Transformation of speech sequences in human sensorimotor circuits

Kathrin Müsch[a,1] , Kevin Himberger[a] , Kean Ming Tan[b] , Taufik A. Valiante[c] , and Christopher J. Honey[a]

[a]Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD 21218; [b]Department of Statistics, University of Michigan, Ann Arbor, MI 48109; and [c]Krembil Research Institute, Toronto Western Hospital, Toronto, ON M5T 2S8, Canada

After we listen to a series of words, we can silently replay them in our mind. Does this mental replay involve a reactivation of our original perceptual dynamics? We recorded electrocorticographic (ECoG) activity across the lateral cerebral cortex as people heard and then mentally rehearsed spoken sentences. For each region, we tested whether silent rehearsal of sentences involved reactivation of sentence-specific representations established during perception or transformation to a distinct representation. In sensorimotor and premotor cortex, we observed reliable and temporally precise responses to speech; these patterns transformed to distinct sentence-specific representations during mental rehearsal. In contrast, we observed less reliable and less temporally precise responses in prefrontal and temporoparietal cortex; these higher-order representations, which were sensitive to sentence semantics, were shared across perception and rehearsal of the same sentence. The mental rehearsal of natural speech involves the transformation of stimulus-locked speech representations in sensorimotor and premotor cortex, combined with diffuse reactivation of higher-order semantic representations.

ECoG | sentence repetition | verbal short-term memory | subvocal rehearsal

Immediately after hearing a series of words, we can silently replay them in our minds. What neural processes support this mental replay? Speech comprehension involves phonological, syntactic, and semantic processing across widespread circuits in temporal, frontal, and parietal cortex (1–5). The more specific function of verbal short-term memory relies on a core of regions that are involved in both speech perception and production: These regions include the posterior temporal cortex, motor and premotor areas, and the inferior frontal gyrus (6).

Within the regions implicated in verbal short-term memory, what kind of neural process supports the replay of recent speech? A natural hypothesis is that mental replay arises from neural replay: When we replay a series of words in our minds, the same neural populations may be activated as during the original auditory perception. This "shared representation" hypothesis is consistent with the common observation that activity patterns from perception may remain continuously active during a memory delay period or may be reactivated following periods of inactivity (7–9). More generally, "reactivation" of complex sequences of perceptual input is observed during vivid imagery of those sequences (10). A shared representation for hearing and rehearsing speech would also be consistent with "mirror" models, in which the imitation of speech actions is supported by a common set of neurons across perception and production (11, 12).

An alternative hypothesis is that, when we silently rehearse a series of words, we employ representations that are distinct from those involved in the original auditory perception. Contrasting with early observations of shared representations (13), recent intracranial and imaging studies have found that ventral sensorimotor circuits respond with distinct activity patterns during the perception and production of the same syllables (e.g., "ba") (14, 15). Moreover, widespread bilateral cortical circuits appear to "transform" between sensory and motor representations when

pseudowords (e.g., "pob") are held in mind and spoken aloud (16). Thus, the process of mentally rehearsing an entire sentence may be supported by circuits that transform between sensory and motor representations.

We set out to determine which representations were shared and which were transformed during the perception and silent rehearsal of many seconds of natural speech. Functional magnetic resonance imaging (fMRI) studies of sentence perception and production lack the spatiotemporal resolution to map word-by-word brain dynamics at natural speech rates. Prior electrocorticography (ECoG) studies have focused on rehearsal of individual items (e.g., single syllables), lacking syntactic or semantic content and posing little demand on verbal short-term memory. Here, we used ECoG to measure time-resolved neural activity across the lateral surface of the human brain during the perception and silent rehearsal of natural spoken sentences of 5 to 11 words.

While our primary goal was to examine how speech sequences vary between perception and rehearsal, we also manipulated the semantic coherence of the speech. Coherent strings of words are recalled better than incoherent strings of words, perhaps because surface features of sentences (e.g., their phonology) may be "regenerated" from more abstract features (e.g., semantic and syntactic cues) as long as the sentences are coherent (17–19). This leads to the prediction that semantically sensitive brain regions would be especially likely to express reactivation of activity across perception and rehearsal (20). Thus, to examine

Significance

When we hear a series of words, it is not only the auditory pathways of our brains that respond, but also the pathways that organize speech and action. What is the purpose of "motor" activity during language perception? We recorded electrical signals from the surfaces of the brain as people heard, mentally rehearsed, and spoke natural sentences. We found that human motor pathways generated precise, time-locked representations of sentences during listening. When listening was complete, the motor activity transformed to a distinct representation while people repeated the sentence in their heads and as they spoke the sentence aloud. Thus, motor circuits generate representations that can bridge from perception to short-term memory of spoken language.

whether the processes of reactivation and transformation differed with sentence coherence, we manipulated this factor in our stimulus set.

Overall, the data supported a model in which "motor" circuitry (sensorimotor cortex and premotor cortex) supports verbal short-term memory via a sensorimotor transformation (16). We observed the strongest sustained activation across sentence perception and silent rehearsal within the ventral sensorimotor cortex (vSMC), dorsal sensorimotor cortex (dSMC), and dorsal premotor cortex (dPMC) of the left hemisphere. Consistent with prior literature (e.g., refs. 14 and 21) the SMC and dPMC responded rapidly during sentence perception, encoding subsecond properties of the input. When sentences were silently rehearsed, SMC and dPMC again exhibited sentence-specific activity patterns, but the activity patterns were distinct from those observed during perception of the same sentences.

We also observed activity in higher-order areas (anterior prefrontal cortex [aPFC] and temporoparietal cortex [TPC]) which resembled a continuous activation or reactivation process supporting verbal short-term memory. Both the aPFC and TPC generated sentence-specific activity, but the signals in these areas were less temporally precise and less reliable than in sensory or motor areas. At the same time, patterns in the aPFC were sensitive to the contextual meaning of the sentence being rehearsed. Finally, the representations in these high-level areas were not transformed across perception and rehearsal: Instead, the prefrontal and temporoparietal dynamics elicited by a specific sentence were shared across perception and rehearsal.

Together, the data suggest that "core" speech rehearsal areas may implement a sensorimotor transformation in support of verbal short-term memory, while more distributed networks, sensitive to semantics, expressed a shared pattern of activity, bridging perception and rehearsal. Moreover, the sensorimotor cortex of the left hemisphere may possess both the sensory and the motor representations required to act as an audio-motor interface supporting short-term memory for natural speech sequences.

Methods

A detailed description of the applied methods is given in *SI Appendix, Methods*. In the following we give a brief account of our procedures.

Study Design. Sixteen patients (11 female; 19 to 50 y old; Table S1) who were evaluated for neurosurgical treatment of medically refractory epilepsy participated in the study. Prior to any experimentation, all participants provided informed consent. The study protocol was approved by the local Research Ethics Board of the University Health Network Research Ethics Board. All procedures followed the Good Clinical Practice of the University Health Network, Toronto.

Participants were asked to memorize and repeat sentences (Fig. 1 *A–D*). Each trial contained three phases: perception, silent rehearsal, and production (Fig. 1*A*). In the perception phase, participants listened to a pair of sentences (sentence 1 [S1] and sentence 2 [S2]). Then, in the silent rehearsal phase, participants were asked to silently rehearse S2. Participants were asked to silently rehearse only once, verbatim, without mouthing the words. Finally, in the production phase, participants were asked to vocalize the passage verbatim, at the same pace that they had heard it. Visual symbols on the screen cued participants to each phase.

On a given trial, participants could hear a single pairing of two sentences from a set of four possible sentences (Fig. 1*B*). The set of four sentences composed a "stimulus group," where two sentences served as S1 and two as S2. On each trial the semantic context and coherence were varied, depending on which combination of S1 and S2 was presented. In half of the trials ("coherent"), S2 was semantically coherent and was a natural semantic extension of S1 (e.g., Fig. 1 *B, Top* row). In the other half of trials ("incoherent"), S2 was semantically incoherent and did not have any obvious semantic relationship with S1 (Fig. 1 *B, Bottom* row). For both coherent and incoherent trials, participants were asked to silently rehearse S2 exactly once and then repeat S2 verbatim.

