Coupled neural systems underlie the production and comprehension of naturalistic narrative speech

Lauren J. Silbert[a], Christopher J. Honey[a,b], Erez Simony[a], David Poeppel[c], and Uri Hasson[a,1]

[a]Department of Psychology and the Neuroscience Institute, Princeton University, Princeton, NJ 08540; [b]Department of Psychology, University of Toronto, Toronto, ON, Canada M5S 3G3; and [c]Department of Psychology, New York University, New York, NY 10012

Neuroimaging studies of language have typically focused on either production or comprehension of single speech utterances such as syllables, words, or sentences. In this study we used a new approach to functional MRI acquisition and analysis to characterize the neural responses during production and comprehension of complex real-life speech. First, using a time-warp based intrasubject correlation method, we identified all areas that are reliably activated in the brains of speakers telling a 15-min-long narrative. Next, we identified areas that are reliably activated in the brains of listeners as they comprehended that same narrative. This allowed us to identify networks of brain regions specific to production and comprehension, as well as those that are shared between the two processes. The results indicate that production of a real-life narrative is not localized to the left hemisphere but recruits an extensive bilateral network, which overlaps extensively with the comprehension system. Moreover, by directly comparing the neural activity time courses during production and comprehension of the same narrative we were able to identify not only the spatial overlap of activity but also areas in which the neural activity is coupled across the speaker's and listener's brains during production and comprehension of the same narrative. We demonstrate widespread bilateral coupling between production- and comprehension-related processing within both linguistic and nonlinguistic areas, exposing the surprising extent of shared processes across the two systems.

speech production | speech comprehension | intersubject correlation | brain-to-brain coupling

Successful verbal communication requires the finely orchestrated interaction between production-based processes in the speaker's brain and comprehension-based processes in the listener's brain. The extent of brain areas involved in the production of real-world speech in a speaker's brain during naturalistic communication is largely unknown. As a result, the degree of overlap between the production and comprehension systems, and the ways in which they interact, remain controversial. This study pursues three aims: (*i*) to map all areas (including but not limited to sensory, motoric, linguistic, and extralinguistic) that are reliably activated during the production of complex, real-world narrative; (*ii*) to map the overlap between areas that respond reliably during the production and the comprehension of real-world narrative; and (*iii*) to assess the coupling between activity in the speaker's brain during naturalistic production and activity in the listener's brain during comprehension of the same narrative. We discuss each challenge in turn.

The functional-anatomic architecture underlying the production of speech in an ecological context is incompletely characterized. Studies investigating production-based brain activity have been mainly restricted to the production of single phonemes (1–5), words (6–8), or short phrases in decontextualized, isolated environments (9–13) (see refs. 14 and 15 for exceptions). These studies report of a set of lateralized brain regions in the left frontal and left temporal–parietal hemisphere that are activated during speech production. These results are in contrast to the extensive bilateral set of brain areas reported to be activated during speech comprehension (16–18). Moreover, during comprehension, long segments of real-life speech activate a set of

extralinguistic midline areas, such as the precuneus and medial prefrontal areas (18, 19). It is not known whether production of real-life speech will also recruit these extralinguistic areas. Thus, in this study we asked whether a similarly extensive and bilateral set of brain areas will be involved in the production of real-life complex narratives, contrary to the prevailing models, which argue for a rather lateralized, dorsal-stream production system (17, 20–22).

Mapping the cortical attributes of the production system during natural speech is further complicated by methodological challenges related to the motor variability across repeated speech acts (retellings) (23). Unlike comprehension, where the same story can be presented repeatedly in an identical manner across listeners, in real-world contexts a speaker will never produce exactly the same utterances without subtle differences in speech rate, intonation, word choice, and grammar. In addition, owing to both the spatiotemporal complexity of natural language and an insufficient understanding of language-related neural processes, it is challenging to use conventional hypothesis-driven functional MRI (fMRI) analysis methods for modeling the brain activity acquired during long segments of natural speech. These challenges have mired our ability to fully characterize the production system. Here, to map motor, linguistic, and extralinguistic areas involved in real-world speech production, we trained speakers [one amateur storyteller (L.J.S.) as well as two professional actors] to precisely reproduce a real-life 15-min-duration narrative in the fMRI scanner. This unique design allowed us to map the reliable responses within the production system as a whole during the production of real-life rehearsed speech and to compare those to the responses recorded during spontaneous speech. Further, to correct for the variability in the motor output across retellings,

Significance

Successful verbal communication requires the finely orchestrated interaction between production-based processes in the speaker's brain and comprehension-based processes in the listener's brain. Here we first develop a time-warping tool that enables us to map all brain areas reliably activated during the production of real-world speech. The results indicate that speech production is not localized to the left hemisphere but recruits an extensive bilateral network of linguistic and extralinguistic brain areas. We then directly compare the neural responses during speech production and comprehension and find that the two systems respond in similar ways. Our results argue that a shared neural mechanism supporting both production and comprehension facilitates communication and underline the importance of studying comprehension and production within unified frameworks.

we applied a dynamic time-warping method to the fMRI data that allowed us to assess the response reliability within and across speakers during the production of real-life speech.

Uncertainty about the extent of the production system has hindered attempts at mapping the full spatial overlap between the production and comprehension systems. This is further complicated by the fact that few neurolinguistic studies measure production and comprehension of speech using the same speech materials (24, 25). Seminal studies have attempted to map the extent of overlap between brain areas dedicated to the production and comprehension of speech (12, 26–32) and have identified intriguing spatial overlaps between the production and comprehension systems in left inferior frontal gyrus (IFG), left medial temporal gyrus (MTG), left superior temporal sulcus (STG), and left Sylvian fissure at the parietal–temporal boundary (SPt). These studies by and large implicate left hemisphere structures but provide an incomplete map of the overlap between the two systems, because it is unknown whether such overlap changes in the context of complex communication.

Finally, mapping the extent of spatial overlap between comprehension and production of real-life speech is necessary but not sufficient for understanding the relationship between the two processes. An area involved in both speech production and speech comprehension can nonetheless perform different functions across the two tasks. In contrast, if similar functions are being recruited during the production and comprehension of speech, then the brain responses over time will be correlated (coupled) across the speaker's and the listener's brains. A better characterization of the temporal dependencies between production-based and comprehension-based neural processes requires a paradigm that allows for the direct comparison of the neural response time courses across both functions. Recently, we reported neural coupling across a speaker and a group of listeners during natural communication (33). However, that study did not map the production system. To assess whether the neural coupling across interlocutors is extensive or confined to a small portion of the production system it is necessary to map the full extent of brain regions that participate in the production of real-life speech. The comprehensive mapping of the production system in this study allowed us to assess the extent of speaker–listener neural coupling and to situate this coupling within the context of the larger human communication system.

Results

Production of Real-Life Story. To map all areas involved in the production of natural speech, we asked a speaker to memorize and pre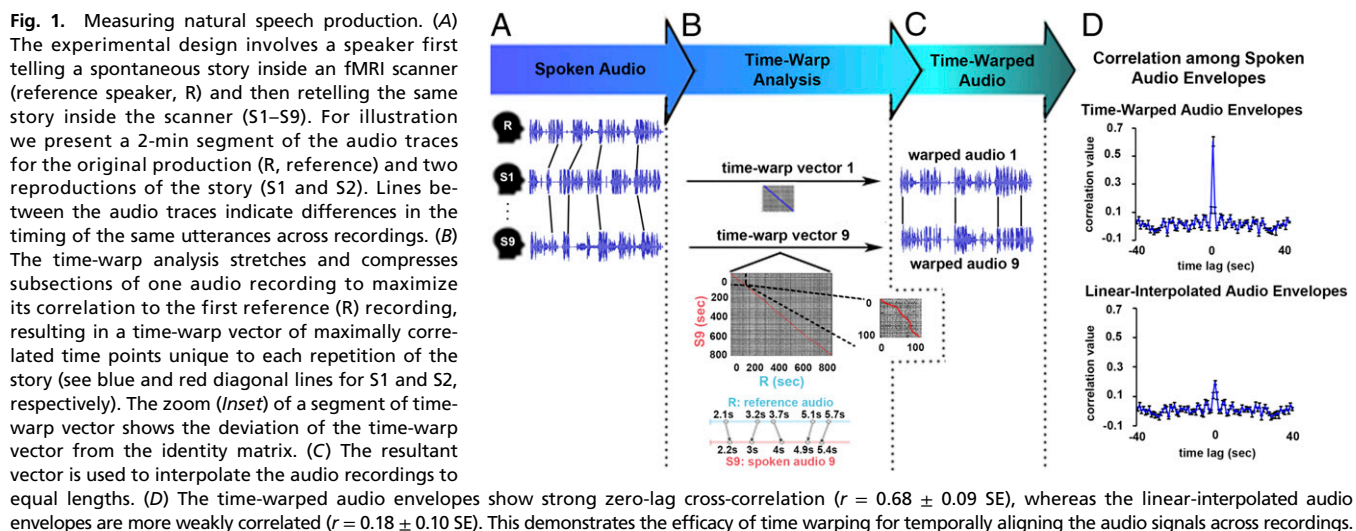cisely reproduce a story she spontaneously produced in a prior study (33). The speaker was instructed to use the exact same utterances during each repetition, as well as to try to preserve the same intent to communicate while maintaining the intonations, pauses, and rate of speech as in the original unrehearsed telling. To assess the speaker's ability to reproduce the story, we cross-correlated the audio envelopes of each recording with the original (first) recording. After the speaker learned to reproduce the story in a precise manner, she was brought back to the scanner to retell the story multiple times ($n = 12$).

