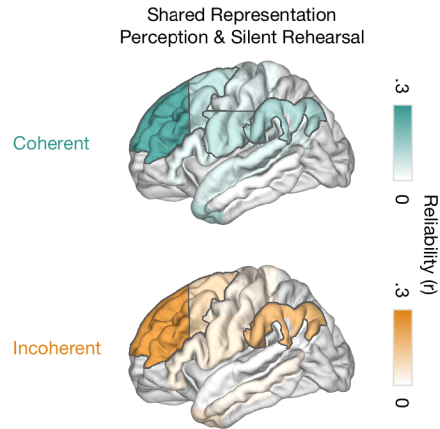
1

2 Figure 8. **Shared Sentence-Specific Time Courses Between Perception and Silent Rehearsal.**
3 Sentence-specific reliability measured by directly correlating the response during perception
4 (sentence 2) and during silent rehearsal. Cross-phase correlations for coherent (top) and
5 incoherent (bottom) sentences are shown in the left hemisphere. Weights for this cross-phase
6 analysis were defined as those that maximized the within-phase reliability in sentence 2. The
7 highlighted areas exhibit a significant difference from the null distribution ($p < 0.05$, two-sided,
8 uncorrected). See also Figure S4C.

Discussion

10 We recorded intracranial neural activity across widespread language-related networks as

11 people heard and then mentally rehearsed spoken sentences. In each brain region, we tested

12 whether silent rehearsal of sentences involved reactivation of sentence-specific

13 representations established during perception, or transformation to a distinct sentence-specific

14 representation. We found evidence for both processes: transformation was most apparent in

15 SMC and PMC, while aPFC and TPJ maintained a more static representation that was shared

16 across phases. In the aPFC and TPJ, where representations were shared across perception and

17 rehearsal, neural representation were also more sensitive to changes in the syntactic structure

18 and contextual semantics of the sentences.

1   The data implicate sensorimotor cortex (both vSMC and dSMC) and PMC as critical bridges

2   between the perception and rehearsal of spoken sentences. These regions were the only ones

3   for which neuronal activity, as indexed by broadband power, was consistently increased during

4   both silent rehearsal and perception (Figure 3A). Moreover, broadband power in motor and

5   premotor areas during rehearsal predicted the accuracy of sentence memory (Figure 3B).

6   Finally, vSMC, dSMC and PMC exhibited stimulus-specific encoding across both perceptual and

7   silent rehearsal phases (Figure 5) and this stimulus-specific encoding during perception was

8   temporally precise (Figure 6). These findings are consistent with prior ECoG reports of short-

9   latency encoding of auditory speech in SMC and PMC (Cheung et al., 2016; Edwards et al., 2009;

10   Glanz et al., 2018), and they extend the prior work by tracking the representations of complex

11   speech in motor areas during the transition from perception to rehearsal.

12   Sentence-specific activity in vSMC, dSMC and PMC was transformed between the perception

13   and silent rehearsal of the same sentence. A moment-by-moment comparison of the sentence-

14   specific activity (Figure 6) revealed a transition between distinct representations in perception

15   and silent rehearsal. Sentence-specific activity patterns defined during perception in the vSMC,

16   dSMC and PMC did not extend beyond the end of the perception of S2; at the same time, a

17   distinct rehearsal-specific activity pattern became reliable, starting from the end of the

18   perception phase and continuing into rehearsal. Thus, the present data connect SMC and PMC

19   to a transformation process that supports short-term memory for natural spoken language.

20   These findings are consistent with a sensorimotor transformation model that has been

21   proposed for working memory (Cogan et al., 2014, 2017); they are not consistent with models

22   that posit a one-to-one mirroring of activity during perception and production (D'Ausilio et al.,

23   2009; see also Liberman et al., 1967).

24   Short-term memory for phonological information is thought to be supported by an auditory-

25   motor interface (Buchsbaum and D'Esposito, 2008; Hickok and Poeppel, 2007; Jacquemot and

26   Scott, 2006; Rauschecker, 2011). The present data indicate that the SMC and PMC could

37

1   possess both the auditory and motor representations necessary for such an interface. The

2   vSMC activity in speech perception likely reflects more low-level and obligatory audio-motor

3   representations (Cheung et al., 2016): this area exhibited reliable sentence-specific activity

4   even for S1 (which did not have to be rehearsed), and its reliability increased only marginally for

5   S2 (which had to be rehearsed). By contrast, the dSMC and PMC were more sensitive to the

6   task relevance of the speech input: these areas exhibited greater sentence-specific reliability for

7   S2 (which had to be rehearsed), relative to S1 (Figures 5C, S5). We tentatively propose that the

8   vSMC representations directly track auditory and motor representations, while the time-locked

9   dSMC and PMC responses are purposed not only toward control of articulatory sequences, but

10   also toward expectation of sensory sequences (e.g., Schubotz and von Cramon, 2003). A short-

11   term memory trace of speech in sensorimotor cortices could also explain why these areas

12   would be engaged for discriminating speech stimuli in noise (Du et al., 2014).

13   In contrast with the sensorimotor circuits discussed above, the prefrontal and temporoparietal

14   cortex exhibited a temporally diffuse sentence-specific activity pattern, which was shared

15   across the perception and silent rehearsal of the same sentence. Although some of the

16   sentence-specific activity in motor sites was also shared across phases (Figures 7,8), these

17   shared effects were inconsistent across conditions and made up only a small proportion of the

18   reliable signal in motor circuits (Figure 6B). By contrast, the cross-phase correlation in aPFC and

19   in the TPJ was as large as the within-phase correlations (Figure 6B, within-phase and cross-

20   phase). At the level of aggregate activation, the prefrontal and temporoparietal circuits

21   exhibited increased broadband power responses for coherent sentences (Figure 4), consistent

22   with their role in a "semantic network" (Binder et al., 2009).