On coherent trials, the precise meaning of S2 depended on the contextual information presented in S1. In particular, by changing only the initial words within S1, we varied the interpretation of S2. Even the final words of S1 were
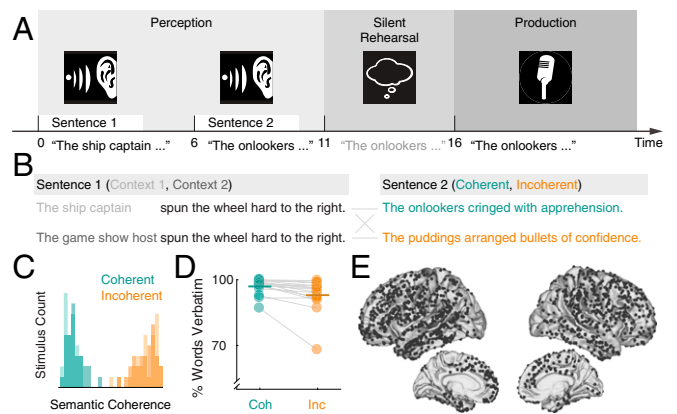

Fig. 1. Experimental design and behavioral performance. (*A*) Participants first listened to two spoken sentences (sentence 1, sentence 2). They then silently rehearsed sentence 2, verbatim, exactly once. Finally, they repeated sentence 2 aloud. (*B*) A "stimulus group" consisted of four possible sentences: 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) (*SI Appendix*, Table S2). The two versions of sentence 1 were always coherent, but the differences in the first part of the sentence resulted in very different semantic contexts for interpreting the coherent sentence 2. (*C*) Projection of sentence norming features via multidimensional scaling reveals a clear separation of the semantics of coherent and incoherent sentences (*SI Appendix, Methods* and Table S3). (*D*) Average number of words spoken verbatim by each participant (Coh, coherent trials; Inc, incoherent trials). (*E*) Combined electrode placement for all 16 subjects on the lateral and medial surfaces of the Freesurfer average brain.

shared across the two contexts. For example, the subject of the sentence in Fig. 1*B* ("ship captain" vs. "game show host") determines whether the "apprehension" in S2 is understood as "suspense about a prize" or "concern about danger."

Data Analysis. To compute a behavioral accuracy score, for each participant we computed the ratio of words recalled verbatim to the number of total words within a sentence based on the individual transcripts from the production phase. Power time courses were estimated using Morlet wavelets between 70 and 200 Hz after standard preprocessing procedures of the ECoG data (see *SI Appendix, Methods* for details). We used paired permutation tests within each participant to assess activation within trials relative to baseline and to compare activation differences of coherent and incoherent trials.

To identify brain regions that encoded stimulus-specific information, we employed correlated component analysis (22). This linear decomposition technique identifies a set of weights across channels that will produce a weighted-average signal whose Pearson correlation coefficient across repeats is maximal. Intuitively, this can be thought of as assigning greater weighting to electrodes whose activity time courses are reliable across repeats of a stimulus. This technique can also be understood as similar to principal component analysis; instead of weighting electrodes such that the variance of a single repeat is maximized, this technique weights electrodes such that the correlation between two repeats is maximized, relative to the within-repeat correlation. For this analysis, only the data from the participants completing both repeats of all 30 stimulus groups totaling 240 trials were used ($N = 8$) to maximize power and numerical stability of this procedure (see *SI Appendix, Methods* for details). The data across those subjects were pooled and grouped into nine anatomical regions of interest (ROIs), including superior temporal gyrus (STG), middle temporal gyrus (MTG), anterior temporal lobe (ATL), TPC, vSMC ($z < 40$), dSMC ($z \geq 40$), inferior frontal gyrus (IFG), aPFC ($y \geq 23$), and dPMC ($y < 23$).

We used a subsampling and cross-validation approach (see Fig. 3 *A* and *B*) to quantify the performance of the correlated component analysis. Weights were estimated for two-thirds of the data within a given task phase (S1, S2, silent rehearsal, production) and condition (coherent, incoherent) and then tested on the remaining data from that task phase. This resulted in a single reliability value ρ for each ROI. The selections of in-sample and out-of-sample data were repeated 5,000 times to obtain a more robust estimate of repeat reliability.

In some analyses (*SI Appendix*, Figs. S5 *B* and *C*, S6, and S7), the estimated weights were also applied to held-out data from a different task phase. In this case, we subsampled 5,000 times from all of the data.

To statistically assess whether the responses were stimulus specific and reliable, we tested whether the ρ values of stimulus repeats were greater than the ρ values for nonmatching (shuffled) trial pairs. We generated a distribution of 5,000 differences (ρ values for matching and nonmatching sentences for each fold of the cross-validation) and tested the proportion of values greater than zero.

To identify whether neural responses were specific to a particular moment within each stimulus, we implemented a cross-temporal analysis. This approach was inspired by cross-temporal decoding methods developed in the working memory literature (23–26). Intuitively, this analysis asks: If a particular ROI has an elevated weighted response at time t_1 on the first repeat, will the response also be elevated at time t_2 within the second repeat? If this is true, only when $t_1 = t_2$, then the neural response is specific not only to the stimulus but also to a particular moment within the stimulus. On the other hand, if the responses are correlated across repeats for all pairs of time points, then the response is stimulus specific but not time-point specific. For each ROI we generated a "response time course" using held-out data weighted according to the weights w from the correlated component analysis. To assess whether the temporal patterns differed between perception of S2 and silent rehearsal, we statistically compared three different areas of the time–time correlation matrix: the matching time points along the diagonal within a task phase (on-diagonal), the nonmatching time points within a task phase (off-diagonal), and the nonmatching time points across task phases (cross-phase; for easier visualization of these areas, see the key in Fig. 4 *D*, *Left*).

Results

Recall Behavior. We measured neural responses while 16 participants performed a sentence repetition task. The participants listened to a pair of sentences and were asked to mentally rehearse the second sentence (S2) and then reproduce it aloud (Fig. 1*A*). Half of these S2s were semantically coherent, while the other half consisted of nonsense sentences that were semantically incoherent (Fig. 1*B* and *SI Appendix*, Table S2). Participants were able to accurately reproduce both coherent and incoherent sentences in their overt recall (Fig. 1*D*), with coherent sentences recalled slightly better than incoherent sentences (coherent, 96.9% words spoken verbatim, SEM 0.9; incoherent, 93.0% words spoken verbatim, SEM 1.9; $t(15) = 3.4$, $p = 0.004$).

Broadband ECoG Activity during Sentence Perception and Rehearsal. To characterize neural activity during sentence perception and rehearsal, we focused on changes in broadband power (70 to 200 Hz) of the local electrical field measured by subdural ECoG electrodes. Amplitude changes in this frequency band are a robust estimate of population firing (27–29), provide reliable responses to audiovisual stimuli across much of the lateral cortex (30), and are sensitive to speech perception and production (16, 31, 32). Aggregating data across participants resulted in dense coverage of the cortical surface in left and right hemispheres, excepting only the occipital cortex (Fig. 1*E*).

Patterns of Sustained Activation during Perception, Silent Rehearsal, and Production. Which regions of the brain were active during either perception or rehearsal? Which regions were active during both perception and rehearsal? To quantify the overall activation, we contrasted broadband power responses during each task phase against a pretrial baseline and identified electrodes exhibiting sustained activation in each phase (lateral view, Fig. 2*E*; medial view, *SI Appendix*, Fig. S2*A*). For S1, broadband power increased above baseline over left and right superior temporal gyrus, middle temporal gyrus, and dorsal motor and premotor cortex, as well as in the left inferior frontal gyrus and anterior frontal cortex. The observation of both sensory and motor activity during passive sentence perception is consistent with prior fMRI and ECoG studies (30, 33–35). When participants listened to S2, which they subsequently rehearsed, the pattern of activation was similar, with one clear difference: a "ring" of below-baseline

activity that was sustained across widespread electrodes in inferior temporal, parietal, and dorsal frontal regions. This ring of below-baseline activity was most pronounced in the left hemisphere and was not observed during S1, suggesting that it was related to participants' active attention as they prepared for the subsequent rehearsal of S2.

Silent rehearsal produced an activation pattern that was different from the perception phases: Broadband activity was sustained above baseline only in the vicinity of the left motor and premotor cortex as well as in a very small number of lateral and inferior frontal electrodes bilaterally and in one left posterior temporal site. Widespread electrodes across frontal, temporal, and parietal cortex reduced their activity below baseline. To the extent that sustained increases in broadband power, an estimate of population firing, index functional activation (36), these data most strongly implicate the left motor and premotor cortices in silent rehearsal.