Temporal Variability in Speech Production. Although the speaker managed to retell the story using similar utterances, we observed small differences in the precision of the production timing across repetitions (Fig. 1*A*). First, each recording was slightly longer or shorter in total length relative to the original recording (~15 min ± 15 s). Second, the variability in speech rate was not evenly distributed within each recording of the story (see lines in Fig. 1*A*). Such inherent variability in the production timing of real-life utterances is a major hurdle for the mapping of the production system during natural speech.

Time-Warp Analysis. To correct for the variability in the motor output, we adapted a dynamic time-warping technique used previously in the context of speech perception (34, 35) to our fMRI data. An audio recording of each spoken story provided us with a precise and objective measure of the speaker's behavior during each retelling. The time-warping technique (Fig. 1*B*) matches the audio envelope of each spoken story to the audio envelope of the original recording (reference audio, R). In contrast to a linear interpolation technique that interpolates at constant rate throughout a dataset, the time-warping technique dynamically stretches or compresses different components of the audio envelope to maximize its correlation to the audio envelope of the original story production (Fig. 1*B*). Examples of time-warped audios are shown in Fig. 1*C* (audio examples are also available upon request).

The time-warp procedure increases the correlation of the speech envelope across retellings, and more so than linear interpolation. Fig. 1*D* presents the autocorrelation plots between the audio envelopes of different speech recordings when using either linear interpolation or the time-warping procedure. The zero-lag peak correlation, attesting to the temporal alignment across recordings, is increased after applying the time-warping procedure compared with the linear interpolation ($r = 0.18 \pm 0.1$ vs. $r = 0.68 \pm 0.09$).

The level of correlation between the time-warped audio and the original recording provided us with an objective measure of the speaker's ability to reproduce the story in a reliable manner. The time-warped audio envelope was significantly correlated with

Fig. 1. Measuring natural speech production. (*A*) The experimental design involves a speaker first telling a spontaneous story inside an fMRI scanner (reference speaker, R) and then retelling the same story inside the scanner (S1–S9). For illustration we present a 2-min segment of the audio traces for the original production (R, reference) and two reproductions of the story (S1 and S2). Lines between the audio traces indicate differences in the timing of the same utterances across recordings. (*B*) The time-warp analysis stretches and compresses subsections of one audio recording to maximize its correlation to the first reference (R) recording, resulting in a time-warp vector of maximally correlated time points unique to each repetition of the story (see blue and red diagonal lines for S1 and S2, respectively). The zoom (*Inset*) of a segment of time-warp vector shows the deviation of the time-warp vector from the identity matrix. (*C*) The resultant vector is used to interpolate the audio recordings to equal lengths. (*D*) The time-warped audio envelopes show strong zero-lag cross-correlation ($r = 0.68 \pm 0.09$ SE), whereas the linear-interpolated audio envelopes are more weakly correlated ($r = 0.18 \pm 0.10$ SE). This demonstrates the efficacy of time warping for temporally aligning the audio signals across recordings.

the original recording in 9 out of the 12 recordings by the original speaker ($r = 0.52 \pm 0.05$), 10 out of 11 recordings by the first actor ($r = 0.29 \pm 0.03$), and 6 out of 10 recordings by the second actor ($r = 0.19 \pm 0.04$). Failing to time-warp a given recording to the original recording indicates that the speaker failed to reproduce the story in a precise manner during a retelling run. Of the 33 datasets originally recorded from all three speakers, we removed 8 outliers, leaving a total of 25 speech production datasets (*Methods*).

Time Warping of the Blood-Oxygen-Level-Dependent Signal. The acoustic time-warping vector for each recording (e.g., blue and red diagonal lines, Fig. 1*B*) was then used to dynamically align the fMRI signals to the same time base as the original reference audio (Fig. 2 *A–C*). The same time warping is applied to all voxels within a given retelling run. Only in cases where the brain responses are time-locked to the speech utterances applying the audio time warping to the blood-oxygen-level-dependent (BOLD) responses will improve the correlation of the BOLD responses across runs. However, given that the time-warping structure is determined based on the similarities among the sound wave forms, it cannot inflate (by overfitting) the neural correlation across repetitions. If two brain signals are uncorrelated with the speech production process, then using an independent time-warping alignment will not increase their similarity. Moreover, the time-warping procedure can only correct for differences in the production timing across recordings but cannot correct for other differences, such as differences in intonation or communicative intent. The presence of these additional behavioral differences can only reduce the reliability of neural responses across retellings, and thus reduce our sensitivity to detect reliable responses across retellings.

Intrasubject Correlation Analysis. Having mapped neural activity from each retelling onto a common time base, we next sought to measure the reliability of the neural time courses in the speaker's brain across retellings. To do so, we implemented an intrasubject correlation (intra-SC) analysis (*Methods*). For illustration purposes Fig. 2*D* presents the cross-correlation between the BOLD response time courses in the motor cortex across different speech recordings, both for linear interpolation (Fig. 2*D*, *Upper*) and for the time-warping procedure (Fig. 2*D*, *Lower*). A clear zero-lag peak correlation is observed after the time-warping procedure, with a clear advantaged for the time-warped signals, attesting to the temporal alignment of neural activity across retellings of the story. To map all areas that were reliably involved in speech production we computed the intra-SC analysis across the entire brain. Statistical significance of the intra-SC analysis was assessed using a nonparametric permutation procedure. All maps were corrected for multiple comparisons by controlling the false discovery rate (FDR).

Shared Brain Responses During Real-World Speech Production. The intra-SC analysis identified an extensive network of brain areas essential for the production of speech in the context of telling a real-life story (see Table S1 for Talairach coordinates). These areas responded in a reliable manner within the speaker's brain across multiple productions of the same 15-min story (Fig. 3*A*). Reliable responses during speech production were seen in motor speech areas, including the left and right motor cortices, as well as both right and left premotor cortex. Reliable responses were also observed in the right and left insula, which are adjacent to the motor cortex and may be crucial for syllabification (36), and in the basal ganglia, which is critically involved in motor coordination (37). Further reliable responses were seen bilaterally in the IFG. Specifically, this included the left pars triangularis and pars orbitalis, whose activity is associated with lexical access (38) and the construction of grammatical structures (39), among other functions (40, 41) (*Discussion*), and the right posterior IFS and pars orbitalis.

Reliable responses during speech production also extended to the left and right STG, left and right temporal pole (TP), left and right MTG, and left and right temporoparietal junction (TPJ) and Sylvian fissure (SPt). These structures have been previously linked to speech comprehension; here they are bilaterally reliable during production (*Discussion*). Significant reliability was also observed in a collection of extralinguistic areas implicated in the processing of semantic and social aspects of the story (19), including the precuneus, dorsolateral prefrontal cortex, posterior cingulate, and medial prefrontal cortex (mPFC), all bilateral.

We quantified the bilaterality of brain responses seen during speech production using a lateralization index (42, 43) (Table S2 and *SI Methods*). The laterality index (LI) yields values between −1 and 1, with +1 being purely left and −1 being purely right reliable responses, and with values between 0.2 and −0.2 widely considered to be bilateral (42, 43). We found significant bilateral responses in the STG, TP, TPJ, angular gyrus (AG), motor cortex, premotor cortex, and precuneus. Within the bilaterality found, AG (−0.198 ± 0.04) and precuneus (−0.173 ± 0.039) were weakly lateralized to the right, and the STG (0.19 ± 0.038), TP (0.09 ± 0.025), TPJ (0.12 ± 0.021), motor cortex (0.046 ± 0.022), and premotor cortex (0.068 ± 0.033) were weakly left lateralized. The MTG showed reliable responses in both hemispheres but was weakly lateralized to the left (0.224 ± 0.034). Considered in its entirety, the IFG response reliability was left-lateralized (0.486 ± 0.056), but when the IFG was segmented into smaller known functional areas bilaterality was observed in its more dorsal posterior segments (0.2 ± 0.035). This overall weak selectivity is in contrast to the current model of left-lateralized activity during speech production (17) and suggests a greater role than previously assumed for the right hemisphere during complex, narrative production.
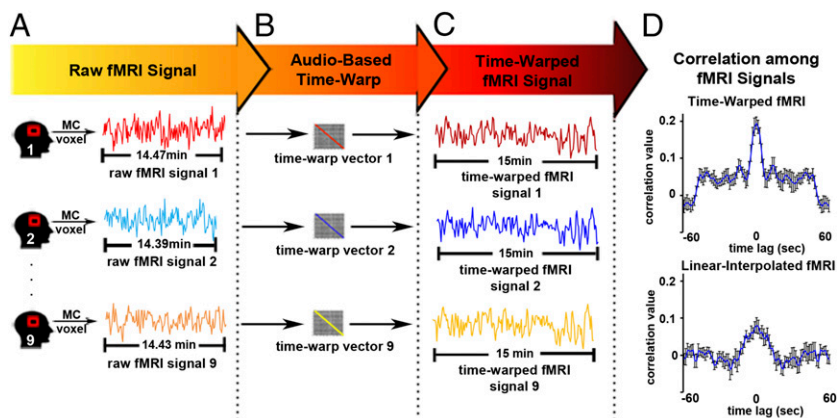