23   Prefrontal and temporoparietal cortex have long temporal receptive windows, exhibiting slower

24   population dynamics than sensory cortices (Honey et al., 2012; Murray et al., 2014) and

25   responding to new input in a manner that depends on prior context (Hasson et al., 2015; Lerner

26   et al., 2011). Consistent with this prediction, the temporal activation pattern in the left aPFC

27   was sensitive to prior contextual information (changes in the high-level situational meaning of

1   S2 based on the context provided by S1). The stable sentence-specific signals in higher order

2   areas may provide a persistent "scaffold" which supports "regeneration" of detailed surface

3   features of the sentence (e.g., phonemes and prosody) from more abstract properties (e.g.,

4   syntax and semantics) (Lombardi and Potter, 1992; Potter and Lombardi, 1990; see also Savill et

5   al., 2017). More generally, the slowly changing prefrontal representations we observed in high

6   level cortical areas are consistent with a distributed, drifting cortical "context" that recursively

7   interacts with episodic retrieval processes (Polyn et al., 2009). We note, however, that the

8   temporal context effects we observed were small relative to the those in prior studies using

9   naturalistic stimuli (e.g., Lerner et al., 2011); rich and extended narrative stimuli likely generate

10   a more powerful contextual modulation than the single preceding sentence (S1) that we used in

11   the present design.

12   Sentence-specific information changed most rapidly in the posterior areas of frontal cortex (i.e.

13   in motor cortex) and changed more slowly toward anterior regions (i.e. premotor and

14   prefrontal cortex, Figure 6). The distinct timescales of these frontal areas parallel the recent

15   observation of distinct working memory functions for populations of neurons with distinct

16   dynamical timescales (Cavanagh et al., 2018; Wasmuht et al., 2018). Prefrontal neurons with

17   intrinsically short timescales, as measured by spontaneous firing dynamics, responded more

18   rapidly during the perceptual phase of a task, whereas those with intrinsically longer timescales

19   encoded information in a more stable way towards a delay period (Wasmuht et al., 2018). In

20   our task, faster dynamics seem to be more important in sensorimotor transformation (from

21   perception to subvocal rehearsal), whereas slower dynamics were associated with areas whose

22   function may be more semantically sensitive. In future work, we will directly characterize the

23   relative timing across these different stages of processing, in order to determine changes in the

24   direction of information flow across perception and rehearsal.

25   Unexpectedly, coherent sentences and incoherent sentences were represented in a stimulus-

26   specific manner within essentially the same sets of brain regions, albeit with different levels of

27   mean broadband power activation. This suggests that even the incoherent sentences possessed

39

1   sufficient high-level structure (e.g., lexical semantics and aspects of syntax), similar to that used

2   to represent the coherent sentences. This minor effect of sentence incoherence is consistent

3   with the general observation that semantic implausibility has less of an effect on sentence

4   processing than strong syntactic violations (Polišenská et al., 2015).

5   During silent rehearsal, we expected to observe, but did not observe, sentence-specific activity

6   in the ATL (Patterson et al., 2007), the posterior STG and planum temporale (Buchsbaum and

7   D'Esposito, 2008), and the IFG (Hickok, 2012). There is evidence of marginally reliable responses

8   during the perception phase in both the ATL and in the IFG in the time-resolved analysis (Figure

9   6B). In addition, one electrode in the posterior temporal cortex was jointly active in both

10   perception and rehearsal, but we did not find stimulus-specific decoding in the superior

11   temporal cortex as a whole during silent rehearsal. It is possible that stimulus-specific encoding

12   could be identified in these areas given a larger pool of patients, denser electrode coverage, or

13   depth electrode coverage targeting sulci (especially the Sylvian fissure). Although we had broad

14   coverage, we cannot rule out the possibility of false negatives; instead, our data speaks most

15   clearly for the reliable and stimulus-specific responses across the lateral cerebral cortex,

16   spanning perception and rehearsal.

17   Overall, our data suggest that sensorimotor and premotor cortices can support an audio-motor

18   interface proposed by leading models of verbal short-term memory. In the SMC and PMC we

19   observed extensive joint activation across perception and rehearsal, and rehearsal activity

20   which predicted the accuracy of later sentence production. In parallel with this possible audio-

21   motor interface, more abstract sentence features were maintained in prefrontal and

22   temporoparietal circuits. To better understand how high-level semantic features constrain and

23   facilitate our inner speech, future work should examine the interactions between sensorimotor

24   circuits and the frontal and temporoparietal cortex as we silently rehearse sequences of words.

25   Acknowledgments

8 References

9 Arsenault, J.S., and Buchsbaum, B.R. (2016). No evidence of somatotopic place of articulation

10 feature mapping in motor cortex during passive speech perception. Psychon. Bull. Rev. *23*,

11 1231–1240.

12 Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and

13 powerful approach to multiple testing. J. R. Stat. Soc. - Ser. B Methodol. *57*, 289–300.

14 Binder, J.R., Desai, R.H., Graves, W.W., and Conant, L.L. (2009). Where is the semantic system?

15 A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex *19*,

16 2767–2796.

17 Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and

18 classification of ERP components — a tutorial. NeuroImage *56*, 814–825.

19 Bonhage, C.E., Fiebach, C.J., Bahlmann, J., and Mueller, J.L. (2014). Brain signature of working

20 memory for sentence structure: enriched encoding and facilitated maintenance. J. Cogn.

21 Neurosci. *26*, 1654–1671.

22 Buchsbaum, B.R., and D'Esposito, M. (2008). The search for the phonological store: from loop to

23 convolution. J. Cogn. Neurosci. *20*, 762–778.

1   Buchsbaum, B.R., Lemire-Rodger, S., Fang, C., and Abdi, H. (2012). The neural basis of vivid

2   memory is patterned on perception. J. Cogn. Neurosci. *24*, 1867–1883.

3   Cavanagh, S.E., Towers, J.P., Wallis, J.D., Hunt, L.T., and Kennerley, S.W. (2018). Reconciling

4   persistent and dynamic hypotheses of working memory coding in prefrontal cortex. Nat.

5   Commun. *9*, 3498.

6   Cheung, C., Hamiton, L.S., Johnson, K., and Chang, E.F. (2016). The auditory representation of

7   speech sounds in human motor cortex. ELife *5*, e12577.

8   Cogan, G.B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., and Pesaran, B. (2014). Sensory-

9   motor transformations for speech occur bilaterally. Nature *507*, 94–98.

10   Cogan, G.B., Iyer, A., Melloni, L., Thesen, T., Friedman, D., Doyle, W., Devinsky, O., and Pesaran,

11   B. (2017). Manipulating stored phonological input during verbal working memory. Nat.

12   Neurosci. *20*, 279–286.

13   Crone, N.E., Boatman, D., Gordon, B., and Hao, L. (2001). Induced electrocorticographic gamma

14   activity during auditory perception. Clin. Neurophysiol. *112*, 565–582.

15   D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The

16   motor somatotopy of speech perception. Curr. Biol. CB *19*, 381–385.