The same patterns observed in the trial-averaged activity were present in time-resolved single-trial responses (Fig. 2 *A*–*D*). For example, a typical electrode in the superior temporal gyrus (Fig. 2*A*) exhibited broadband power modulations that reflected auditory responses: They increased during the presentation of each sentence, showed little modulation during silent rehearsal, and increased again during production, when the participants heard their own voice (37). An example electrode in ventral motor cortex (Fig. 2*C*) expressed slightly elevated activity during both perception and rehearsal, with the strongest modulations during production. An electrode in the premotor cortex (Fig. 2*D*), the area active during both perception and rehearsal, exhibited consistent single-trial responses during all phases. Consistent with the observations of Cheung et al. (14), broadband power responses in the superior temporal gyrus and premotor cortex tracked the stimulus acoustics. For example, the sample electrodes shown in Fig. 2 *A* and *C* exhibited single-trial correlations with the auditory envelope of the sentence of $r = 0.47$ (superior temporal gyrus) and $r = 0.57$ (premotor cortex).

To visualize the time-resolved activation patterns at a regional level, we grouped the electrodes into nine anatomically defined ROIs (Fig. 2 *F*, *Middle*). Electrodes exhibiting above- or below-baseline activity in at least one task phase in any stimulus condition (coherent or incoherent) were included in the regional summary. The average time courses of the ROIs reflect the overall activation pattern in Fig. 2*E*. Although we observed above-baseline responses during sentence perception, the mean broadband power in all ROIs trended below baseline during rehearsal, just until the onset of production (Fig. 2*F*).

Sensorimotor and Premotor Sites Coactive during Perception and Silent Rehearsal. Neural circuits that exhibit increased activity during both perception and silent rehearsal of a sentence are candidates for supporting the short-term memory of that sentence. Across participants, motor and premotor sites appear to be the areas most consistently involved in both perception and silent rehearsal phases of the task (Fig. 2*E*). To spatially refine this finding, and to rule out the possibility that perceptual and rehearsal activations occurred in different subsets of participants, we measured the conjunction of sites within patients that were statistically active above or below baseline in both task phases (149 sites in total). This analysis confirmed that almost all sites active above baseline in both task phases (17 of the 20 sites) were clustered around sensorimotor and premotor cortex of the left hemisphere (Fig. 2*G*, red electrodes).

Beyond the motor cortices, we also observed sustained coactivation in one electrode in the left posterior superior temporal cortex, one in the right middle superior temporal gyrus, and one in the right inferior frontal gyrus. Although there are few such electrodes, their locations are consistent with prior literature implicating posterior temporal and inferior frontal regions in
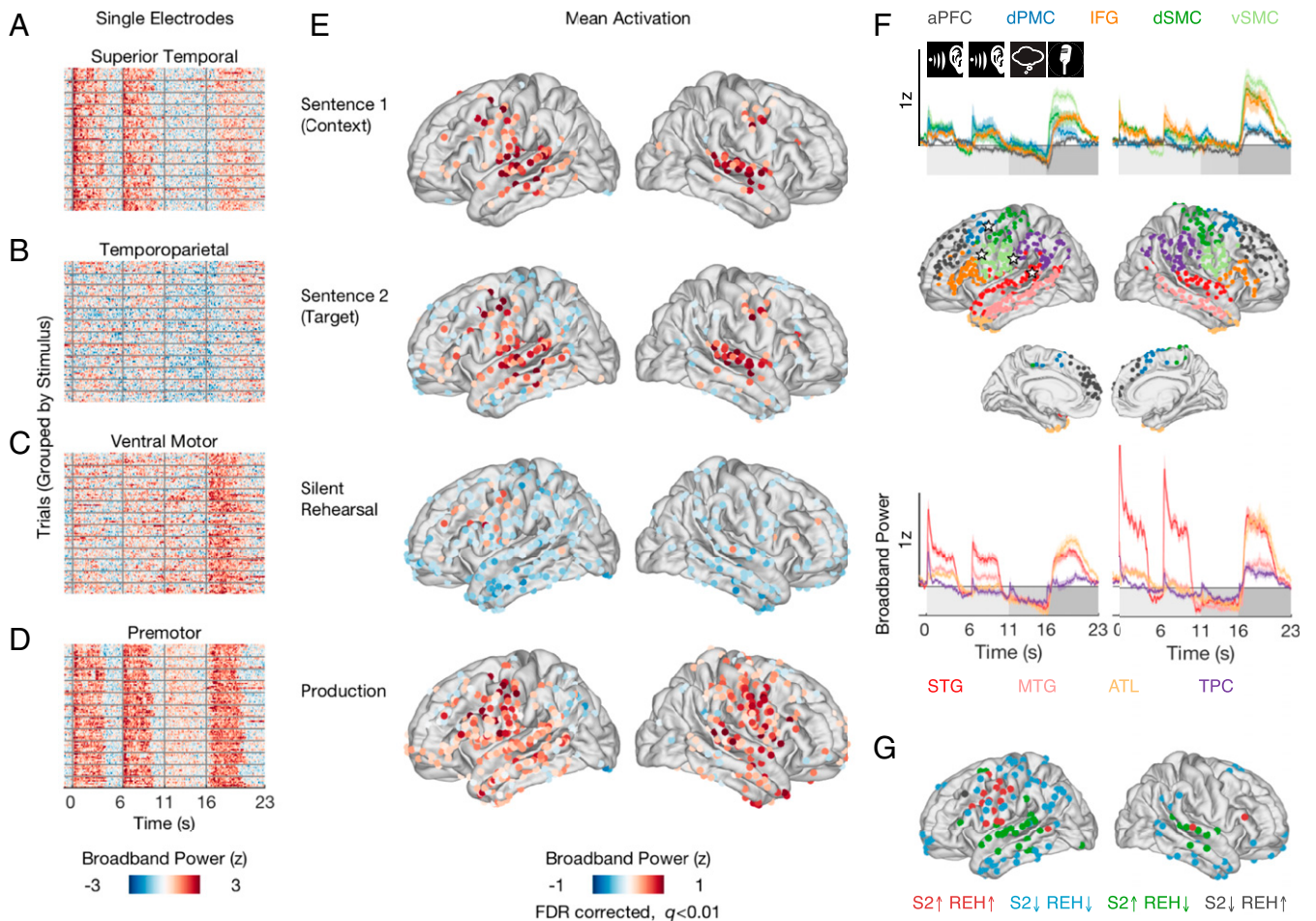
Fig. 2. Patterns of mean activation during perception, silent rehearsal, and production. (*A–D*) Stacked single-trial responses for example electrodes in superior temporal (*A*), temporoparietal (*B*), ventral motor (*C*), and premotor cortex (*D*). Each row illustrates broadband power for a single trial. Broadband power changes were consistent across trials within an electrode but differed across brain regions. Locations of the example electrodes are indicated by stars in *F, Middle*. Onset of sentence 1 and sentence 2, silent rehearsal, and production are indicated by vertical gray bars. Horizontal gray bars separate trials into "stimulus groups" (*SI Appendix*, Table S2). (*E*) Spatial distribution of broadband power responses aggregated across participants and conditions for each task phase. Only electrodes activated at the level $q < 0.01$ after FDR correction are shown. (*F*) Average time courses and electrode groupings within ROIs, split by left and right hemispheres. *Top* shows the average time courses for aPFC (gray), dPMC (dark blue), vSMC (light green), dSMC (dark green), and IFG (orange). *Bottom* shows the average time courses for STG (red), MTG (pink), ATL (yellow), and TPC (purple). Onset of perception (sentence 1, sentence 2), silent rehearsal, and production are indicated by icons and shading. Note that some electrodes appear outside their designated ROI: The imprecision in location arises in the process of mapping from individual anatomy to a standardized MNI coordinate system. (*G*) 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 (S2) and silent rehearsal phase (REH). Sites that are activated above baseline in both task phases (red) cluster around sensorimotor and premotor cortex.

verbal working memory tasks (6, 38). Moreover, several other temporal and parietal sites exhibited transient activations at the onset of the rehearsal phase (*SI Appendix*, Fig. S8), which may indicate a role in initiating the rehearsal process.

Sites that activated during S2 and deactivated during rehearsal (36 sites, green electrodes) clustered around superior and middle temporal cortex. Sites that were deactivated in both task phases (92 sites, blue electrodes) were mostly distributed around temporal and motor cortex (see deactivated sites during S2 in Fig. 2*E*). Finally, there was one site in left frontal cortex that was activated during rehearsal but deactivated during perception (gray electrode). The overall results of this analysis are: first, activation sustained across both perception and rehearsal in sensorimotor and premotor areas (21, 39) and, second, widespread suppression of rehearsal activity in a ring of surrounding temporal and parietal sites.