Fig. 2. Time warping the fMRI signal. (*A*) For illustration we present raw fMRI time courses from a given voxel in the motor cortex (MC) during speech production of the same story. (*B*) The time-warp vectors generated by time warping a given audio signal to the original (reference) recording (from Fig. 1*B*) are used to transform the fMRI signals measured during story retellings to a common time base. (*C*) Each fMRI response is interpolated individually according to the time-warping template of its corresponding audio envelope, thus improving the alignment between the brain responses and the envelope of the audio across repetitions of the story. (*D*) The time-warped fMRI signals show strong zero-lag cross-correlation ($r = 0.20 \pm 0.05$ SE), whereas the linear-interpolated audio envelopes are more weakly correlated ($r = 0.07 \pm 0.05$ SE). This demonstrates the efficacy of time warping for temporally aligning the fMRI time courses across recordings.
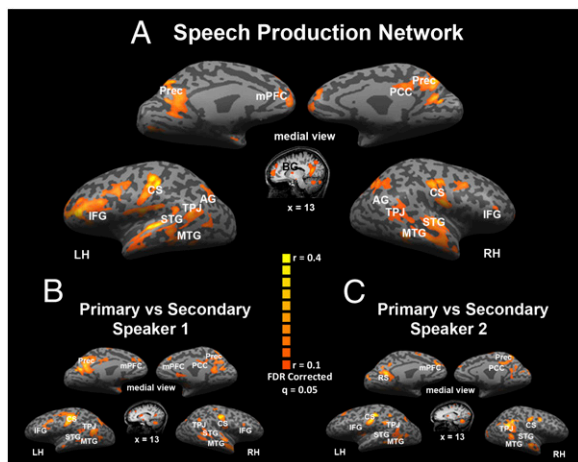
Fig. 3. (*A*) Areas that exhibit reliable neural responses across runs (*n* = 9) during which the first primary speaker produced a 15-min real-life story. The results are presented on lateral and medial views of inflated brains and one sagittal slice at Talairach coordinate *x* = 13. (*B*) Areas that exhibit reliable responses during the production of the same 15-min story between the primary speaker and secondary speaker 1. (*C*) Areas that exhibit reliable responses during the production of the same 15-min story between the primary speaker and secondary speaker 2. Anatomical abbreviations: AG, angular gyrus; BG, basal ganglia; CS, central sulcus; IFG, inferior frontal gyrus; mPFC, medial prefrontal cortex; PCC, posterior cingulate cortex; Prec, precuneus; STG, superior temporal gyrus; TPJ, temporal–parietal junction. See Table S1 for a complete list of areas that respond reliably during speech production.

Spontaneous vs. Rehearsed Speech. To ensure that our results can be generalized to spontaneous speech, we correlated the brain responses recorded during spontaneous speech with the brain responses recorded during the rehearsed speech using intra-SC (Fig. S1). Intra-SC revealed that the same extensive, bilateral network of brain areas recruited during rehearsed speech production is also shared between spontaneous and rehearsed speech production. Although our methods show extensive overlap between spontaneous and rehearsed speech, they do not allow us to assess differences between these two forms of speech production, because the spontaneous telling of a story is by definition a singular event that cannot be replicated. Nonetheless, the large extent of brain areas reliably activated during the production of spontaneous and rehearsed speech suggests the validity of using rehearsed speech to measure the entire network active during speech production in real-world contexts. Differences between spontaneous and rehearsed speech would manifest as an even more extensive network of brain areas recruited during complex speech production.

Brain Areas Are Aligned Across Speech Acts Only When Speakers Produce the Exact Same Speech Utterances. To test whether reliable responses during speech production are tied to the content of the story and not a function of the production of natural speech in general, we asked the speaker to tell another spontaneous, real-life story in the scanner. In this case, the same speaker is producing spontaneous and complex speech during both stories; however, the content of the speech varies across the two stories. We used intra-SC to compare the speaker's brain activity while telling the first story to her brain activity while telling the second story. We find no significant reliable responses between these two different speech productions. This suggests that the network reliably involved during speech production is tightly tied to the content of the produced speech.

Reliability of Brain Activity Across Speech Acts Is Tied to the Semantic and Grammatical Content of the Speech. To ensure that the reliability of brain responses seen across speech production acts is

not solely the result of low-level motor output, we asked the speaker to reproduce nonsense speech multiple times in the scanner. Specifically, the speaker uttered the phrase "goo da ga ba la la la la fee foo fa" for 5 min, eight different times. The phrase was uttered to the beat of a metronome to ensure maximal correlation across speech acts (*Methods*). We used intra-SC to compare the speaker's brain responses while telling each 5-min nonsense speech act. We find reliable brain responses in early auditory cortex and motor cortex but no significant reliability in brain areas that process upper-level aspects of speech production (Fig. S2*A*). This suggests that the extensive shared brain responses seen during narrative speech production is indeed tied to the semantic and grammatical content of the speech produced.

Reproduction of Real-Life Story by Secondary Speakers. To replicate our findings using additional speakers we trained two secondary speakers (SSs) to precisely reproduce the original real-life story told by the primary speaker. Once the secondary speakers learned to repeat the story with adequate precision (*Methods*), each was brought into the fMRI scanner to retell the story multiple times (*n* = 11 for SS 1, *n* = 10 for SS 2). Fig. S3 presents the quantified behavioral analysis of the precision with which each SS performed the story inside the fMRI scanner. The time-warped audio envelope of the SS reproductions was significantly correlated with the primary speaker's original recording in 10 out of 11 recordings by the first SS and 6 out of 10 recordings by the second SS. The successful production runs were entered into subsequent neural analysis (*Methods*).

Shared Brain Responses Between Multiple Speakers During Real-World Speech. To measure the reliability of the response time courses between the primary speaker and each secondary speaker we used a variation of the previously described intra-subject correlation analysis, an intersubject correlation (inter-SC) analysis (44, 45). For each brain region, the inter-SC analysis now measures the correlation of the response time courses across different speakers producing the same story. As with the intra-SC map, significance was assessed using a nonparametric permutation procedure and the map was corrected for multiple comparisons using an FDR procedure. The inter-SC between the primary speaker and each of the secondary speakers was similar to the production map obtained within the primary speaker (Fig. 3 *B* and *C*). We observed reliable responses in motor cortex, TPJ, MTG, TP, precuneus, left medial prefrontal gyrus, and dorsolateral prefrontal cortex. Although areas showing reliable activation during speaking were consistent among all speakers (especially in the motor cortex along the central gyrus, the temporal partial junction, middle temporal gyrus, temporal pole, precuneus, left medial prefrontal gyrus, and the dorsolateral prefrontal cortex), the strength of the results in the two secondary speakers was weaker than in the primary speaker, possibly related to the memorization of a story as opposed to the recollection of an episodic experience.

Overlapping and Distinct Elements of Production and Comprehension Networks: Shared Brain Responses During Speech Comprehension. To measure the overlap between the production and comprehension systems, we first measured the shared brain responses during speech comprehension. The comprehension map was based on the 11 subjects who listened, for the first time, to a recording of the story in a separate study (33), and the map was corrected for multiple comparisons using FDR. Successful comprehension was also monitored using a postscan questionnaire (*Methods* and ref. 33).

Reliability of Brain Activity During Comprehension Is Tied to the Content of the Story. To ensure that the reliable brain activity seen during speech comprehension is not the result of low-level acoustic input we asked a listener to listen to the nonsense speech produced by the speaker (discussed above) multiple times in the scanner (*n* = 8). In this case, the listener is listening to the same auditory input but cannot form any interpretation of meaning

over time. We used intra-SC to compare the listener's brain responses while listening to each 5-min nonsense speech act. We found reliable brain responses in early auditory cortex but no significant reliability in brain areas that process upper-level aspects of language comprehension (Fig. S2B). This suggests that the extensive shared brain responses seen during narrative speech comprehension are tied to the content of the story. This finding also reinforces previous findings that demonstrate that shared meaningless auditory input (e.g., reversed speech) does not result in the extensive reliable activity shared across listeners during the comprehension of meaningful auditory input (18). Moreover, the opposite is also true: Shared content without shared form has been shown to evoke reliable responses in high-order brain areas but not in low-order brain areas, such as correlations found between Russian listeners who listened to a Russian story and English listeners who listened to a translation of the story (46). As a result, we are confident that the widespread areas that responded reliably during speech comprehension and production are tied to the processing of linguistic and extralinguistic content and not to the processing of low-level audio features.