17   Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human

18   cortical gyri and sulci using standard anatomical nomenclature. NeuroImage *53*, 1–15.

19   Dmochowski, J.P., Sajda, P., Dias, J., and Parra, L.C. (2012). Correlated Components of Ongoing

20   EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement? Front. Hum.

21   Neurosci. *6*.

22   Du, Y., Buchsbaum, B.R., Grady, C.L., and Alain, C. (2014). Noise differentially impacts phoneme

23   representations in the auditory and speech motor systems. Proc. Natl. Acad. Sci. *111*, 7126–

24   7131.

Edwards, E., Soltani, M., Kim, W., Dalal, S.S., Nagarajan, S.S., Berger, M.S., and Knight, R.T. (2009). Comparison of time–frequency responses and the event-related potential to auditory speech stimuli in human cortex. J. Neurophysiol. *102*, 377–386.

Fedorenko, E., Scott, T.L., Brunner, P., Coon, W.G., Pritchett, B., Schalk, G., and Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. Proc. Natl. Acad. Sci. *113*, E6256–E6262.

Fegen, D., Buchsbaum, B.R., and D'Esposito, M. (2015). The effect of rehearsal rate and memory load on verbal working memory. NeuroImage *105*, 120–131.

Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E., and Knight, R.T. (2010). Single-trial speech suppression of auditory cortex activity in humans. J. Neurosci. Off. J. Soc. Neurosci. *30*, 16643–16650.

Flinker, A., Korzeniewska, A., Shestyuk, A.Y., Franaszczuk, P.J., Dronkers, N.F., Knight, R.T., and Crone, N.E. (2015). Redefining the role of Broca's area in speech. Proc. Natl. Acad. Sci. U. S. A. *112*, 2871–2875.

Glanz, O., Derix, J., Kaur, R., Schulze-Bonhage, A., Auer, P., Aertsen, A., and Ball, T. (2018). Real-life speech production and perception have a shared premotor-cortical substrate. Sci. Rep. *8*, 8898.

Groppe, D.M., Bickel, S., Dykstra, A.R., Wang, X., Mégevand, P., Mercier, M.R., Lado, F.A., Mehta, A.D., and Honey, C.J. (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. J. Neurosci. Methods *281*, 40–48.

Hasson, U., Chen, J., and Honey, C.J. (2015). Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. *19*, 304–313.

Hickok, G. (2012). Computational neuroanatomy of speech production. Nat. Rev. Neurosci. *13*, 135–145.

1   Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. Nat. Rev.

2   Neurosci. *8*, 393–402.

3   Honey, C.J., Thesen, T., Donner, T.H., Silbert, L.J., Carlson, C.E., Devinsky, O., Doyle, W.K., Rubin,

4   N., Heeger, D.J., and Hasson, U. (2012). Slow cortical dynamics and the accumulation of

5   information over long timescales. Neuron *76*, 423–434.

6   Humphries, C., Binder, J.R., Medler, D.A., and Liebenthal, E. (2007). Time course of semantic

7   processes during sentence comprehension: an fMRI study. NeuroImage *36*, 924–932.

8   Jacquemot, C., and Scott, S.K. (2006). What is the relationship between phonological short-term

9   memory and speech processing? Trends Cogn. Sci. *10*, 480–486.

10   King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the

11   temporal generalization method. Trends Cogn. Sci. *18*, 203–210.

12   Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (2011). Topographic mapping of a hierarchy

13   of temporal receptive windows using a narrated story. J. Neurosci. *31*, 2906–2915.

14   Lewis-Peacock, J.A., and Postle, B.R. (2008). Temporary activation of long-term memory

15   supports working memory. J. Neurosci. *28*, 8765–8771.

16   Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967). Perception

17   of the speech code. Psychol. Rev. *74*, 431–461.

18   Lombardi, L., and Potter, M.C. (1992). The regeneration of syntax in short term memory. J.

19   Mem. Lang. *31*, 713–733.

20   Manning, J.R., Jacobs, J., Fried, I., and Kahana, M.J. (2009). Broadband shifts in local field

21   potential power spectra are correlated with single-neuron spiking in humans. J. Neurosci. Off. J.

22   Soc. Neurosci. *29*, 13613–13620.

1   Meltzer, J.A., Rose, N.S., Deschamps, T., Leigh, R.C., Panamsky, L., Silberberg, A., Madani, N.,

2   and Links, K.A. (2016). Semantic and phonological contributions to short-term repetition and

3   long-term cued sentence recall. Mem. Cognit. *44*, 307–329.

4   Miller, K.J., Hermes, D., Honey, C.J., Sharma, M., Rao, R.P., den Nijs, M., Fetz, E.E., Sejnowski,

5   T.J., Hebb, A.O., Ojemann, J.G., et al. (2010). Dynamic modulation of local population activity by

6   rhythm phase in human occipital cortex during a visual search task. Front Hum Neurosci *4*, 197.

7   Miller, K.J., Foster, B.L., and Honey, C.J. (2012). Does rhythmic entrainment represent a

8   generalized mechanism for organizing computation in the brain? Front. Comput. Neurosci. *6*,

9   85.

10 Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. Science

11 *319*, 1543–1546.

12 Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C.,

13 Pasternak, T., Seo, H., Lee, D., et al. (2014). A hierarchy of intrinsic timescales across primate

14 cortex. Nat. Neurosci. *17*, 1661–1663.

15 Pallier, C., Devauchelle, A.D., and Dehaene, S. (2011). Cortical representation of the constituent

16 structure of sentences. Proc Natl Acad Sci U A *108*, 2522–2527.

17 Patterson, K., Nestor, P.J., and Rogers, T.T. (2007). Where do you know what you know? The

18 representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. *8*, 976–987.

19 Pei, X., Leuthardt, E.C., Gaona, C.M., Brunner, P., Wolpaw, J.R., and Schalk, G. (2011).

20 Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert

21 word repetition. NeuroImage *54*, 2960–2972.

22 Polišenská, K., Chiat, S., and Roy, P. (2015). Sentence repetition: what does the task measure?

23 Int. J. Lang. Commun. Disord. *50*, 106–118.

1 Polyn, S.M., Norman, K.A., and Kahana, M.J. (2009). A context maintenance and retrieval model

2 of organizational processes in free recall. Psychol. Rev. *116*, 129–156.

3 Potter, M.C., and Lombardi, L. (1990). Regeneration in the short-term recall of sentences. J.

4 Mem. Lang. *29*, 633–654.

5 Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y.