We additionally analyzed the relationship between activation of each electrode and the accuracy of speech production behavior on each trial (*SI Appendix*, Fig. S3B). These behavioral

analyses indicated that motor and premotor sites are important sites for memory-dependent behavior. However, to our surprise, larger responses in motor and premotor sites were associated with a greater probability of error. Because of the use of naturalistic speech stimuli in this experiment, the behavioral associations may be confounded with stimulus properties; e.g., sentences that are easier to remember because of their semantic content may also be sentences which produce less broadband power activation during perception. Thus, while the broadband power in motor and premotor sites was correlated with behavioral performance, we cannot unambiguously interpret this connection.

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. In addition, the mean activation maps (Fig. 2*E*)
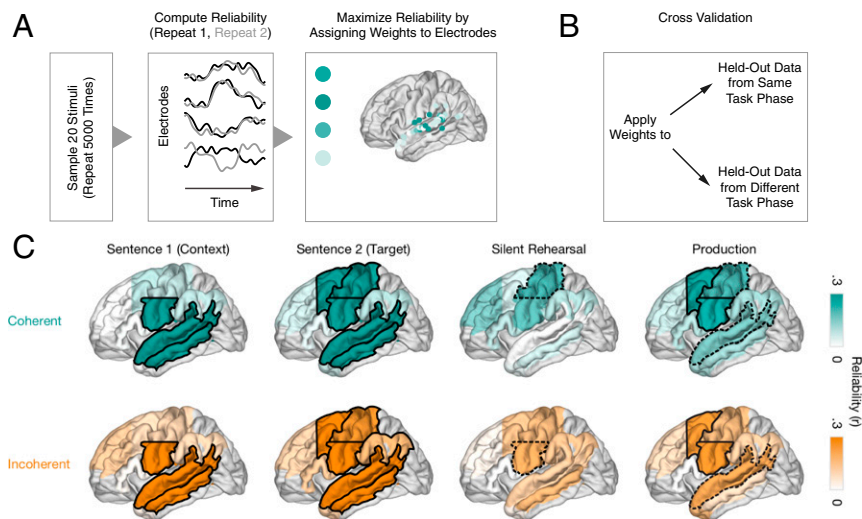
Müsch et al.

Fig. 3. Weighted reliability analysis reveals sentence-specific encoding in sensory and sensorimotor cortices. (*A*) Feature selection based on a subset of data. Reliability is maximized in a ROI by assigning different weights to electrodes. Data from eight participants who completed all trials was entered into this analysis. (*B*) Cross-validation using the weights from *A* on out-of-sample data from the same task phase (Figs. 4 and 5 and *SI Appendix*, Figs. S5A and S6) or from a different task phase (Fig. 5 and *SI Appendix*, Figs. S5 *B* and *C* and S7). (*C*) Sentence-specific encoding for coherent (teal, *Top*) and incoherent (orange, *Bottom*) sentences during perception (sentence 1, sentence 2), silent rehearsal, and production by ROI. The solid black line indicates significance at *q* < 0.05 (FDR corrected across ROIs), and the dashed black line indicates significance at *P* < 0.05 (uncorrected).

emphasize areas exhibiting sustained activation changes, while not considering the temporal profile of each electrode's responses. Therefore, we tested whether activity time courses in each region were sentence specific. To this end, we applied a technique for measuring region-level response reliability (Fig. 3 and *Methods*, *Data Analysis*) to quantify sentence-specific responses in nine anatomical ROIs (Fig. 2*F*). In each ROI, and separately for coherent and incoherent sentences, we used a subsample of stimuli to estimate a set of electrode weights (i.e., a spatial filter) that maximizes the reliability of responses across stimulus repeats (Fig. 3*A*). Then, to obtain an unbiased estimate of how much sentence-specific information was contained in the time courses of each ROI, we measured the reliability of the weighted responses during the same task phase, but in different (out-of-sample) sentences and their repeats (Fig. 3*B*).

During perception, we observed that auditory, linguistic, and motor cortices of the left hemisphere exhibited the most reliable and sentence-specific responses to both coherent and incoherent sentences (Fig. 3*C*; see *SI Appendix*, Fig. S4 for electrode-level reliability). The median reliability, ρ, and associated *P* values for each of the ROIs are listed in *SI Appendix*, Table S5. During the auditory presentation of S1 (which did not have to be rehearsed), sentence-specific response time courses were observed in the left STG and MTG as well as in left vSMC. During the presentation of S2 (the target sentence for rehearsal), we additionally observed sentence-specific information in left dSMC and in left dPMC. These effects were observed for both coherent and incoherent sentences. The left TPC also exhibited sentence-specific reliability for incoherent but not coherent sentences. In the right hemisphere (*SI Appendix*, Fig. S5A), sentence-specific responses were observed in the STG both during S1 and S2 and for coherent and incoherent sentences. Sentence-specific information was also present in the right vSMC, but this effect achieved statistical significance only for the coherent sentences.

During silent rehearsal, we observed weakly reliable sentence-specific responses in left sensorimotor cortices (*P* < 0.05, uncorrected; Fig. 3). The effect in the left dSMC achieved statistical significance only for the coherent sentences (reliability ρ = 0.20, *P* = 0.019) but was similar for incoherent sentences (ρ = 0.19, *P* = 0.054). Likewise, the effect in the left vSMC achieved

statistical significance only for the coherent sentences (ρ = 0.19, *P* = 0.033) but was similar for incoherent sentences (ρ = 0.15, *P* = 0.072). Although the reliability in these sites is weaker than what was observed in the perceptual phases, it is sufficiently robust to replicate across the independent sets of coherent and incoherent sentences. No sentence-specific responses were observed in the right hemisphere (*SI Appendix*, Fig. S5A)

During production, we observed reliable sentence-specific responses in the left vSMC and dPMC. Weakly reliable responses (*P* < 0.05) for both coherent and incoherent sentences were observed in the left and right STG (*SI Appendix*, Fig. S5A), the right vSMC, and right TPJ. Responses in the right MTG approached significance for incoherent sentences.

The sensitivity of our sentence-locked reliability measure may be reduced by trial-to-trial temporal jitter during the rehearsal and production of speech. If participants mentally rehearse faster or slower on some trials, this could reduce the match of neural dynamics across trials. To reduce the ambiguity introduced by trial-to-trial behavioral jitter, we 1) conducted a time-resolved analysis to distinguish time-locked and temporally diffuse aspects of the reliable responses (Fig. 4 and *SI Appendix*, Fig. S6) and 2) implemented cross-phase reliability metrics, which have greater statistical power to detect reliability (Fig. 5 and *SI Appendix*, Fig. S7).

Temporal Precision of Sentence-Specific Neural Activity. The previous analysis revealed sentence-specific neural activity that was most robust in the temporal lobe and in sensorimotor and premotor cortices—But is this sentence-locked activity specific to individual moments within each stimulus? To answer this question, we compared the sentence-specific activity across all pairs of individual time points during perception of S2 and during rehearsal. Using the weights optimized for each ROI in each fold of our cross-validation procedure, we computed the correlation of a specific moment's activity level in the first presentation of a stimulus, comparing it against all possible time points in the second presentation of that stimulus for all of the held-out sentences.