Spatial Overlap Between Speech Production and Speech Comprehension.
A spatial comparison of the speech production network and the speech comprehension network revealed substantial overlap (Fig. S4, orange) as well as sets of areas that were specific to production (red) and comprehension (yellow) processes. Some aspects of speech comprehension and production may be unique to each process, and indeed there were brain areas selective for only one or the other function. Areas involved in speech production included bilateral motor cortex, bilateral premotor cortex, left anterior dorsal section of the IFG, and a subset of bilateral areas along the temporal lobe (see Table S1 for a complete list). Areas involved in speech comprehension (but not production) included bilateral parietal lobule, the right pars orbitalis), and a set of areas bilaterally along the temporal lobe. However, many of the brain areas involved in speech production were also reliably activated during speech comprehension. The overlapping areas included bilateral TPJ and subsets of regions along the STG and MTG, as well as the precuneus, posterior cingulate, and mPFC. A t test ($\alpha = 0.05$) between the production-related and comprehension-related reliability values indicated that none of the overlapping voxels exhibited significantly greater reliability during the production or comprehension of the narrative. The overlap suggests, plausibly, that many production-related and comprehension-related computations are performed within the same brain areas.

Widespread Coupling of Neural Activity in Speaker and Listener During Real-World Communication. Finally, taking advantage of our measurements of neural responses during production and comprehension of the same story, we directly compared the neural time courses elicited during production and comprehension. To measure the coupling between production and comprehension mechanisms, we formulated, in a previous publication, a model of the expected responses in the listener's brain during speech comprehension based on the speaker's responses during speech production (*Methods* and ref. 33). The coupling model allows us to test the hypothesis that the speaker's brain responses during production are spatially and temporally coupled with the brain responses measured across listeners during comprehension. During communication we expect significant production–comprehension couplings to occur if the neural responses during the production of speech in the speakers are similar to the neural responses in the listener's brain during the comprehension of the same speech utterances (47–49).

Significant coupling between the speaker's and listeners' brain responses was found in "comprehension-related" areas along the left anterior and posterior MTG, bilateral TP, bilateral STG, bilateral AG, and bilateral TPJ; "production-related" areas in the dorsal posterior section of the left IFG, the bilateral insula, the left premotor cortex, and the supplementary motor cortex; and a collection of extra linguistic areas known to be involved in
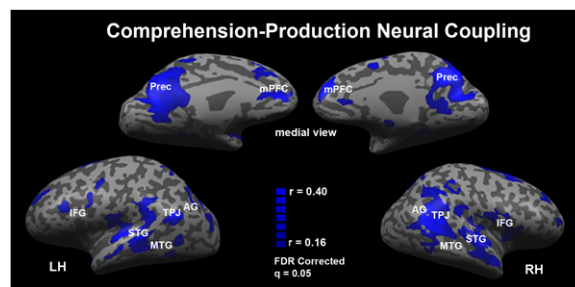

Fig. 4. Areas in which the responses during speech production are coupled to the responses during speech comprehension. The comprehension–production coupling includes bilateral temporal cortices and linguistic and extralinguistic brain areas (see Table S1 for a complete list). Significance was assessed using a nonparametric permutation procedure and the map was corrected for multiple comparisons using an FDR procedure.

narrative comprehension, including the precuneus and mPFC (Fig. 4). Thus, these results replicate our previous findings using a new dataset (33). Moreover, the results indicate that the extent of coupling between the speaker and listener is greater than previously reported. This is probably due to the increase in signal-to-noise ratio when averaging BOLD signal across multiple speech production runs. In addition, the extensive coupling further relieves the methodological challenge of using rehearsed speech to model spontaneous speech, because the rehearsed speech production was more extensively coupled to the listeners, all of whom listened only once and for the first time. Most importantly, this study places the idea of a speaker-listener coupling within the context of the human communication system as a whole by assessing the extent of coupling relative to the extent of segregation between the production and comprehension systems. Our results suggest that only a subset of the human communication system is dedicated to either the production or the comprehension of speech, whereas the majority of brain areas exhibit responses that are shared, hence similar, across the speakers and the listeners (Fig. 5).

Discussion
In this study we mapped the network of brain areas involved in complex, real-world speech production. The production of real-world speech recruited a network of bilateral brain areas within the motor system, the language system, as well as extralinguistic areas (Fig. 3). The bilateral symmetry observed in the production network challenges the suggestion that the dorsal linguistic stream, which is associated with the production system, is strongly lateralized to the left hemisphere (17, 50). The lack of lateralized responses is in agreement, however, with several recent publications that report bilaterality in speech production (12, 20, 50–52). A major difference between this study and all others investigating language processing is the production task. Here the speakers produced a long, unconstrained, real-life narrative, which is likely to recruit a larger network of brain areas relative to prior studies that have focused mainly on the production of short and unrelated utterances (12, 23). In addition, analytical methods differ greatly between this study and previous ones: Here we measure response reliability, whereas the majority of previous studies measure signal amplitude using event-related averaging methods (12, 23). As we demonstrated before, intersubject and intrasubject correlation can uncover reliable responses that cannot be detected using standard event-related averaging methods (53). Our data-driven approach seems to increase our sensitivity to detect production-related responses in the speaker's brain (especially in the right hemisphere) that were not reported in prior studies.

A growing number of right hemisphere homologs have been reported for left hemispheric areas involved in speech comprehension and speech production; these reports are in apparent conflict with lesion data that persistently indicate lateralization to the left hemisphere. Resolving this puzzle is beyond the scope of this
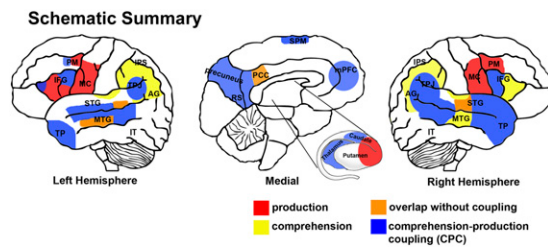
Fig. 5. Schematic summary of the networks of brain areas active during real-life communication. Areas that exhibited reliable time courses only during the production of speech are marked in red and include the right and left motor cortex, right premotor cortex, left anterior section of the IFG, right anterior inferior temporal (IT), and the caudate nucleus of the striatum. Areas that exhibited reliable time courses only during the comprehension of speech are marked in yellow, and include the right and left IPS, the left and right posterior STG, and the right anterior IFG. Areas that exhibited reliable time courses during both the production and comprehension of speech (overlapping areas) but in which the response time courses during the production and comprehension of speech did not correlate are marked in orange. These areas include sections of the left and right MTG, sections of the left and right IPS, and the PCC. Areas in which the response time courses during the production and comprehension of speech are coupled are marked in blue. These areas include comprehension related areas along the left and right anterior and posterior STG, left anterior and posterior MTG, left and right TP, left and right AG, and bilateral TPJ; production-related areas in the dorsal posterior section of the left IFG, the left and right insula, and the left premotor cortex; and a collection of extra linguistic areas in the precuneus and medial prefrontal cortices.

study; our study was not designed to discern the nature of the right hemisphere contribution. Moreover, against the background of predominantly bilateral production-related neural activity we did note some laterality in the production system. For example, the anterior portion of the left IFG exhibited stronger production-related reliability than in the right hemisphere. In addition, we observed slightly more production and comprehension responses in the left MTG than in the right (Table S1). The left MTG seems to be crucial in the retrieval of lexical items (discussed below), a process that may be necessary for both the production and comprehension of words.

The comprehensive mapping of the production network allowed us to assess the extent of overlap between brain areas involved in the production and comprehension of the same complex, real-world speech. Fig. 5 provides a schematic summary of the main results. We found areas of activity specific to either speech production (Fig. 5, red) or speech comprehension (Fig. 5, yellow), as well as regions that responded reliably during both the production and comprehension of real-world speech (Fig. 5, orange and blue). The areas of convergence included regions classically associated with the comprehension system (including anterior and posterior MTG and STG, as well as the TPJ and AG, all bilaterally) as well as regions classically associated with the production system (including the left dorsal IFG and the insula bilaterally). Furthermore, we observed extensive coupling in extralinguistic areas such as the precuneus and mPFC. Finally, a large subset of the overlapping areas did not only respond reliably during both speech production and comprehension but also exhibited temporally coupled (correlated) activity profiles across the two tasks (Fig. 5, blue).