6 (2006). Motor cortex maps articulatory features of speech sounds. Proc. Natl. Acad. Sci. U. S. A.

7 *103*, 7865–7870.

8 Rauschecker, J.P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor

9 control and integration. Hear. Res. *271*, 16–25.

10 Ray, S., and Maunsell, J.H.R. (2011). Different origins of gamma rhythm and high-gamma

11 activity in macaque visual cortex. PLoS Biol. *9*, e1000610.

12 Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. Annu Rev Neurosci *27*, 169–

13 192.

14 Savill, N., Ellis, R., Brooke, E., Koa, T., Ferguson, S., Rojas-Rodriguez, E., Arnold, D., Smallwood,

15 J., and Jefferies, E. (2017). Keeping it together: semantic coherence stabilizes phonological

16 sequences in short-term memory. Mem. Cognit. 1–12.

17 Schubotz, R.I., and von Cramon, D.Y. (2003). Functional–anatomical concepts of human

18 premotor cortex: evidence from fMRI and PET studies. NeuroImage *20*, S120–S131.

19 Stokes, M.G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding

20 framework. Trends Cogn. Sci. *19*, 394–405.

21 Tibshirani, R.J., and Taylor, J. (2011). The solution path of the generalized lasso. Ann. Stat. *39*,

22 1335–1371.

1   Towle, V.L., Yoon, H.-A., Castelle, M., Edgar, J.C., Biassou, N.M., Frim, D.M., Spire, J.-P., and

2   Kohrman, M.H. (2008). ECoG gamma activity during a language task: differentiating expressive

3   and receptive speech areas. Brain *131*, 2013–2027.

4   Wasmuht, D.F., Spaak, E., Buschman, T.J., Miller, E.K., and Stokes, M.G. (2018). Intrinsic

5   neuronal dynamics predict distinct functional roles during working memory. Nat. Commun. *9*,

6   3499.

7

1 Supplement

2 Table S1. **Clinical and Demographic Information for the Participating Patients.** Related to

3 Methods, Participants.

Participant	Gender	Age at Surgery/ Seizure Onset	Language Laterality	Language (Age of Onset)	Seizure Type	Seizure Focus
P1	F	50.7/8	L	English	CPS	L Post Mesial
P2	M	28.4/16	L	English	CPS	Multifocal
P3	F	24.8/12	R	English	CPS	L Fronto-Temporal
P4	F	28.7/17	L	English	CPS	Bilateral Occipital
P5	F	26.2/19	NA	English	CPS	L Inferior Temporal
P6	F	32.4/18	Bilateral (R>L)	English (0), Greek (0)	CPS	Unclear
P7	M	28.0/1.5	NA	English	CPS	Unclear
P8	M	36.9/14	L	English	Deja-vu, GCS	L (poorly localized)
P9	M	35.2/4	Unclear	English	CPS	Bilateral Temporal
P10	F	25.1/5	L	Portuguese (0), English (3)	CPS	Diffuse
P11	F	19.7/<1	Bilateral (L>R)	Rumanian (0), English (<13)	CPS	L Temporal
P12	F	45.4/25	L	English	CPS	R Anterior Temporal
P13	F	39.4/19	Bilateral (R>L)	English	CPS, GTS	Unclear
P14	F	18.8/12	NA	English	CPS	R Frontal
P15	M	28.1/7	L	English	NA	Multifocal
P16	F	28.0/21	L	English	CPS, GTCS	R Temporal

4 *Note:* CPS, complex partial seizure; F, female; GCS, generalized clonic seizure; GTCS, generalized

5 tonic clonic seizure; GTS, generalized tonic seizure; L, left; M, male; NA, not available; R, right.

48

1 Table S2. **Full Sentence List.** Related to Figure 1B.

Stimulus Group #	Context Sentence: Context 1	Context Sentence: Context 2	Target Sentence: Coherent	Target Sentence: Incoherent
1	Gripping her drink, she looked around at everyone watching.	Gripping her camera, she looked around at everyone watching.	She took the shot, hoping she wouldn't regret it.	But then a pillowcase sat the mouse and brownies.
2	Neil Armstrong had been feeling depressed.	Michael Jackson had been feeling depressed.	He wanted to be known for more than his moonwalk.	Only to sit in the high tooth of a landscape.
3	His beloved poodle might not make it through another day.	His phone's battery might not make it through another day.	He made the decision to put it down.	It began once a work tree with the cloth.
4	Healthy food was the focus of the Californian charity.	Computers were the focus of the Californian charity.	They hoped to provide every child with an Apple.	Because baskets fished at the paint cup of the bagel.
5	The boat began to make noises as it entered the port.	The USB key began to make noises as it entered the port.	They felt something dangerous was happening.	When the neighbors are really grazed in something.
6	She stopped and stared at the cereal box, her eyes glazed over.	She stopped and stared at her fiancée, her eyes glazed over.	Finally, she continued down the aisle.	They mainly buttered around green hair.
7	The gameshow host spun the wheel hard to the right.	The ship captain spun the wheel hard to the right.	The onlookers cringed with apprehension.	The puddings arranged bullets of confidence.
8	Travelling to space had been her dream since childhood.	Travelling to Spain had been her dream since childhood.	She looked outside, eager to leave the shuttle.	When a problem was stuffing the loaded phone.
9	The bus was moving quickly on the other side of the field.	The baseball was moving quickly on the other side of the field.	He sprinted across the grass and jumped to catch it.	The ideas were to be impolite but not fresh.
10	Her laptop had not been used for the past six months.	Her apartment had not been used for the past six months.	She opened a few windows and looked around.	The snow was waving in the shell and injustice.
11	The weary pilgrims celebrated the achievement of their goal.	The hockey team celebrated the achievement of their goal.	It was amazing how far they had come.	His minor constitution is late to the wall.
12	The speech was only ready five minutes before the guests arrived.	The meal was only ready five minutes before the guests arrived.	It was very spicy, but everyone enjoyed it.	The sheet is voting mud into better charges.
13	The formal dresses had to be ready before the large party arrived.	The bowling lanes had to be ready before the large party arrived.	She watched him place each pin precisely.	All the fires were levelled on their softness.