The time-specific correlation analysis produces a 2D time–time correlation matrix (left hemisphere, Fig. 4; right hemisphere, *SI Appendix*, Fig. S6). The diagonal of the matrix reveals correlation
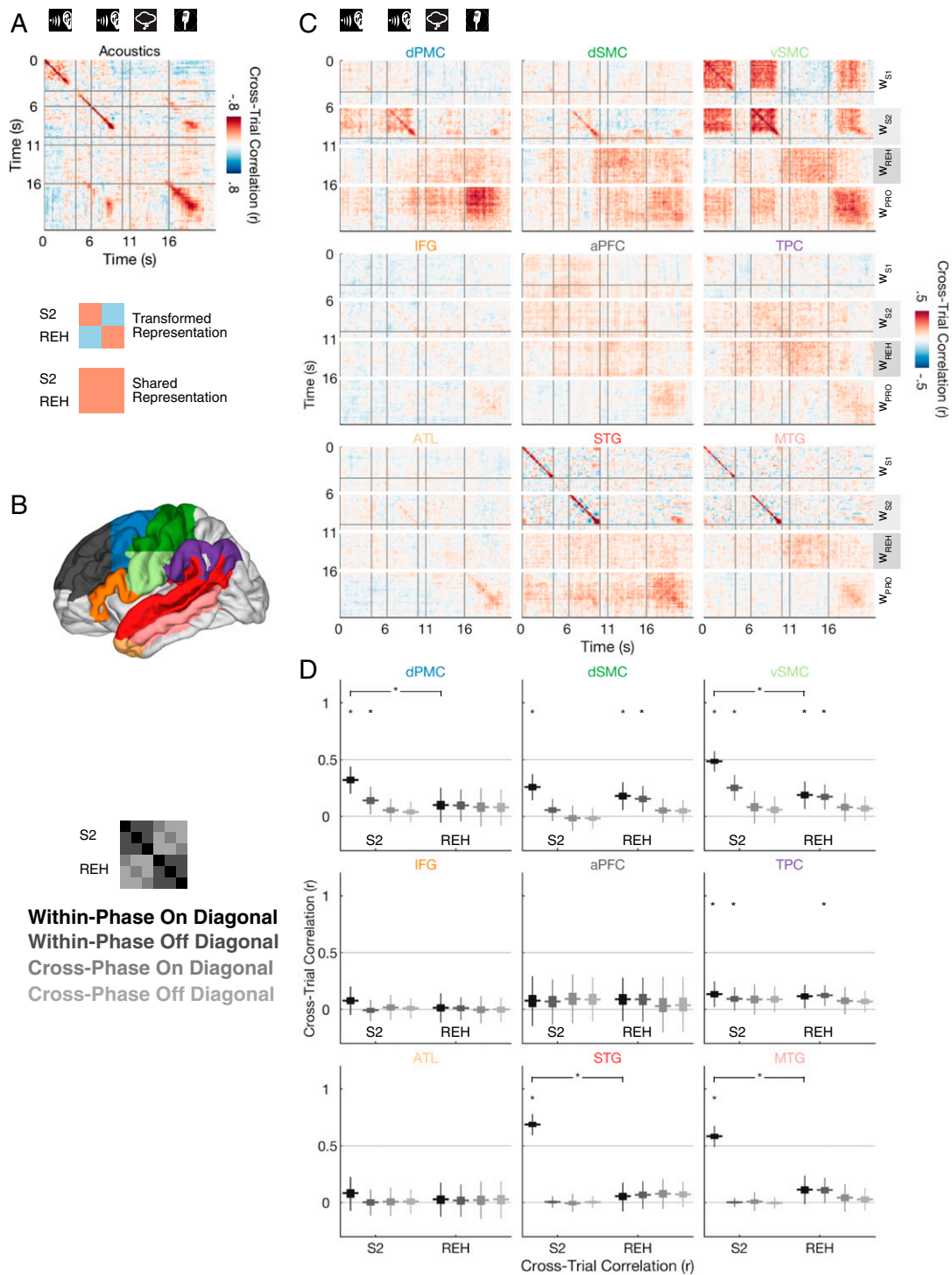
Müsch et al.

Fig. 4. Temporal specificity of sentence-locked activity. (*A*) Cross-time, cross-trial correlation maps for the amplitude of the recorded acoustics. A positive correlation at time (t_1, t_2) indicates that higher amplitude values at time t_1 on the first presentation of a sentence were associated with higher amplitude values at time t_2 on the second presentation of the sentence. Thus, a positive correlation along the diagonal ($t_1 = t_2$) indicates that amplitudes are not only sentence specific but also locked to individual time points within a stimulus. Correlation values are averaged across coherent and incoherent sentences. The vertical gray lines indicate the onsets of the silent gap following S1 and S2, silent gap following S2, and the silent rehearsal cue, respectively. (*B*) Schematic of color-coded ROIs. (*C*) Cross-time, cross-trial correlation maps of the neural data within each ROI (similar to *A*, but with correlations computed for weighted cortical broadband power timecourses). All correlations along one row of the time–time correlation matrix are computed using the region-specific weights derived from a single phase (e.g., perception, rehearsal, or production phase). W_{S1}, weights from the S1 perception phase; W_{S2}, weights from the S2 perception phase; W_{REH}, weights from the rehearsal phase; W_{PRO}, weights from the production phase. (*D*) Statistical summary of the cross-time, cross-trial correlation values shown in *A* ($q < 0.05$, FDR corrected across ROIs and number of comparisons within phase). The distribution of correlation values is shown for the within-phase on-diagonal component ($t_1 = t_2$, black), the within-phase off-diagonal component (dark gray), the cross-phase on-diagonal component (lighter gray), and the cross-phase off-diagonal component (lightest gray) of the correlation matrix. Results are summarized separately for the S2 phase and the REH phase for each ROI. The thicker 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.
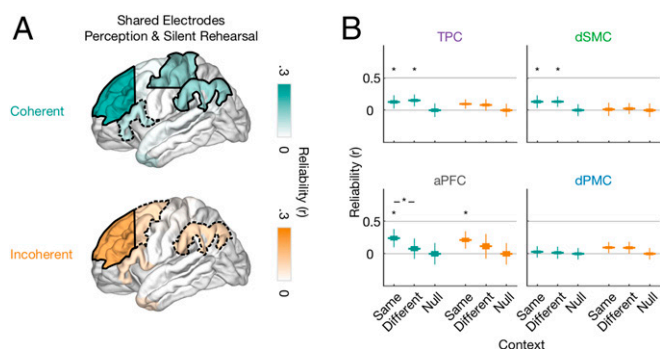
Fig. 5. Prefrontal electrodes encode sentence-specific high-level semantics, shared between perception and rehearsal. (*A*) Reliability for shared electrodes during perception (sentence 2) and silent rehearsal that encode sentence-specific information for coherent (teal, *Top*) and incoherent (orange, *Bottom*) sentences. Weights from sentence 2 were applied to time courses from silent rehearsal. The solid black line indicates significance at $q < 0.05$ (FDR corrected across ROIs), and the dashed black line indicates significance at $P < 0.05$ (uncorrected). (*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, and differences between same vs. different contexts are indicated by lines, FDR corrected at $q < 0.05$ across all four ROIs and three tests performed in each of the conditions.

at matching time points (time t in repeat 1 matched against the same point in repeat 2). If sentence-specific activity in a region varies from moment to moment, then the time–time correlation values along the diagonal of the matrix will be higher than the off-diagonal values. Conversely, if the sentence-specific activity pattern evolves more slowly and is not specific to individual moments in the stimulus, or if the neural signal has unreliable timing across trials, then the time–time correlation pattern will be more diffuse. In this case, nonmatching time points will also be correlated across repeats, and off-diagonal values will be as high as the on-diagonal values (simulations in *SI Appendix*, Figs. S9 and S10). For the statistics, in each fold we averaged the values within a specific area of the correlation matrix (within-phase on-diagonal, within-phase off-diagonal, cross-phase on-diagonal, cross-phase off-diagonal; see key in Fig. 4*D*, *Left* for visualization) for the S2 phase and the silent rehearsal phase, and we compared the distributions of foldwise values across phases (*Methods*).

To confirm the temporal reliability of the stimulus presentation and the participants' speech production, we began by applying a time–time correlation analysis to the acoustic signals from the experiment room (Fig. 4*A*). The sounds in the experiment room include those generated by the computer speakers (generating the auditory stimuli S1 and S2) and also those generated by the human participants (who spoke their verbatim memory of the stimuli in the production phase). The cross-trial correlation of auditory loudness (Fig. 4*A*) revealed three key features. First, the diagonal entries in S1 and S2 are the highest values in the correlation map ($r \sim 0.96$), consistent with the time-locked presentation of auditory stimuli. Second, participants can reliably produce speech in the same temporal profile across trials, as shown in the on-diagonal correlations in the production phase. The blur around the diagonal in the production phase also provides a measure of the cross-trial jitter in the production of speech. Third, the auditory envelope of the speaker's own production matches the envelope of the auditory speech the speaker was instructed to remember: There is an on-diagonal correlation across trials, across the S2 perceptual phase, and across the

production phase. Together, these data demonstrate that task presentation and performance were sufficiently temporally precise for detecting time-locked response, even in the presence of jitter arising from the participant's production variability.

In the left hemisphere, we observed three main types of sentence-specific response patterns, with varying degrees of temporal specificity (Fig. 4 *C* and *D*; with statistics for all ROIs in *SI Appendix*, Table S6). First, the STG and MTG exhibited sentence-locked activity only along the time–time diagonal and only during sentence perception. Thus, the responses in STG and MTG regions are locked to specific moments within a sentence during perception, in line with its proposed role in extracting lower-level perceptual and linguistic features (14).