The response time courses evoked during the production and comprehension of the same story were coupled in many areas previously known to be crucial for either comprehension or production of speech. The IFG is of particular interest in this respect. Whereas it was traditionally associated with speech production (54, 55), more recent work has identified finer parcellations in this region (40, 56) that associate it with functions including language comprehension. In terms of the role of this area in speech production, left BA 44–45 has been shown to be involved in the articulatory loop and to increase its response

amplitude as a function inner speech rate (57), verbal fluency (58), and generation of pseudowords relative to familiar words (59). In agreement with these studies the left IFG, including BA44 and BA45, exhibits highly reliable activation patterns during speech production in all three speakers. However, the left IFG was also found to be involved with semantic, syntactic, and phonological processes that are relevant to both the production and comprehension of speech (41). These processes include semantic retrieval (60), lexical decision (61), syntactic transformations (62, 63), and verbal working memory (64–69). Based on such reports it was argued that the left IFG is involved with both production and comprehension processes (21, 70, 71). Our study goes beyond these findings by revealing a production–comprehension coupling in the left dorsal BA44–45. Thus, the present data indicate that a portion of the left IFG is not only involved in production and comprehension of speech but also exhibits a common pattern of activity when people produce or comprehend the same natural speech patterns. Moreover, the insular cortex, a region involved with coordination of complex articulatory movements (36) and also implicated as a core lesion area for Broca's aphasia (72, 73), also exhibited strong production–comprehension coupling. The robust coupling in this area is especially noteworthy when compared against the more mixed pattern of results observed across subregions of the IFG.

The production–comprehension coupling observed here suggests that many aspects of language processing are shared across both functions, but it also argues against a strong version of the motor theory of speech perception (MTSP) (48). MTSP argues that the recruitment of the articulatory motor system is necessary for speech comprehension (47). In contrast, our results indicate that the articulatory-based motor areas along the left and right ventral precentral gyrus responded reliably only during speech production (Fig. S4, red), but not during speech comprehension. The finding that speech comprehension does not rely on the articulatory system is in agreement with the observation that young infants can comprehend speech before they are able to produce speech (74). However, our study did find production–comprehension coupling in the left premotor cortex and the left and right insula, which are adjacent to early motor areas, as well as in the dorsal and ventral linguistic streams (75). Thus, our findings argue for robust interconnected relationships between action (production)–perception (comprehension) circuits at the syntactic and semantic levels but not at the articulatory level (76).

Production–comprehension coupling was also found in areas traditionally thought to be involved in speech comprehension, such as the STG, MTG, and AG (17, 31, 72, 77–90). Although it is initially surprising to observe coupled responses in traditionally comprehension-based areas, these results are consistent with recent findings, such as the implication of the left posterior MTG in lexical access, which is needed for both production and comprehension of speech (91). Using diffusion imaging of white matter pathways, the left MTG has been indicated to be interconnected with many parietal and frontal language-related regions (92). These data suggest a central role for the left posterior MTG in both production and comprehension of speech, because the region seems to function as a focal point for both processes (13, 92). The bilateral STG has also been shown to be involved in speech production and speech comprehension, specifically in the external loop of self-monitoring, based on studies that distorted the subjects' feedback of their own voices or presented the subjects with alien feedback while they spoke (93, 94). More recently, the AG has been shown to be involved with the construction of meaning in its temporal parts (31, 50) and with sensorimotor speech transformation in its more parietal parts along the posterior Sylvian fissure (24). In agreement with the notion that these functions are integral to the production as well as the comprehension of speech, our data indicate they share similar response time courses during the processing of the same story.

We found the most extensive production–comprehension coupling in the mPFC and precuneus. These extralinguistic areas are argued to be necessary for a range of social functions that should indeed be shared across interlocutors, such as the ability to infer

the mental states of others (95–97). For example, recent functional imaging findings suggest a central role for the precuneus in a wide spectrum of highly integrated tasks, including episodic memory retrieval, first-person perspective taking, and the experience of agency (98, 99). Studies assessing the role of the mPFC assign functions ranging from reward-based learning and memory to empathy and theory of mind (96, 100, 101). The ability of a listener to relate to a speaker and therefore understand the content of a complex real-world narrative seems to rest in these higher-level processing centers. For example, stories that require a clear understanding of second-order belief reasoning (understanding how another person can hold a belief about someone else's mental state) are frequently used as localizers for areas involved in theory of mind [mPFC, precuneus, and PCC (102–104)]. Inferring someone's intentions plays a valuable role during the exchange of information across interlocutors and may therefore be integral to the success of real-world communication.

Not all areas recruited by both speech comprehension and speech production responded in a similar way (orange areas in Fig. 5 and Fig. S4). This suggests that a subset of overlapping brain areas perform different computations during the production and the comprehension of speech. Interestingly, parts of the mid superior temporal sulcus, which are adjacent to early auditory cortical areas, responded reliably during production and comprehension processes separately but were not coupled between the two processes. During comprehension, these areas may receive input from early auditory areas and serve as an entry point for the analysis of the speech sounds properties, whereas during production these areas might be needed to monitor the precision of the speech output. Furthermore, some auditory processing areas seem to be inhibited during speech production (105), which would decrease the coupling of activity between speakers and listeners in these regions (106).

To map the entire production system using our methodologies, a speaker had to repeat the same story multiple times, thus potentially introducing adaption effects. Adaptation is the reduction in neural activity when stimuli are repeated. It has been reported robustly and at multiple spatial scales from individual cortical neurons to hemodynamic changes using fMRI. If repeated storytelling were affected by adaptation, brain activity involved in speech production should decrease over time, as should the shared brain responses across reproductions of the same story. In other words, adaptation effects work against our ability to assess the brain activity involved in speech production. The extensive reliable responses observed during the reproduction of the story suggest, however, that adaptation effects do not mask our ability to map many of the production-related brain areas active during rehearsed speech.

Another potential caveat with our coupling model is the possibility that the speaker–listener coupling results from the speaker also playing the role of listener, because she is likely to listen to herself while speaking. We were able to address this caveat in a previous publication with the analysis of the temporal structure of the data (33). In our model of brain coupling the speaker–listener temporal coupling is reflected in the model's weights, where each weight multiplies a temporally shifted time course of the speaker's brain responses relative to the moment of vocalization (synchronized alignment, zero shift). We found the shared responses among the listeners to be time-locked to the moment of vocalization. In contrast, in most areas the responses in the listeners' brains lagged behind the responses in the speaker's brain by 1–3 s. These results allay the methodological concern that the speaker–listener neural coupling is induced simply by the fact that the speaker is listening to her own speech, because the dynamics between the speaker and listener significantly differ from the dynamics among listeners (33).

In this study we applied new methodological and analytical tools to map the production and comprehension systems during real-world communication and to measure functional coupling across systems. Our time-warp-based intraspeaker correlation analysis provides a comprehensive map of the human speech production

system in its entirety and introduces a new experimental tool for studying speech production under more ecologically valid conditions. Moreover, we have shown extensive coupling between the response time courses in the speaker's brain and the listener's brain during the production and comprehension of the same story. The robust production–comprehension coupling observed here underlines the importance of studying comprehension and production within a unified framework. Just as one cannot study the processes by which information is transmitted at the synaptic level by focusing solely on the presynaptic or postsynaptic compartments, one cannot fully characterize the communication system by focusing on the processes within the border of an isolated brain (107).

Methods

Subject Population. Three speakers and 11 listeners, ages 21–40, participated in one or more of the experiments. All participants were right-handed native English speakers. Procedures were in compliance with the safety guidelines for MRI research and approved by the University Committee on Activities Involving Human Subjects at Princeton University. All participants provided written informed consent.

Production Procedure. In a previous study (33), we used fMRI to record the brain activity of a speaker spontaneously telling a real-life story (15 min long). The speaker had three practice sessions inside the scanner telling real-life stories to familiarize her with the conditions of storytelling inside the scanner and to practice minimizing head movements during natural storytelling. In her fourth fMRI session, the speaker told a new, unrehearsed, real-life account of an experience she had as a freshman in high school. The speaker then told another, unrehearsed, real-life story of an experience she had while mountain climbing to be used as a control. Both stories were recorded using an MR-compatible microphone (discussed below). The speech recording was aligned with the transistor–transistor logic (TTL) signal produced by the scanner at the onset of the volume acquisition.

In the current study, the same speaker was asked to rehearse her first story, learning to reproduce the same narrative with the same utterances, keeping intonation and speech rate as similar as possible while maintaining the same intent to communicate. Communicative intent was maintained by instructing the speaker to produce each retelling of the story as if for a different friend or audience, much like one might share a story with multiple people in her daily life. In focusing on a personally relevant experience we strove both to approach the ecological setting of natural communication and to ensure the speaker's intention to communicate. After the speaker learned to precisely reproduce the story, she retold it in the scanner 12 more times.

Although the reliance on rehearsed speech is not fully natural, it serves as a compromise for natural speech by preserving the complexity of real-life narrative, and as such it provides a step forward from experimental protocols that rely on the production of single sentences or words in a decontextualized setup. Thus, in this study we were able to map all areas that respond reliably during the production of real-life rehearsed speech. This design, however, is not suitable for mapping areas that are uniquely activated during the production of spontaneous speech. The design does allow for the comparison of rehearsed and spontaneously produced speech, however, which can be used to measure the similarity in brain activity recruited during the two distinct processes.