14	Mandarin was the elective he had chosen to take in high school.	Auto shop was the elective he had chosen to take in high school.	It proved helpful when he was stranded outside Beijing.	The creek looked at them while catching the shining bells.
15	The pilot had an unusual safety check procedure.	The electrician had an unusual safety check procedure.	He liked to kiss the ground for good luck.	Where the wall shoe, a bin was talking.
16	The Formula One driver changed his tunings at the last minute.	The guitar player changed his tunings at the last minute.	Sadly, those changes produced the disaster.	They named the disappearing recent smiles.
17	The record company discovered her in the middle of nowhere.	The rescue helicopter discovered her in the middle of nowhere.	It was as if her life had found a new meaning.	She is the quadrangle or the time box on top.
18	The pilot's voice coming from the speakers was strongly accented.	The music coming from the speakers was strongly accented.	She was upset because she couldn't follow it.	Pages of the zeal were near the kitchen table.
19	She wanted to pay the bill, but was completely unsuccessful.	She wanted to walk the trapeze, but was completely unsuccessful.	Unfortunately, her balance was insufficient.	Anything without distance wants everything believed.
20	I wondered why the football was growing bigger and bigger.	I wondered why the phone bill was growing bigger and bigger.	Then, a few seconds later, it suddenly hit me.	The pie of sleepy video on her expectations.
21	The bus driver had always followed the rules exactly.	The accountant had always followed the rules exactly.	That day she decided to take a shortcut.	As the chocolate set for the country.
22	She lectured until lunch most days, but not Tuesdays.	They met there on most days, but not Tuesdays.	She forgot and ran over without thinking.	Only the market sang the pointing with blue.
23	The desert survivor had reached his goal after much effort.	The soccer team's captain had reached his goal after much effort.	He teared up, when he was handed the cup.	Also, a new file does prick in the comb.
24	The union was unsure of where they should strike.	The robbers were unsure of where they should strike.	They decided to stage it at the mall.	Even the power about in its today.
25	The Australian tour had been recommended by his friends.	The Australian wine had been recommended by his friends.	In the end, it was far too dry for his liking.	He befriends the fish and the bag was to be stared.
26	The frying pan was knocked over due to absolute carelessness.	The ocean was contaminated due to absolute carelessness.	Oil was everywhere and difficult to clean up.	Then feel spoken papers more for the memory.
27	A wedding ring was not the anniversary gift she had been expecting.	Divorce papers were not the anniversary gift she had been expecting.	Stunned, she reached for her ring finger.	Clocks that see miles were clean enough.
28	His daughter, Emily, was right there on the sofa when he finally returned home.	His guitar, Emily, was right there on the sofa when he finally returned home.	He ran anxiously across the room and hugged her.	I confirm to book the sky what the product.

29	He wanted to publish his memoires but they advised against it.	He wanted to publish a bank account but they advised against it.	There just wasn't enough interest in the account.	The milk basically triumphed from the other.
30	The fishing gear lay beside him in the passenger seat.	The investment plan lay beside him in the passenger seat.	He grabbed it and walked down to the bank.	The run while in the plan of first serve.

1 *Note*: Each row represents a complete stimulus group, i.e., four sentences composed of two context sentences (context 1, context 2)

2 and two target sentences (coherent, incoherent).

3 Table S3. **Questions Used in and Results of Norming Task.** Related to Figure 1C.

Category	Question (Rating Scale)	Context 1		Context 2			ANOVA results[1]	
		Coh, M(SEM)	Inc, M(SEM)	Coh, M(SEM)	Inc, M(SEM)	Semantic Context, F	Semantic Coherence, F	Interaction, F
Understandability	How understandable were the sentences? (very confusing/very understandable)	4.7(0.0)	2.2(0.1)	4.8(0.0)	2.0(0.1)	0.38	1430.70***	4.18*
Complexity	How complex is the sentence structure? (very simple/very complex)	2.3(0.1)	3.5(0.1)	2.5(0.1)	3.6(0.1)	2.58	221.35***	2.58
Visual Vividness	How vivid were the mental sounds described by the sentences? (very vague/very vivid)	4.1(0.1)	2.7(0.1)	4.2(0.1)	2.7(0.1)	0.07	302.42***	0.88
Auditory Vividness	How vivid were the mental images described by the sentences? (very vague/very vivid)	3.6(0.1)	2.7(0.1)	3.8(0.1)	2.8(0.1)	2.19	119.26***	0.83
Arousal	How energetic were the events described by the sentences? (very calm/very energetic)	3.1(0.1)	2.6(0.1)	3.3(0.1)	2.6(0.1)	0.34	30.99***	0.54
Empathy	How much empathy do you feel about the events described by the sentences? Very indifferent/very empathetic)	3.7(0.1)	2.5(0.1)	3.5(0.1)	2.3(0.1)	2.33	119.65***	0.07
Surprise	How surprising were the events described by the sentences?	2.7(0.1)	3.6(0.1)	2.7(0.1)	3.1(0.1)	5.20*	29.86***	6.30*

(very unsurprising/very surprising)

|								
Pleasantness	How pleasant were the events described by the sentences? (very unpleasant /very pleasant)	3.0(0.2)	2.7(0.1)	2.9(0.2)	2.6(0.1)	0.31	2.68	0.01

1   *Note*: [1] Significance levels indicated by * and *** corresponding to $p < .05$ and $p < .001$, respectively, degrees of freedom were (1,

2   120) for all tests; ANOVA, analysis of variance; Coh, coherent sentences; *F*, *F* value; Inc, incoherent sentences; *M*, mean; *SEM*,

3   standard error of the mean.

1   Table S4. **Included Areas in Destrieux Atlas per Region of Interest (ROI).** Related to Figure 2C.

| ROI	List of Included Areas in Destrieux Atlas
Superior Temporal Gyrus (STG)	G_temp_sup-Lateral, G_temp_sup-Plan_polar, G_temp_sup-Plan_tempo, Lat_Fis-post, S_temporal_sup
Middle Temporal Gyrus (MTG)	G_temporal_middle,S_temporal_inf
Anterior Temporal Lobe (ATL)	Pole_temporal
Temporoparietal Junction (TPJ)	G_pariet_inf-Angular, G_pariet_inf-Supramar
Inferior Frontal Gyrus	G_front_inf-Opercular, G_front_inf-Orbital, G_front_inf-Triangul
Ventral Sensory Motor Cortex (vSMC)	G_and_S_paracentral, G_and_S_subcentral, G_postcentral, G_precentral, S_central, S_postcentral, S_precentral-inf-part, S_precentral-sup-part
Dorsal Sensory Motor Cortex (dSMC)	G_and_S_paracentral, G_and_S_subcentral, G_postcentral, G_precentral, S_central, S_postcentral, S_precentral-inf-part, S_precentral-sup-part
Dorsal Premotor Cortex (dPMC)	G_front_middle, G_front_sup, S_front_inf, S_front_middle, S_front_sup
Anterior Prefrontal Cortex (aPFC)	G_front_middle, G_front_sup, S_front_inf, S_front_middle, S_front_sup