A second pattern of stimulus-specific dynamics was observed in the TPC: Here the sentence-specific responses were temporally diffuse and present on the on- and off-diagonal, for both perception and rehearsal phases. Moreover, there was an indication that the reliable activity patterns were shared across perception and silent rehearsal: The signals during the S2 phase correlated diffusely with those during the rehearsal phase (Fig. 4*D*, TPC). Although the data numerically indicate a cross-phase effect ($r = 0.1$, $P = 0.04$ on-diagonal; $r = 0.1$, $P = 0.01$ off-diagonal), it did not surpass the statistical threshold for multiple comparisons in this analysis. In subsequent analyses with greater statistical power (Fig. 5), we reexamined the possibility of shared signals perceptual–rehearsal in the TPC and aPFC.

Finally, in motor and premotor regions (vSMC, dSMC, and dPMC) we observed a temporal encoding pattern that indicated a transformation of representations from the perception of S2 to silent rehearsal (Fig. 4 *C* and *D*). During perception of S2, sensorimotor circuits exhibited robust time-locked sentence-specific activity (i.e., elevated correlations along the diagonal), combined with off-diagonal, temporally diffuse activity that significantly differed from zero in all motor areas except the dSMC. During rehearsal, sentence-specific activity in sensorimotor circuits was more temporally diffuse. In the dSMC and vSMC, for example, activity was locked to the specific sentence (within-phase off-diagonal correlations >0), but it was not locked to the specific moment within the sentence when it was rehearsed (within-phase on-diagonal values were not higher than within-phase off-diagonal values).

Crucially, the sentence-specific activity in motor and premotor regions was distinct between perception and rehearsal phases: The reliable activity during perception was not correlated with the stimulus-locked activity during rehearsal in any of the three regions (Fig. 4*D*, vSMC, dSMC, and dPMC, cross-phase). This pattern of results is not consistent with a simple mirroring of perceptual activity during rehearsal; instead the data suggest that the sensorimotor representations of the same sentence differ across task phases. Indeed, examining the cross-phase S2-rehearsal correlations, we did not find a time-locked replay (on-diagonal correlations greater than off-diagonal correlations) in any region (all Ps > 0.08). Additionally, in the vSMC and dSMC, the cross-phase S2-rehearsal correlations were less than the within-phase correlations (on-diagonal comparison, cross-phase versus within-phase Ps < 0.001). Thus, we found no evidence of time-locked perceptual dynamics being replayed during rehearsal. Finally, we used simulations to confirm that a "perceptual replay" process could not account for the pattern of data observed in sensorimotor cortex (*SI Appendix*, Fig. S9, *Methods*, and *Simulation Results*). In particular, if speech rehearsal was a replay of the perceptual process, either temporally jittered or exact, then the cross-phase (perception-to-rehearsal) correlations should not be zero at the same time as the within-phase correlations (rehearsal-to-rehearsal) are greater than zero.

Few areas in the right hemisphere elicited reliable responses (*SI Appendix*, Fig. S6). The right STG and vSMC exhibited sentence-locked activity along the diagonal during perception,

with the STG and vSMC also showing a significant difference between perception of S2 and silent rehearsal. This suggests that the right STG and vSMC respond to auditory and phonological information in the stimulus, as do their left hemispheric counterparts.

High-Level Semantic Representations Shared across Perception and Rehearsal. Given the temporally nonspecific responses in frontal and parietal electrodes, are these areas encoding high-level syntactic or semantic features? Some models of verbal short-term memory have suggested that higher-level sentence features (semantics and syntax) are stored as the primary short-term memory trace and later used to regenerate information patterns at the time of sentence recall (17, 18). Therefore, we tested whether the shared sentence-specific representations in each ROI were sensitive to the high-level semantics of the stimuli. We did so in three steps. First, we conducted an analysis to directly confirm what is suggested by Fig. 4: that common sets of prefrontal and parietal electrodes encode sentence-specific information during both perception and rehearsal. Second, we measured whether this sentence-specific activity in each region during silent rehearsal was the same across the two different semantic contexts that were generated by the nonrehearsed 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 Fig. 1*B* for both contexts). Finally, we confirmed that the same spatiotemporal patterns were generated across perception and rehearsal phases in the prefrontal and temporoparietal cortex (*SI Appendix*, *SI Text* and Fig. S7).

We first confirmed that prefrontal and parietal electrodes represent sentence-specific information in common sets of electrodes across the perceptual and rehearsal phases. To directly test for this phenomenon, we applied our cross-validation procedure to the rehearsal phase using weights that were identified during perception of S2. Since the perceptual and rehearsal phases are completely separate data, this also enabled us to increase the power of our cross-validation procedure: We used the average weights estimated on 20 stimulus groups from the perceptual phase (Fig. 3*C* and *SI Appendix*, Fig. S5*A*) and measured the repeat reliability, ρ, in the full set of 30 (instead of 10) held-out stimulus groups from the rehearsal phase. Using this approach, we identified sentence-specific activity with shared electrode weights in aPFC. This cross-phase effect was observed separately for both coherent and incoherent sentences (Fig. 5*A*). Thus, even though the sentence-specific activity in these high-level areas is weak and temporally diffuse (Fig. 4*C*), it is expressed in a common set of electrodes across perception and silent rehearsal of S2. We also observed significant sentence-specific information (with shared electrode weights) in TPC and dSMC for coherence sentences. Weakly sentence-specific activity with shared electrode weights was observed in the left IFG for coherent sentences only and in the left dPMC and TPC for incoherent sentences. No sentence-specific activity with shared electrode weights was observed in the right hemisphere (*SI Appendix*, Fig. S5*B*).

Having observed that parietal and prefrontal electrodes encode sentence-specific information in common electrodes across task phases, we asked whether this information tracked the contextual meaning of the sentences. Neuroimaging studies of semantic representation (1) reliably implicate prefrontal and temporoparietal areas, and some electrodes in these areas did increase their activation when participants processed coherent (rather than incoherent) sentences (*SI Appendix*, Fig. S3*A*). Thus, because the information in S1 altered the meaning of S2 that was later rehearsed, we tested whether each ROI's representation during rehearsal was sensitive to the contextual information provided in S1. We found that, as participants rehearsed S2, only the aPFC was sensitive to the contextual information

provided by S1 ($q < 0.05$, Fig. 5*B*). While this context effect survived false discovery rate (FDR) correction only for coherent sentences (Fig. 5*B*, teal), the pattern of activity was similar for incoherent sentences (Fig. 5*B*, orange). In all other ROIs the sentence-specific dynamics were not detectibly affected by the contextual information of S1 (Fig. 5*B*). This suggests that the regions most robustly encoding sentence-specific information during silent rehearsal (i.e., motor and premotor regions) were not representing high-level sentence context. Instead, the high-level semantic content of each sentence appeared to affect only the temporally diffuse representations of the aPFC.

Discussion

When we rehearse a spoken sentence in our mind, do our brains replay the neural dynamics from the earlier sentence perception? We recorded intracranial neural activity across widespread language-related networks as people heard and then mentally rehearsed spoken sentences. In each brain region, we tested whether silent rehearsal of sentences involved reactivation of sentence-specific representations established during perception or transformation to a distinct sentence-specific representation. We found evidence for both processes: Transformation was most apparent in sensorimotor and premotor cortex, while prefrontal and temporoparietal cortices maintained a more static representation that was shared across perception and rehearsal. In the aPFC and TPC, where representations were shared across phases, neural representations were also more sensitive to changes in the syntactic structure and contextual semantics of the sentences.

The data implicate sensorimotor cortex (both vSMC and dSMC) and PMC as active bridges between the perception and rehearsal of spoken sentences. These regions were the only ones for which neuronal activity, as indexed by broadband power, was consistently increased during both silent rehearsal and perception (Fig. 2*G*). Moreover, vSMC and dSMC exhibited stimulus-specific encoding during both the perceptual phase and the silent rehearsal phase (Fig. 3*C*). The stimulus-specific encoding in these sensorimotor cortices was temporally precise during perception (Fig. 4). Finally, broadband power in motor and premotor areas during rehearsal was related to the accuracy of sentence memory (*SI Appendix*, Fig. S3*B*). These data build on prior ECoG reports of short-latency encoding of auditory speech in SMC and PMC (14, 21, 40), now revealing how the representations of complex speech in motor areas persist and transform between perception and rehearsal.

Sentence-specific activity in vSMC, dSMC, and PMC was transformed between the perception and silent rehearsal of the same sentence. A moment-by-moment comparison of the sentence-specific activity (Fig. 4) revealed a transition between distinct representations in perception and silent rehearsal. Sentence-specific activity patterns identified in the vSMC, dSMC, and PMC during perception did not extend beyond the end of that phase. Instead, as the perceptual phase ended, a distinct rehearsal-specific activity pattern became reliable in sensorimotor cortices. Thus, the present data connect SMC and PMC to a transformation process that supports short-term memory for natural spoken language. These findings are consistent with a sensorimotor transformation model that has been proposed for working memory (16, 41); they are not consistent with models that posit a one-to-one mirroring of activity during perception and production (11, 42).