Next, we recruited two secondary speakers (SSs), both accomplished actors, to repeat the experiment. Each SS was asked to first learn and then perform the real-life story told by the original speaker in a precise manner. Specifically, the SSs were instructed to tell the story as if it were their own. Once they could repeat the story with adequate precision (discussed below) they began retelling the story inside the fMRI scanner. Each SS had two practice storytelling sessions inside of the scanner to familiarize her with the scanner environment and learn to minimize head movement. They then each retold the story multiple times ($n = 11$ for SS 1, $n = 10$ for SS 2), each time with the intention to communicate the story to a different audience so as to maintain as natural an environment as possible.

Finally, we had the primary speaker perform a control experiment whereby she repeated the nonsense phrase "goo da ga ba la la la fee foo fa" repeatedly in 5-min segments and then repeated the entire 5-min nonsense production a total of eight times. The speaker repeated the nonsense syllabic phrase to the beat of a metronome, which allowed for temporal precision and maximal correlation across speech acts and eliminated the need for time warping. This procedure served as a control for low-level speech production as the speaker was reliably producing sound that contained no meaning.

Comprehension Procedure. Next, we measured the listeners' brain responses during audio playback of the original recorded story. We synchronized the

functional MRI signals to the speaker's vocalization using the scanner's TTL pulse, which precedes each volume acquisition. Eleven listeners listened to the recording of the story. Our experimental design thus captured both the production and comprehension sides of the simulated communication. Following the fMRI scan, each listener was asked to freely recall the content of the story. Six independent raters scored each of these listener records according to a 115-point true/false questionnaire, and the resulting score provided a quantitative measure of each listener's understanding.

We then measured a listener's brain responses during audio playback of nonsense speech. The listener listened to all eight 5-min productions of the nonsense phrase "goo da ga ba la la la fee foo fa" told by the original speaker above. This served as a control for the role that content plays in comprehension, because the listener was hearing the same vocalizations but had no ability to interpret meaning.

Recording System. We recorded the speaker's speech during the fMRI scan using a customized MR-compatible recording system (FOMRI II; Optoacoustics Ltd). The MR recording system uses two orthogonally oriented optical microphones. The reference microphone captures the background noise, whereas the source microphone captures both background noise and the speaker's speech utterances (signal). A dual-adaptive filter subtracts the reference input from the source channel (using a least mean square approach). To achieve an optimal subtraction, the reference signal is adaptively filtered where the filter gains are learned continuously from the residual signal and the reference input. To prevent divergence of the filter when speech is present, a voice activity detector is integrated into the algorithm. Finally, a speech enhancement spectral filtering algorithm further preprocesses the speech output to achieve a real-time speech enhancement.

MRI Acquisition. Subjects were scanned in a 3T head-only MRI scanner (Allegra; Siemens). A custom radio frequency coil was used for the structural scans (NM-011 transmit head coil; Nova Medical). For fMRI scans, a series of volumes was acquired using a T2*-weighted EPI pulse sequence [repetition time (TR) 1,500 ms; echo time (TE) 30 ms; flip angle 80°]. The volume included 25 slices of 3-mm thickness with a 1-mm interslice gap (in-plane resolution 3×3 mm^2). T1-weighted high-resolution ($1 \times 1 \times 1$mm) anatomical images were acquired for each observer with an MPRAGE pulse sequence to allow accurate cortical segmentation and 3D surface reconstruction. To minimize head movement, subjects' heads were stabilized with foam padding. Stimuli were presented using Psychophysics toolbox (108) in MATLAB (The MathWorks, Inc.). High-fidelity MRI-compatible headphones (MR Confon) were fitted to present the audio stimuli to the subjects while attenuating scanner noise.

Data Preprocessing. fMRI data were preprocessed with the BrainVoyager software package (Brain Innovation) and with additional software written with MATLAB Preprocessing of functional scans included linear trend removal and high-pass filtering (up to three cycles per experiment). All functional images were transformed into a shared Talairach coordinate system so that corresponding brain regions are roughly spatially aligned. To further overcome misregistration across subjects the data were spatially smoothed with a Gaussian filter of 8 mm full width at half maximum value. To remove transient nonspecific signal elevation effects at beginning of the experiment and some preprocessing artifacts at the edges of the time courses, we excluded the first 15 and the last 5 time points of the experiment from the story analysis. Further, voxels with a low mean BOLD signal (i.e., voxels outside the brain, defined as <400 activity units) were excluded from all subsequent analysis.

Correcting for Speech-Related Motion. To minimize head motion during speech, the speakers were first trained to speak with minimal head motion while lying on the scanner board by participating in two practice sessions each. To correct for any head motion, we used a 3D algorithm that adjusts for small head movements by rigid body transformations of all slices to the first reference volume. Detected head motions were less than 1mm, which is well within the range of typical movements observed in imaging studies.

Signal from individually defined regions of interest (ROIs) was regressed out of the data at each run, based on the assumption that signal from noise ROIs in white matter and scalp can be used to accurately model physiological fluctuations as well as residual motion artifacts. For each subject we defined three "noise ROIs" (nROIs): (*i*) white matter nROI, defined based on the MPRAGE of each subject; (*ii*) scalp nROI, defined in the interface between the scalp and the brain; and (*iii*) eyeball nROI, defined in the eye region. Signals in these out-of-brain regions are unlikely to be modulated by neural activity and thus primarily reflect physiological noise (109) and head movement due to speech. For each nROI in each subject we calculated the first 10 principal components, which explained more than 50% of the vari-

ance across all voxels within each response time course in a given nROI. We then used linear regression to remove these components from each voxel in each subject. This method of regressing noise-related signal was more conservative than the standard method in which the mean response (which explains only 6–8% of the variance) in each nROI is regressed out (109).

Time-Warp Analysis. To correct for temporal differences in the rate of speech production across story retellings, we implemented a time-warp analysis that aligns each retelling to a common time base (Figs. 1 and 2). An audio recording of each retelling provided us with a precise and objective measure of the speaker's behavior during that retelling. The time-warping technique first matches the audio recording of each retelling to the first spontaneous audio recording. In contrast to standard linear interpolation that interpolates at constant rate throughout a dataset, the time-warping technique dynamically stretches or compresses different components of the audio envelope to maximize its correlation to the audio envelope of the original story production. To compute the dynamic warping, we modified an algorithm to fit to fMRI data (34, 110, 111). This algorithm produces a mapping from time points in the audio of each retelling to time points in the audio of the original telling. This audio-based mapping vector was then used to dynamically align the fMRI signals of each retelling to the common time base of the original story production. Each fMRI response was interpolated individually according to its matching audio time-warping vector, and the same interpolation was applied to all voxels within a given retelling run.

To ensure that each retelling of the story was recorded with adequate precision, we correlated the audio envelope of the time-warped audio with the original recording. We identified outliers as those runs where the correlation between the time-warped audio recording and the original recording was unusually low. Specifically, outlier runs were defined as those runs whose audio correlation with the original, after time-warping, were more than twice the interquartile range from the median of the correlations. Thus, if $IQR = Q_3 - Q_1$, the outliers had correlation values below $Q_1 - 2(IQR)$ or above $Q_3 + 2(IQR)$. Such low correlation values between the envelopes of the original audio and reproduced audio indicates that the speakers failed to precisely reproduce the original story in these runs, and they were excluded from further analyses.

We note that the temporal alignment of the fMRI signals was performed entirely using information from the auditory recordings. Therefore, this procedure cannot artificially inflate (by overfitting) the correlation across the fMRI datasets. However, in cases where the brain responses are time-locked to the speech utterances, using the speech time warping on the BOLD signal can improve the correlation of the brain responses across recordings.

Intra-SC Analysis. To measure the reliability of the response time courses between corresponding regions during the production of speech, we used an intra-SC analysis. This analysis provides a measure of the reliability of the responses to complex speech production by comparing the BOLD response time courses across different storytelling repetitions (45). Correlation maps were constructed on a voxel-by-voxel basis (in Talairach space) among each storytelling repetition by comparing the responses across all storytelling repetitions in the following manner. For each voxel, we computed the Pearson product-moment correlation $r_j = \text{corr}(TC_j, \overline{TC}_{All-j})$ between that voxel's BOLD time course TC_j in one repetition and the average $\overline{TC}_{All-j}$ of that voxel's BOLD time courses in the remaining repetitions. The mean correlation $R = \frac{1}{N}\sum_{j=1}^{N} r_j$ was then calculated at every voxel, and this defined the intra-SC.

Intra-SC analysis was also used to measure both the reliability of response time courses between corresponding regions during the production of spontaneous vs. rehearsed speech and during the production of an additional unrelated story versus the original story.

Inter-SC Analysis. To measure the reliability of the response time courses both between the speaker and each SS and between the listeners' brain, we used an inter-SC analysis. The inter-SC analysis was computed in an analogous manner to the intra-SC analysis above, except that instead of correlating across repetitions of speech production within the same speaker we correlated across different speakers or across different listeners hearing the same story.