2
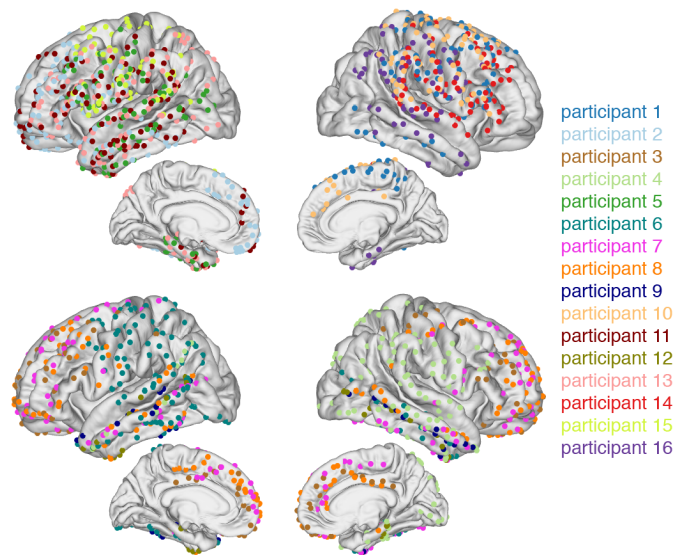
3

4   Figure S1. **Electrode Placement for Individual Participants Coded by Color.** Related to Figure
5   1D.
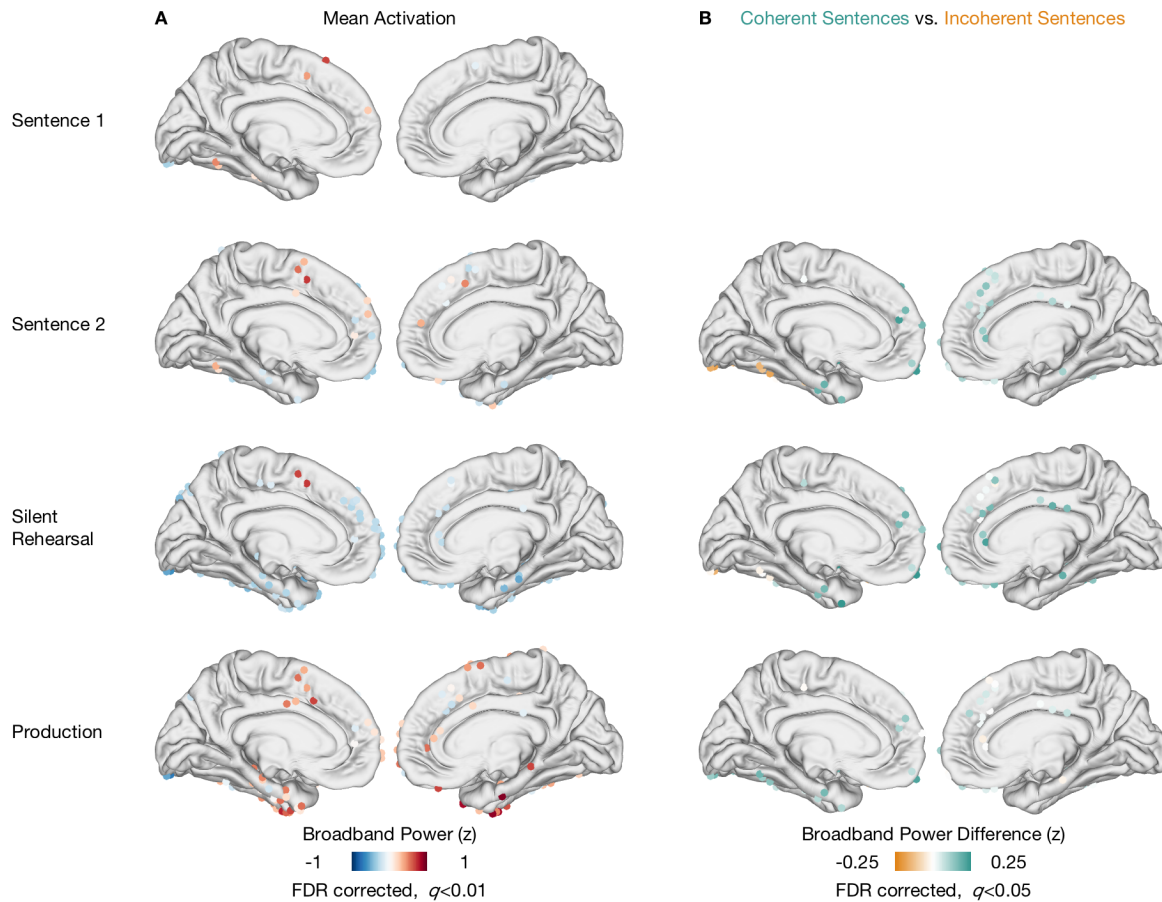
Figure S2. **Spatial Distributions of Broadband Power Responses in Medial Cortex.** Related to Figures 2B and 4A. **A**, Spatial distribution of broadband power responses aggregated across participants and conditions for each task phase in the left and right hemispheres. Only electrodes activated at the level q < 0.01 after FDR correction are shown. **B**, Spatial distribution of broadband power differences between coherent and incoherent sentences across task phases. Depicted are all sites showing a difference during either sentence 2, silent rehearsal or production. Only electrodes activated at the level q < 0.05 after FDR correction are shown.
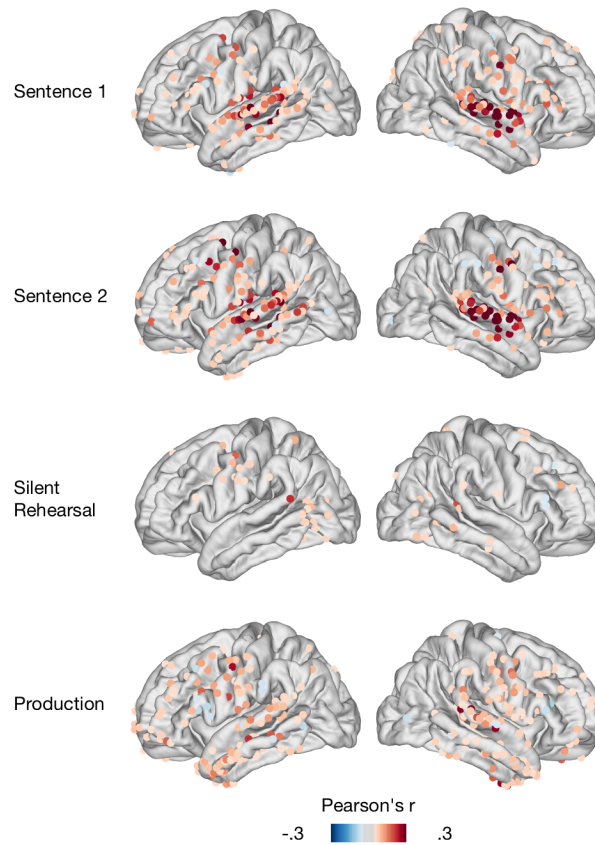
1