Short-term memory for phonological information is thought to be supported by an auditory–motor interface (2, 43–45). The present data indicate that the SMC and PMC could possess both the auditory and motor representations necessary for such an interface. The vSMC activity in speech perception likely reflects more low-level and obligatory audio-motor representations (14): This area exhibited reliable sentence-specific activity even for S1

Müsch et al.

(which did not have to be rehearsed), and its reliability increased only marginally for S2 (which had to be rehearsed; Fig. 4*C*). By contrast, the dSMC and PMC were more sensitive to the task relevance of the speech input: These areas exhibited greater sentence-specific reliability for S2 (which had to be rehearsed), relative to S1 (Fig. 4*C*). We tentatively propose that the vSMC representations directly track auditory and motor representations, while the time-locked dSMC and PMC responses are purposed not only toward control of articulatory sequences, but also toward expectation of sensory sequences (e.g., ref. 46). A short-term memory trace of speech in sensorimotor cortices could also explain why these areas would be engaged for discriminating speech stimuli in noise (47).

In contrast with the sensorimotor circuits discussed above, the prefrontal and temporoparietal cortex exhibited a temporally diffuse sentence-specific activity pattern, which was shared across the perception and silent rehearsal of the same sentence. Although some of the sentence-specific activity in motor sites was also shared across phases (Fig. 5 and *SI Appendix*, Fig. S7), these shared effects were inconsistent across conditions and made up only a small proportion of the reliable signal in motor circuits (Fig. 4*D*). By contrast, the cross-phase correlation in aPFC and in the TPC was almost as large as the within-phase correlations (Fig. 4*D*, compare within-phase and cross-phase). At the level of aggregate activation, the prefrontal and temporoparietal circuits exhibited increased broadband power responses for coherent sentences (*SI Appendix*, Fig. S3*A*), consistent with their role in a "semantic network" (1).

Prefrontal cortex and temporoparietal cortex have long temporal receptive windows, exhibiting slower population dynamics than sensory cortices (30, 48) and responding to new input in a manner that depends on prior context (35, 49). Consistent with this functional profile, the temporal activation pattern in the left aPFC was sensitive to prior contextual information (changes in the high-level situational meaning of S2 based on the context provided by S1). The slowly changing sentence-locked signals in higher-order areas may provide a persistent "scaffold" which supports "regeneration" of detailed surface features of the sentence, e.g., phonemes and prosody, from more abstract properties, e.g., syntax and semantics (17, 18, 50). Models of cortical dynamics suggest that a static (or slowly changing) spatial pattern of higher-order drive regions can be used to elicit complex dynamics in lower-level regions, as long as the lower-level regions contain the appropriate recurrent connectivity (51). Thus, slowly changing higher-order dynamics may provide a control signal for faster sequence generation, which may unfold in lower-level motoric representations. More generally, the slowly changing prefrontal representations we observed in high-level cortical areas are consistent with a distributed, drifting cortical "context" that recursively interacts with episodic retrieval processes (52). We note, however, that the temporal context effects we observed (Fig. 5*B*) were small relative to those in prior studies using naturalistic stimuli (e.g., ref. 35); rich and extended narrative stimuli likely generate a more powerful contextual modulation than the single preceding sentence (S1) that we used in the present design.

Sentence-specific information changed most rapidly in the posterior areas of frontal cortex (i.e., in motor cortex) and changed more slowly toward anterior regions (i.e., premotor and prefrontal cortex, Fig. 4). The distinct timescales of these frontal areas parallel the recent observation of distinct working memory functions for populations of neurons with distinct dynamical timescales (53, 54). Prefrontal neurons with intrinsically short timescales, as measured by spontaneous firing dynamics, responded more rapidly during the perceptual phase of a task, whereas those with intrinsically longer timescales encoded information in a more stable way toward a delay period (54). In our task, faster dynamics seem to be more important in sensorimotor transformation (from perception to subvocal rehearsal),

whereas slower dynamics were associated with areas whose function may be more semantically sensitive. In future work, we will directly characterize the relative timing across these different stages of processing, to determine changes in the direction of information flow across perception and rehearsal.

The stimulus acoustics were delivered in a temporally precise manner from trial to trial during the perception phases, but the timing of stimulus rehearsal could jitter from trial to trial, due to natural human variations in speech rate and prosody. Using simulations (*SI Appendix, SI Text* and Fig. S9), we confirmed that trial-to-trial variability in rehearsal rates could generate temporally diffuse time–time maps. Thus, the diffuse time–time maps in the motor cortices during rehearsal may arise from trial-to-trial jitter in the motoric signals. However, the time–time correlation analysis can still detect the presence of temporally precise (e.g., syllable-locked) signal subject to natural behavioral jitter. This is demonstrated in the time–time correlations of the speech envelope (Fig. 4*A*), which incorporate the influence of trial-to-trial jitter as participants speak the same sentence aloud on separate trials. In this acoustic analysis, the time-resolved correlation analysis still revealed speech acoustics clustered around the diagonal of the time–time correlation map (Fig. 4*A*, production) as well as time-locked "replay" of the acoustics from perception during production (Fig. 4*A*, cross-phase S2 production). Moreover, rehearsal timing cannot explain all observations of temporally diffuse signals, because we observed slow neural signals in the time-resolved analysis even during the perception phase (Fig. 4*C*, TPC and aPFC maps during S2 phase). Since temporally nonspecific neural dynamics may also be at play during rehearsal, future studies could use alternative behavioral methods (such as asking participants to sing sequences of words) which can sharpen the temporal reliability of speech production. However, in the present analysis our simulations (*SI Appendix*, Fig. S9) confirmed that a simple "jittered replay" of perceptual signals could not account for the motor cortical dynamics during rehearsal.

Unexpectedly, we observed only minor differences in the neural processes supporting short-term memory for coherent and incoherent sentences (Fig. 3 and *SI Appendix*, Fig. S5). Both coherent and incoherent sentences were represented in a stimulus-specific manner within essentially the same sets of brain regions, albeit with different levels of mean broadband power activation. This suggests that even the incoherent sentences possessed sufficient high-level structure (e.g., lexical semantics and aspects of syntax), similar to that used to represent the coherent sentences. This minor effect of sentence incoherence is consistent with the general observation that semantic implausibility has less of an effect on sentence processing than strong syntactic violations (55).

During silent rehearsal, we expected to observe, but did not observe, sentence-specific activity in the ATL (4), the posterior STG and planum temporale (43), and the IFG (56). In addition, one electrode in the posterior temporal cortex was jointly active in both perception and rehearsal, but we did not find stimulus-specific decoding in the superior temporal cortex as a whole during silent rehearsal. It is possible that stimulus-specific encoding could be identified in these areas given a larger pool of patients, denser electrode coverage, or depth electrode coverage targeting sulci (especially the Sylvian fissure). Another possibility is that, when participants rehearse temporally extended stimuli (in this case 3 to 4 s of speech), the verbal working memory signals in temporoparietal and inferior frontal areas are more transient, locked to the onset of rehearsal, while those in the motor cortex are more sustained throughout rehearsal. Indeed, onset transients are apparent in the regional activations of premotor cortex (dPMC), TPC, and IFG at the onset of the rehearsal phase (Fig. 2*F* and *SI Appendix*, Fig. S7). Thus, in addition to the sustained rehearsal-related activity in motor cortices

(Fig. 2 *E* and *G*), there may be more transient signals at the onset of rehearsal in temporoparietal sites. A transient temporoparietal response, followed by more sustained motor-mediated rehearsal, would be consistent with neuroimaging studies (38) and lesion studies (57) which implicate temporoparietal sites in verbal working memory.

We also did not analyze phase-locked field potentials, which are an important area for future analysis of these data: Rhythmic activity in the beta band in the IFG is reliable and specific during maintenance of a simpler set of verbal stimuli (26). Analyzing the broadband power changes in the IFG, thought to index firing-rate shifts, we observed borderline effects: There were seemingly reliable responses during the perception phase, but these did not exceed our statistical threshold in the right IFG (*SI Appendix*, Fig. S6) or the left IFG (Fig. 4*D*). We further note the most inferior electrodes in the vSMC region are near the Sylvian fissure and thus could receive some signal generated in the auditory cortex; although our reliability analyses are surprisingly robust to noise (*SI Appendix*, Fig. S10), we cannot rule out the possibility that the vSMC dynamics we measured were biased by auditory signals. Finally, although we have broad coverage, we cannot rule out the possibility of false negatives. Thus, our data speak most clearly for the stimulus-specific broadband power responses of lateral cerebral cortex during perception and rehearsal of natural

language and especially for the highly reliable sentence-locked signals in sensorimotor circuits.