Phase-Randomized Bootstrapping. Because of the presence of long-range temporal autocorrelation in the BOLD signal (112), the statistical likelihood of each observed correlation was assessed using a bootstrapping procedure based on phase randomization. The null hypothesis was that the BOLD signal in each voxel in each individual was independent of the BOLD signal values in the corresponding voxel in any other individual at any point in time (i.e., that there was no inter-SC between any pair of subjects).

Phase randomization of each voxel time course was performed by applying a discrete Fourier transform to the signal then randomizing the phase of each

Fourier component and inverting the Fourier transformation. This procedure scrambles the phase of the BOLD time course but leaves its power spectrum intact. A distribution of 1,000 bootstrapped average correlations was calculated for each voxel in the same manner as the empirical correlation maps described above, with bootstrap correlation distributions calculated within subjects and then combined across subjects. The distributions of the bootstrapped correlations for each subject were approximately Gaussian, and thus the mean and SDs of the r_j distributions calculated for each subject under the null hypothesis were used to analytically estimate the distribution of the average correlation, R, under the null hypothesis. Finally, P values of the empirical average correlations (R values) were computed by comparison with the null distribution of R values.

FDR Correction. To correct for multiple comparisons, the Benjamini–Hochberg–Yekutieli false-discovery procedure that controls the FDR under assumptions of dependence was applied (113, 114). Following the procedure, P values were sorted in ascending order and the value p_{q*} was chosen. This value is the P value corresponding to the maximum k such that $p_k < (k/N)q*$, where $q* = 0.05$ is the FDR threshold and N is the total number of voxels included in the analysis.

Direct Statistical Testing Between Production and Comprehension Maps. To identify areas that show increase in response reliability for one condition over the other, a t test ($\alpha = 0.05$) was performed within each voxel that exceeded the threshold in both of the conditions (i.e., voxels that seems to participate in both the comprehension and production of the speech signal). Thus, the t test was performed by comparing the correlation values of the speaker's retellings during production $\{r_j, r_{j+1} \dots r_n\}$ to the correlation values of the listeners during comprehension $\{r_j, r_{j+1} \dots r_n\}$, within each voxel. The resulting map of t test result/voxel was corrected for multiple comparisons using the FDR analysis outlined above where $q* = 0.05$.

Neural Coupling Analysis. To measure the direct coupling between production and comprehension-based processing, we formed a spatially local general linear model in which temporally shifted voxel time series in one brain are linearly summed to predict the time series of the spatially corresponding voxel in another brain. Thus, the activity at one moment in time in the listener's brain is described as a function of past, present, and future activity in in the speaker's brain. The coupling model is written as

$$v_{\text{listener}}^{\text{model}}(t) = \sum_{\tau=-\tau_{\max}}^{\tau=\tau_{\max}} \beta_i v_{\text{speaker}}(t+\tau),$$

where the weights $\vec{\beta}$ are determined by minimizing the rms error and are given by $\vec{\beta} = (C)^{-1}\langle \vec{v} \ v_{\text{listener}}\rangle$. Here C is the covariance matrix $C_{mn} = \langle v_m v_n\rangle$ and $\vec{v}$ is the vector of shifted voxel times series, $v_{\text{m}} = v_{\text{speaker}}(t-m)$. Here we choose $\tau_{\max} = 4$, which is large enough to capture important temporal processes while also minimizing the overall number of model parameters to maintain statistical power. We obtain similar results with $\tau_{\max} = (3, 5)$ (for a complete assessment of the model's stability see ref. 33).

We identified statistically significant neural couplings by assigning P values through a Fisher F test. Specifically, the model equation above has $\delta_{\text{model}} = 9$ degrees of freedom whereas $\delta_{\text{null}} = T - \delta_{\text{model}} - 1$, where T is the number of time points in the experiment. For each model fit we construct the F statistic and associated P value $P = 1 - f(F, \delta_{\text{model}}, \delta_{\text{null}})$, where f is the cumulative distribution function of the F statistic. We also assigned nonparametric P values by using a null model based on randomly phase-scrambled permuted data ($n = 1,000$) at each brain location. The nonparametric null model produced P values very close to those constructed from the F statistic. We then corrected for multiple statistical comparisons as above by controlling the FDR.

1. Schiller NO, Bles M, Jansma BM (2003) Tracking the time course of phonological encoding in speech production: An event-related brain potential study. *Brain Res Cogn Brain Res* 17(3):819–831.
2. Abdel Rahman R, Sommer W (2003) Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Brain Res Cogn Brain Res* 16(3):372–382.
3. Rodriguez-Fornells A, Schmitt BM, Kutas M, Münte TF (2002) Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming. *Neuropsychologia* 40(7):778–787.
4. Hanulová J, Davidson DJ, Indefrey P (2011) Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Lang Cogn Process* 26(7):902–934.
5. Camen C, Morand S, Laganaro M (2010) Re-evaluating the time course of gender and phonological encoding during silent monitoring tasks estimated by ERP: Serial or parallel processing? *J Psycholinguist Res* 39(1):35–49.
6. Rahman RA, Sommer W (2008) Seeing what we know and understand: How knowledge shapes perception. *Psychon Bull Rev* 15(6):1055–1063.
7. Aristei S, Melinger A, Abdel Rahman R (2011) Electrophysiological chronometry of semantic context effects in language production. *J Cogn Neurosci* 23(7):1567–1586.
8. Huettig F, Hartsuiker RJ (2008) When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Mem Cognit* 36(2):341–360.
9. Haller S, Radue EW, Erb M, Grodd W, Kircher T (2005) Overt sentence production in event-related fMRI. *Neuropsychologia* 43(5):807–814.
10. Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002) FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17(2):73–88.
11. Garrod S, Pickering MJ (2004) Why is conversation so easy? *Trends Cogn Sci* 8(1):8–11.
12. Indefrey P (2011) The spatial and temporal signatures of word production components: a critical update. *Front Psychol* 2:255.
13. Indefrey P, Levelt WJ (2004) The spatial and temporal signatures of word production components. *Cognition* 92(1–2):101–144.
14. Braun AR, Guillemin A, Hosey L, Varga M (2001) The neural organization of discourse: An H2 15O-PET study of narrative production in English and American sign language. *Brain* 124(Pt 10):2028–2044.
15. Bavelier D, et al. (1997) Sentence reading: A functional MRI study at 4 tesla. *J Cogn Neurosci* 9(5):664–686.
16. Jung-Beeman M (2005) Bilateral brain processes for comprehending natural language. *Trends Cogn Sci* 9(11):512–518.
17. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8(5):393–402.
18. Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* 31(8):2906–2915.
19. Xu J, Kemeny S, Park G, Frattali C, Braun A (2005) Language in context: Emergent features of word, sentence, and narrative comprehension. *Neuroimage* 25(3):1002–1015.
20. Federmeier KD, Wlotko EW, Meyer AM (2008) What's "right" in language comprehension: ERPs reveal right hemisphere language capabilities. *Lang Linguist Compass* 2(1):1–17.
21. Friederici AD (2012) The cortical language circuit: From auditory perception to sentence comprehension. *Trends Cogn Sci* 16(5):262–268.
22. Hagoort P, Levelt WJ (2009) Neuroscience. The speaking brain. *Science* 326(5951):372–373.
23. Pickering MJ, Garrod S (2004) Toward a mechanistic psychology of dialogue. *Behav Brain Sci* 27(2):169–190, discussion 190–226.
24. Okada K, Hickok G (2006) Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain Lang* 98(1):112–117.
25. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7(7):701–702.
26. Menenti L, Gierhan SM, Segaert K, Hagoort P (2011) Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol Sci* 22(9):1173–1182.
27. Segaert K, Menenti L, Weber K, Petersson KM, Hagoort P (2012) Shared syntax in language production and language comprehension—an FMRI study. *Cereb Cortex* 22(7):1662–1670.
28. Heim S (2008) Syntactic gender processing in the human brain: A review and a model. *Brain Lang* 106(1):55–64.
29. Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62(2):816–847.
30. Indefrey P, Hellwig F, Herzog H, Seitz RJ, Hagoort P (2004) Neural responses to the production and comprehension of syntax in identical utterances. *Brain Lang* 89(2):312–319.
31. Price CJ (2010) The anatomy of language: A review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci* 1191:62–88.
32. Awad M, Warren JE, Scott SK, Turkheimer FE, Wise RJ (2007) A common system for the comprehension and production of narrative speech. *J Neurosci* 27(43):11455–11464.
33. Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci USA* 107(32):14425–14430.
34. Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456(7219):189–194.
35. Gupta L, Molfese DL, Tammana R, Simos PG (1996) Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Trans Biomed Eng* 43(4):348–356.
36. Baldo JV, Wilkins DP, Ogar J, Willock S, Dronkers NF (2011) Role of the precentral gyrus of the insula in complex articulation. *Cortex* 47(7):800–807.
37. Kotz SA, Schwartze M, Schmidt-Kassow M (2009) Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45(8):982–990.
38. Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc Natl Acad Sci USA* 94(26):14792–14797.

39. Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326(5951):445–449.
40. Friederici AD (2009) Pathways to language: Fiber tracts in the human brain. *Trends Cogn Sci* 13(4):175–181.
41. Liakakis G, Nickel J, Seitz RJ (2011) Diversity of the inferior frontal gyrus—a meta-analysis of neuroimaging studies. *Behav Brain Res* 225(1):341–347.
42. Seghier ML (2008) Laterality index in functional MRI: Methodological issues. *Magn Reson Imaging* 26(5):594–601.
43. Wilke M, Lidzba K (2007) LI-tool: A new toolbox to assess lateralization in functional MR-data. *J Neurosci Methods* 163(1):128–136.
44. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303(5664):1634–1640.
45. Hasson U, Malach R, Heeger D (2010) Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* 14(1):40–48.
46. Honey CJ, Thompson CR, Lerner Y, Hasson U (2012) Not lost in translation: Neural responses shared across languages. *J Neurosci* 32(44):15277–15283.
47. Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychon Bull Rev* 13(3):361–377.
48. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21(1):1–36.
49. Pickering MJ, Garrod S (2007) Do people use language production to make predictions during comprehension? *Trends Cogn Sci* 11(3):105–110.
50. Poeppel D, Emmorey K, Hickok G, Pylkkänen L (2012) Towards a new neurobiology of language. *J Neurosci* 32(41):14125–14131.
51. Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci* 6(1):37–46.
52. Giraud AL, et al. (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56(6):1127–1134.
53. Ben-Yakov A, Honey CJ, Lerner Y, Hasson U (2012) Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *Neuroimage* 63(1):501–506.
54. Geschwind N (1970) The organization of language and the brain. *Science* 170(3961): 940–944.
55. Geschwind N (1979) Anatomical and functional specialization of the cerebral hemispheres in the human. *Bull Mem Acad R Med Belg* 134(6):286–297.
56. Amunts K, et al. (2010) Broca's region: Novel organizational principles and multiple receptor mapping. *PLoS Biol* 8(9):e1000489.
57. Shergill SS, et al. (2002) Modulation of activity in temporal cortex during generation of inner speech. *Hum Brain Mapp* 16(4):219–227.
58. Curtis VA, et al. (1999) Attenuated frontal activation in schizophrenia may be task dependent. *Schizophr Res* 37(1):35–44.
59. Papoutsi M, et al. (2009) From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. *Cereb Cortex* 19(9):2156–2165.
60. Moss HE, Tyler LK (1995) Investigating semantic memory impairments: The contribution of semantic priming. *Memory* 3(3–4):359–395.
61. Xiao Z, et al. (2005) Differential activity in left inferior frontal gyrus for pseudowords and real words: An event-related fMRI study on auditory lexical decision. *Hum Brain Mapp* 25(2):212–221.
62. Ben-Shachar M, Hendler T, Kahn I, Ben-Bashat D, Grodzinsky Y (2003) The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychol Sci* 14(5):433–440.
63. Ben-Shachar M, Palti D, Grodzinsky Y (2004) Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *Neuroimage* 21(4):1320–1336.
64. Decety J, Chaminade T (2003) Neural correlates of feeling sympathy. *Neuropsychologia* 41(2):127–138.
65. Hennenlotter A, et al. (2005) A common neural basis for receptive and expressive communication of pleasant facial affect. *Neuroimage* 26(2):581–591.
66. Shin LM, et al. (2000) Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biol Psychiatry* 48(1):43–50.
67. Rämä P, et al. (2001) Working memory of identification of emotional vocal expressions: An fMRI study. *Neuroimage* 13(6 Pt 1):1090–1101.
68. Honey GD, Bullmore ET, Sharma T (2000) Prolonged reaction time to a verbal working memory task predicts increased power of posterior parietal cortical activation. *Neuroimage* 12(5):495–503.
69. Rogalsky C, Matchin W, Hickok G (2008) Broca's area, sentence comprehension, and working memory: An fMRI Study. *Front Hum Neurosci* 2:14.
70. Rogalsky C, Hickok G (2011) The role of Broca's area in sentence comprehension. *J Cogn Neurosci* 23(7):1664–1680.
71. Hickok G, Poeppel D (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4(4):131–138.
72. Bates E, et al. (2003) Voxel-based lesion-symptom mapping. *Nat Neurosci* 6(5): 448–450.
73. Dronkers NF (1996) A new brain region for coordinating speech articulation. *Nature* 384(6605):159–161.
74. Eimas PD, Miller JL, Jusczyk PW (1987) On infant speech perception and the acquisition of language. *Categorical Perception: The Groundwork of Cognition*, ed Harnad S (Cambridge Univ Press, Cambridge, UK), pp 161–195.
75. Hickok G, Poeppel D (2004) Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92(1-2):67–99.
76. Pulvermüller F, Fadiga L (2010) Active perception: Sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci* 11(5):351–360.
77. Dronkers NF, Wilkins DP, Van Valin RD, Redfern BB, Jaeger JJ (1994) A reconsideration of the brain areas involved in the disruption of morphosyntactic comprehension. *Brain Lang* 47(3):461–463.
78. Dronkers NF, Redfern BB, Knight RT (2000) The neural architecture of language disorders. *The New Cognitive Neurosciences*, ed Gazzaniga MS (MIT Press, Cambridge, MA), pp 949–958.
79. Dronkers NF, Wilkins DP, Van Valin RD, Jr, Redfern BB, Jaeger JJ (2004) Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92(1–2): 145–177.
80. Caplan D, Hildebrandt N, Makris N (1996) Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* 119(Pt 3):933–949.
81. Caplan D, et al. (2007) A study of syntactic processing in aphasia II: Neurological aspects. *Brain Lang* 101(2):151–177.
82. Damasio AR, Damasio H (2002) *Principles of Behavioral and Cognitive Neurology* (Oxford Univ Press, New York).
83. Binder JR, et al. (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17(1):353–362.
84. Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19(12):2767–2796.
85. Tyler LK, Marslen-Wilson W (2008) Fronto-temporal brain systems supporting spoken language comprehension. *Philos Trans R Soc Lond B Biol Sci* 363(1493):1037–1054.
86. Price CJ (2000) The anatomy of language: Contributions from functional neuroimaging. *J Anat* 197(Pt 3):335–359.
87. Bookheimer S (2002) Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25:151–188.
88. Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6(2):78–84.
89. Vigneau M, et al. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage* 30(4):1414–1432.
90. Ferstl EC, Neumann J, Bogler C, von Cramon DY (2008) The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Hum Brain Mapp* 29(5):581–593.
91. Peelle JE, Johnsrude IS, Davis MH (2010) Hierarchical processing for speech in human auditory cortex and beyond. *Front Hum Neurosci* 4:51.
92. Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Front Syst Neurosci* 5:1.
93. McGuire PK, Silbersweig DA, Frith CD (1996) Functional neuroanatomy of verbal self-monitoring. *Brain* 119(Pt 3):907–917.
94. Hirano S, et al. (1997) Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport* 8(9-10):2379–2382.
95. Mar RA (2011) The neural bases of social cognition and story comprehension. *Annu Rev Psychol* 62:103–134.
96. Amodio DM, Frith CD (2006) Meeting of minds: The medial frontal cortex and social cognition. *Nat Rev Neurosci* 7(4):268–277.
97. Gallagher HL, et al. (2000) Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38(1):11–21.
98. Cavanna AE, Trimble MR (2006) The precuneus: A review of its functional anatomy and behavioural correlates. *Brain* 129(Pt 3):564–583.
99. Buckner RL, Carroll DC (2007) Self-projection and the brain. *Trends Cogn Sci* 11(2): 49–57.
100. Adolphs R (2009) The social brain: Neural basis of social knowledge. *Annu Rev Psychol* 60:693–716.
101. Decety J, Sommerville JA (2003) Shared representations between self and other: A social cognitive neuroscience view. *Trends Cogn Sci* 7(12):527–533.
102. Saxe R (2006) Uniquely human social cognition. *Curr Opin Neurobiol* 16(2):235–239.
103. Fletcher PC, et al. (1995) Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition* 57(2):109–128.
104. Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. *Neuroimage* 30(4):1077–1087.
105. Numminen J, Salmelin R, Hari R (1999) Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett* 265(2):119–122.
106. Tian X, Poeppel D (2013) The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *J Cogn Neurosci* 25(7):1020–1036.
107. Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012) Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends Cogn Sci* 16(2):114–121.
108. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.
109. Behzadi Y, Restom K, Liau J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37(1):90–101.
110. Ellis D Dynamic time warp (DTW) in Matlab (Columbia Univ Laboratory for Recognition and Organization of Speech and Audio, New York). Available at http://labrosa.ee.columbia.edu/matlab/dtw/).
111. Turetsky RJ, Ellis DPW Ground-truth transcriptions of real music from force-aligned MIDI syntheses (Columbia Univ, New York). Available at www.ee.columbia.edu/~dpwe/pubs/ismir03-midi.pdf.
112. Zarahn E, Aguirre GK, D'Esposito M (1997) Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage* 5(3):179–197.
113. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57(1):289–300.
114. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4):1165–1188.