2 Figure S3. **Spatial Distributions of Average Within-Trial Repeat Reliability in All Task Phases.**
3 Related to Figures 5C and S4A. Pearson's *r* correlation values were computed separately for
4 each trial and its repeat and then averaged. For perception (sentence 1, sentence 2), the first
5 and last 500 ms were discarded to avoid on- and offset effects on reliability. For production, the
6 first 1000 ms was discarded to avoid onset effects. Only *r* values smaller than -0.05 and
7 exceeding 0.05 are shown.
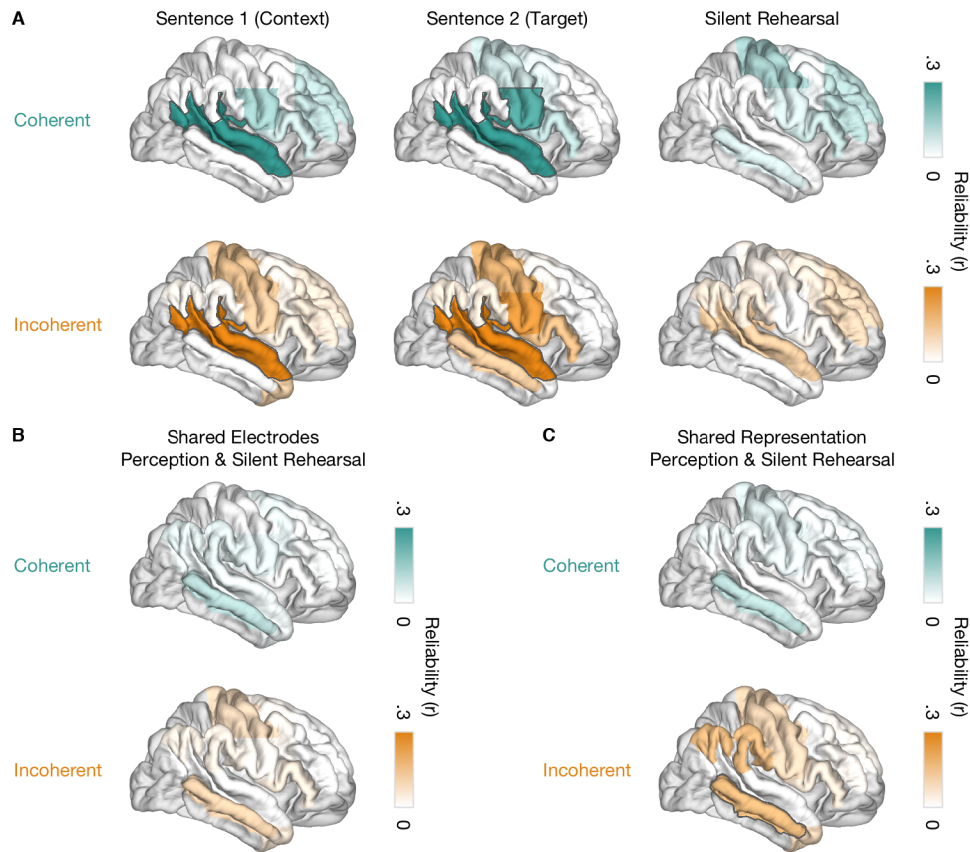
1

2   Figure S4. **Results of the Correlated Component Analysis in the Right Hemisphere.** Related to
3   Figures 5C, 7A and 8. **A**, Sentence-specific encoding in the right hemisphere during perception
4   (sentence 1, sentence 2) and silent rehearsal. Reliability is shown for each ROI for coherent
5   (top) and incoherent sentences (bottom). The highlighted areas exhibit a significant difference
6   from the null distribution ($p < 0.05$, two-sided, uncorrected). See also Figures S3 and S5B. **B**,
7   Reliability for shared electrodes in the right hemisphere. No ROI shares electrodes encoding
8   sentence-specific information for coherent (top) and incoherent (bottom) sentences during
9   perception (sentence 2) and silent rehearsal. Weights from sentence 2 were applied to time
10   courses from silent rehearsal. **C**, Sentence-specific reliability directly correlating perception
11   (sentence 2) and silent rehearsal for coherent (top) and incoherent (bottom) sentences.
12   Weights from sentence 2 were used for feature selection. The highlighted area exhibits a
13   significant difference from the null distribution ($p < 0.05$, two-sided, uncorrected).