Overall, our data suggest that sensorimotor and premotor cortices can support an audio-motor interface proposed by leading models of verbal short-term memory. In the SMC and PMC we observed extensive joint activation across perception and rehearsal and rehearsal activity which predicted the accuracy of later sentence production. In parallel with this possible audio-motor interface, more abstract sentence features were maintained in prefrontal and temporoparietal circuits. To better understand how high-level semantic features constrain and facilitate our inner speech, future work should examine the interactions between sensorimotor circuits and the frontal and temporoparietal cortex as we silently rehearse sequences of words.

1. J. R. Binder, R. H. Desai, W. W. Graves, L. L. Conant, Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
2. G. Hickok, D. Poeppel, The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
3. C. Pallier, A. D. Devauchelle, S. Dehaene, Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2522–2527 (2011).
4. K. Patterson, P. J. Nestor, T. T. Rogers, Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* **8**, 976–987 (2007).
5. X. Pei *et al.*, Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *Neuroimage* **54**, 2960–2972 (2011).
6. D. Fegen, B. R. Buchsbaum, M. D'Esposito, The effect of rehearsal rate and memory load on verbal working memory. *Neuroimage* **105**, 120–131 (2015).
7. J. A. Lewis-Peacock, B. R. Postle, Temporary activation of long-term memory supports working memory. *J. Neurosci.* **28**, 8765–8771 (2008).
8. G. Mongillo, O. Barak, M. Tsodyks, Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008).
9. M. G. Stokes, 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405 (2015).
10. B. R. Buchsbaum, S. Lemire-Rodger, C. Fang, H. Abdi, The neural basis of vivid memory is patterned on perception. *J. Cogn. Neurosci.* **24**, 1867–1883 (2012).
11. A. D'Ausilio *et al.*, The motor somatotopy of speech perception. *Curr. Biol.* **19**, 381–385 (2009).
12. G. Rizzolatti, L. Craighero, The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192 (2004).
13. F. Pulvermüller *et al.*, Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7865–7870 (2006).
14. C. Cheung, L. S. Hamiton, K. Johnson, E. F. Chang, The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).
15. J. S. Arsenault, B. R. Buchsbaum, No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. *Psychon. Bull. Rev.* **23**, 1231–1240 (2016).
16. G. B. Cogan *et al.*, Sensory-motor transformations for speech occur bilaterally. *Nature* **507**, 94–98 (2014).
17. L. Lombardi, M. C. Potter, The regeneration of syntax in short term memory. *J. Mem. Lang.* **31**, 713–733 (1992).
18. M. C. Potter, L. Lombardi, Regeneration in the short-term recall of sentences. *J. Mem. Lang.* **29**, 633–654 (1990).
19. E. Race, D. J. Palombo, M. Cadden, K. Burke, M. Verfaellie, Memory integration in amnesia: Prior knowledge supports verbal short-term memory. *Neuropsychologia* **70**, 272–280 (2015).
20. C. E. Bonhage, C. J. Fiebach, J. Bahlmann, J. L. Mueller, Brain signature of working memory for sentence structure: Enriched encoding and facilitated maintenance. *J. Cogn. Neurosci.* **26**, 1654–1671 (2014).
21. O. Glanz Iljina *et al.*, Real-life speech production and perception have a shared premotor-cortical substrate. *Sci. Rep.* **8**, 8898 (2018).
22. J. P. Dmochowski, P. Sajda, J. Dias, L. C. Parra, Correlated components of ongoing EEG point to emotionally laden attention–A possible marker of engagement? *Front. Hum. Neurosci.* **6**, 112 (2012).
23. J.-R. King, S. Dehaene, Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).
24. R. Quentin *et al.*, Differential brain mechanisms of selection and maintenance of information during working memory. *J. Neurosci.* **39**, 3728–3740 (2019).
25. S. Michelmann, H. Bowman, S. Hanslmayr, The temporal signature of memories: Identification of a general mechanism for dynamic memory replay in humans. *PLoS Biol.* **14**, e1002528 (2016).
26. J. Gehrig *et al.*, Low frequency oscillations code speech during verbal working memory. *J. Neurosci.* **39**, 6498–6512 (2019).
27. J. R. Manning, J. Jacobs, I. Fried, M. J. Kahana, Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* **29**, 13613–13620 (2009).
28. K. J. Miller *et al.*, Dynamic modulation of local population activity by rhythm phase in human occipital cortex during a visual search task. *Front. Hum. Neurosci.* **4**, 197 (2010).
29. S. Ray, J. H. R. Maunsell, Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
30. C. J. Honey *et al.*, Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* **76**, 423–434 (2012).
31. N. E. Crone, D. Boatman, B. Gordon, L. Hao, Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clin. Neurophysiol.* **112**, 565–582 (2001).
32. A. Flinker *et al.*, Redefining the role of Broca's area in speech. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2871–2875 (2015).
33. E. Fedorenko *et al.*, Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci.* **113**, E6256–E6262 (2016).
34. C. Humphries, J. R. Binder, D. A. Medler, E. Liebenthal, Time course of semantic processes during sentence comprehension: An fMRI study. *Neuroimage* **36**, 924–932 (2007).
35. Y. Lerner, C. J. Honey, L. J. Silbert, U. Hasson, Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* **31**, 2906–2915 (2011).
36. K. J. Miller, B. L. Foster, C. J. Honey, Does rhythmic entrainment represent a generalized mechanism for organizing computation in the brain? *Front. Comput. Neurosci.* **6**, 85 (2012).
37. A. Flinker *et al.*, Single-trial speech suppression of auditory cortex activity in humans. *J. Neurosci.* **30**, 16643–16650 (2010).
38. T. L. Scott, T. K. Perrachione, Common cortical architectures for phonological working memory identified in individual brains. *Neuroimage* **202**, 116096 (2019).
39. V. L. Towle *et al.*, ECoG gamma activity during a language task: Differentiating expressive and receptive speech areas. *Brain* **131**, 2013–2027 (2008).
40. E. Edwards *et al.*, Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J. Neurophysiol.* **102**, 377–386 (2009).
41. G. B. Cogan *et al.*, Manipulating stored phonological input during verbal working memory. *Nat. Neurosci.* **20**, 279–286 (2017).
42. A. M. Liberman, F. S. Cooper, D. P. Shankweiler, M. Studdert-Kennedy, Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
43. B. R. Buchsbaum, M. D'Esposito, The search for the phonological store: From loop to convolution. *J. Cogn. Neurosci.* **20**, 762–778 (2008).
44. C. Jacquemot, S. K. Scott, What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* **10**, 480–486 (2006).
45. J. P. Rauschecker, An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* **271**, 16–25 (2011).
46. R. I. Schubotz, D. Y. von Cramon, Functional-anatomical concepts of human premotor cortex: Evidence from fMRI and PET studies. *Neuroimage* **20** (suppl. 1), S120–S131 (2003).

47. Y. Du, B. R. Buchsbaum, C. L. Grady, C. Alain, Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7126–7131 (2014).

48. J. D. Murray *et al.*, A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).

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

50. N. Savill *et al.*, Keeping it together: Semantic coherence stabilizes phonological sequences in short-term memory. *Mem. Cognit.* **46**, 426–437 (2018).

51. D. J. Heeger, W. E. Mackey, Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22783–22794 (2019).

52. S. M. Polyn, K. A. Norman, M. J. Kahana, A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* **116**, 129–156 (2009).

53. S. E. Cavanagh, J. P. Towers, J. D. Wallis, L. T. Hunt, S. W. Kennerley, Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat. Commun.* **9**, 3498 (2018).

54. D. F. Wasmuht, E. Spaak, T. J. Buschman, E. K. Miller, M. G. Stokes, Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nat. Commun.* **9**, 3499 (2018).

55. K. Polišenská, S. Chiat, P. Roy, Sentence repetition: What does the task measure? *Int. J. Lang. Commun. Disord.* **50**, 106–118 (2015).

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

57. T. Shallice, G. Vallar, "The impairment of auditory-verbal short-term storage" in *Neuropsychological Impairments of Short-Term Memory*, G. Vallar, T. Shallice, Eds. (Cambridge University Press, 1990), pp. 11–53.

NEUROSCIENCE

PSYCHOLOGICAL AND COGNITIVE SCIENCES