1
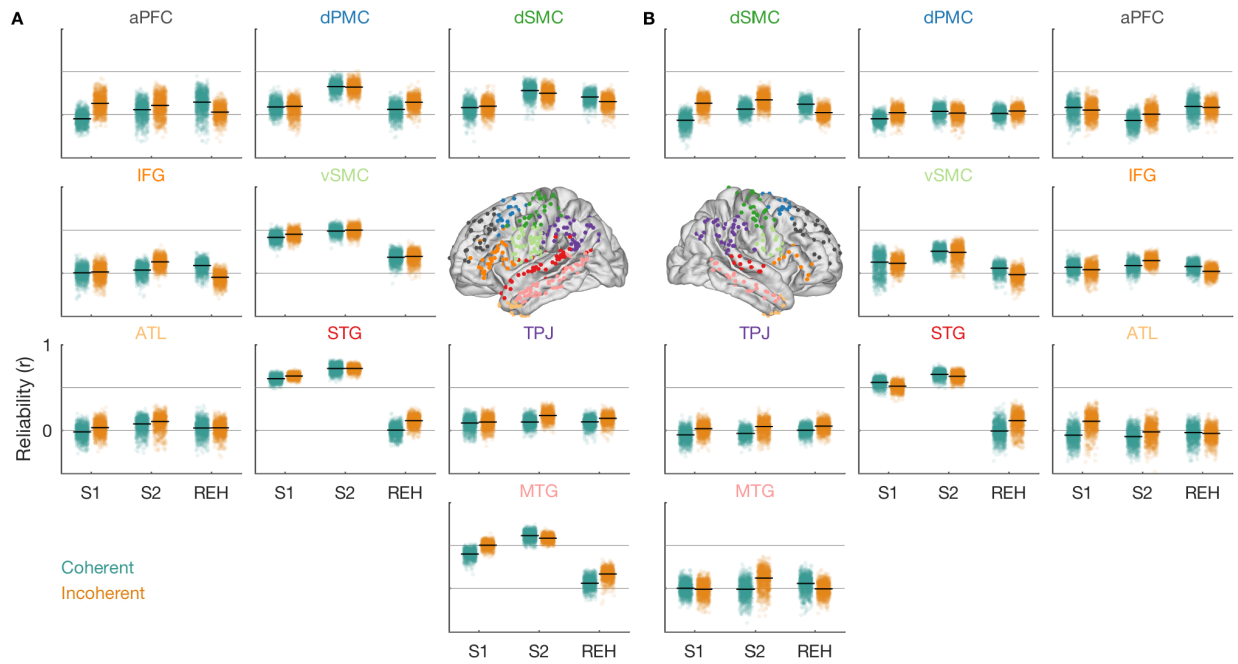
Figure S5. **Stimulus-Specific Encoding During Sentence 1 (S1), Sentence 2 (S2) and Silent Rehearsal (REH).** Related to Figures 5C and S4A. Shown is the reliability for each of the 1,000 permutations in the nine ROIs in the **A**, left and **B**, right hemispheres. The horizontal black line indicates the mean across all permutations, which is identical to the reliability shown in Figures 5 and S4A.
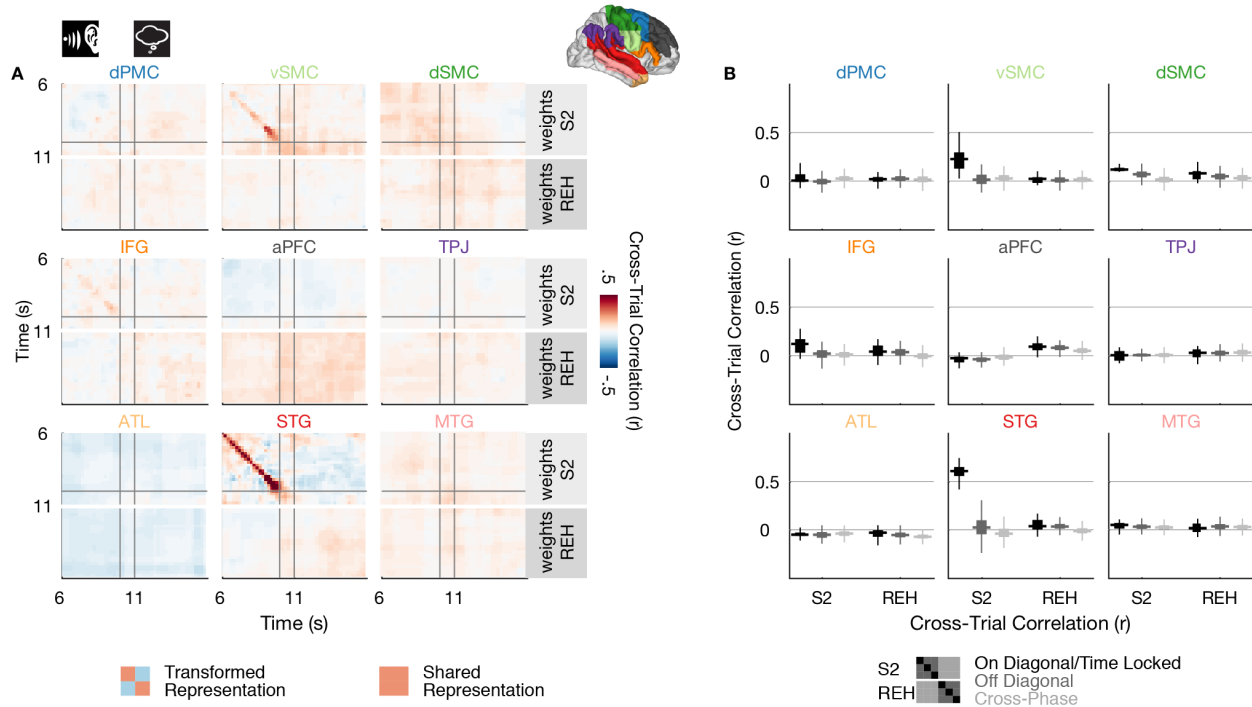
Figure S6. **Temporal Specificity of Sentence-Locked Activity in the Right Hemisphere.** Related to Figure 6. **A**, Cross-time cross-trial correlation maps during perception and silent rehearsal. Within each ROI, a positive correlation at time (t1, t2) indicates that higher (spatially-weighted) activation values at time t1 on the first presentation of a sentence were associated with higher activation values at time t2 on the second presentation of the sentence. Thus, a positive correlation along the diagonal (t1=t2) indicates that responses are not only sentence-specific but also locked to individual timepoints within a stimulus. The correlations are computed either using the region-specific weights derived from the sentence 2 phase (S2, top part of each panel) or the weights derived from the rehearsal phase (REH, bottom part of each panel). Correlation values are averaged across coherent and incoherent sentences. The gray lines indicate the onsets of the silent gap following sentence 2 and the silent rehearsal cue, respectively. **B**, Statistical summary of the cross-time cross-trial correlation values shown in Panel A. The distribution of correlation values is shown for the on-diagonal component (t1=t2, black), the off-diagonal component (dark gray) and the cross-phase component (light gray) of the correlation matrix. Results are summarized separately for the sentence 2 and silent rehearsal phases and for each ROI. The fat horizontal line represents the median of the out-of-sample correlation distribution, while the box width represents its interquartile range. The vertical thin lines extend to the minimum/maximum of the out-of-sample correlation distribution, excluding outliers.