

Predictive Brain Mechanisms in Sound-to-Meaning Mapping during Speech Processing

Bingjiang Lyu,^{1,2,3*} Jianqiao Ge,^{1,2,3*} Zhendong Niu,⁴ Li Hai Tan,⁵ and Jia-Hong Gao^{1,2,3,5}

¹Beijing City Key Laboratory for Medical Physics and Engineering, School of Physics, ²Center for MRI Research, Academy for Advanced Interdisciplinary Studies, and ³McGovern Institute for Brain Research, Peking University, Beijing, China 100871, ⁴School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China 100081, and ⁵Center for Language and Brain, Shenzhen Institute of Neuroscience, Shenzhen, China 518057

Spoken language comprehension relies not only on the identification of individual words, but also on the expectations arising from contextual information. A distributed frontotemporal network is known to facilitate the mapping of speech sounds onto their corresponding meanings. However, how prior expectations influence this efficient mapping at the neuroanatomical level, especially in terms of individual words, remains unclear. Using fMRI, we addressed this question in the framework of the dual-stream model by scanning native speakers of Mandarin Chinese, a language highly dependent on context. We found that, within the ventral pathway, the violated expectations elicited stronger activations in the left anterior superior temporal gyrus and the ventral inferior frontal gyrus (IFG) for the phonological–semantic prediction of spoken words. Functional connectivity analysis showed that expectations were mediated by both top-down modulation from the left ventral IFG to the anterior temporal regions and enhanced cross-stream integration through strengthened connections between different subregions of the left IFG. By further investigating the dynamic causality within the dual-stream model, we elucidated how the human brain accomplishes sound-to-meaning mapping for words in a predictive manner.

Key words: Chinese; context; expectation; fMRI; speech; top-down

Significance Statement

In daily communication via spoken language, one of the core processes is understanding the words being used. Effortless and efficient information exchange via speech relies not only on the identification of individual spoken words, but also on the contextual information giving rise to expected meanings. Despite the accumulating evidence for the bottom-up perception of auditory input, it is still not fully understood how the top-down modulation is achieved in the extensive frontotemporal cortical network. Here, we provide a comprehensive description of the neural substrates underlying sound-to-meaning mapping and demonstrate how the dual-stream model functions in the modulation of expectations, allowing for a better understanding of how the human brain accomplishes sound-to-meaning mapping in a predictive manner.

Introduction

The human brain responds to speech in a rapid manner, typically recognizing words and accessing the corresponding representations within 100–200 ms (Herrmann et al., 2011; MacGregor et al., 2012). The rapid processing of speech relies on both the

recognition of individual words and the contextual information that generates expectations, which entails the use of linguistic context and interaction between context-driven top-down expectations and data-driven bottom-up perceptions (Marslen-Wilson and Welsh, 1978). However, the influence of prior expectations on this highly efficient cortical processing is still not fully understood.

Decades of neuroimaging studies have demonstrated that the neural substrates of speech processing involve distributed frontoparietotemporal cortical areas (Friederici, 2011; Price, 2012), which consist of a dorsal pathway responsible for transforming acoustic speech signals into articulatory codes and a ventral pathway underlying sound-to-meaning mapping (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). In contemporary views, evoked cortical responses can be considered as the brain's attempts to minimize prediction error between expectations and external sensory inputs (Friston, 2005), in which top-down constraints provide guidance for bottom-up sensory observations

Received Feb. 16, 2016; revised July 26, 2016; accepted Aug. 27, 2016.

Author contributions: B.L., J.G., Z.N., L.H.T., and J.-H.G. designed research; B.L., J.G., and J.-H.G. performed research; B.L., J.G., and J.-H.G. analyzed data; B.L., J.G., L.H.T., and J.-H.G. wrote the paper.

This work was supported by China's National Strategic Basic Research Program (973; Grant 2012CB720700), the National Natural Science Foundation of China (Grants 31200761, 31421003, 81227003, and 81430037), Beijing Municipal Science & Technology Commission (Grant Z161100000216152), and Shenzhen Peacock Plan (Grant KQTD2015033016104926).

The authors declare no competing financial interests.

*B.L. and J.G. contributed equally to this work.

Correspondence should be addressed to either Jia-Hong Gao or Jianqiao Ge, Center for MRI Research, Integrated Science Research Building, Peking University, 5 Yiheyuan Road, Haidian District, Beijing, China 100871; E-mail: jgao@pku.edu.cn or gejiao@pku.edu.cn.

DOI:10.1523/JNEUROSCI.0583-16.2016

Copyright © 2016 the authors 0270-6474/16/3610813-10\$15.00/0

(Rao and Ballard, 1999; Friston, 2010). In language processing, such constraints could be prior expectations drawn upon phonological, semantic, or syntactic knowledge based on the context. Readers have been found to anticipate upcoming words based on those appearing earlier in the sentence and exhibit significant preactivated electrical potentials (DeLong et al., 2005). The N400 component has been consistently identified in event-related potential studies when unrelated target words or anomalous sentence endings are presented (Lau et al., 2008). Moreover, semantic expectancy has been shown to constrain the decoding processes of speech, in which an intelligible sentence with low semantic expectancy evokes greater brain activity in the fronto-temporal cortex (Oblener and Kotz, 2010). However, it remains unclear how prior expectations influence the sound-to-meaning mapping at the neuroanatomical level, especially in a natural language context (Gagnepain et al., 2012).

Experimentally induced violations of phonological–semantic expectation can be used to address this question. However, many languages present difficulties in manipulating such violations based on individual words. For instance, although English idiomatic phrases such as “mortal coil” convey unique meanings, the phonological–semantic expectation is weakly embedded in the initial portion of the phrases. It is difficult to predict the word “coil” by only hearing the preceding word “mortal.” However, Chinese idioms of similar syllable length generally have a high transitional probability. For example, phrases consisting of monosyllabic Chinese characters such as “班门弄斧” (meaning showing one’s slight skill before an expert) would be highly predicted by a native Mandarin Chinese speaker after hearing the first portion “班门”. Therefore, by manipulating the last portion of a Chinese idiom, expectation violations can naturally be induced in native Mandarin Chinese speakers.

A recent fMRI meta-analysis demonstrated a processing gradient for phonetic stimuli along the human auditory ventral stream, with word-length stimuli localized to the left anterior superior temporal cortex (DeWitt and Rauschecker, 2012). Moreover, the dorsal–ventral streams architecture for speech processing highlights the role of the prefrontal cortex, particularly the inferior prefrontal gyrus (IFG), in cross-stream integration and mediation of top-down feedback across streams (Bornkessel-Schlesewsky et al., 2015). Because phrases with violated phonological–semantic expectation are likely to increase the processing demands on spoken words, particularly word recognition and combination, we hypothesized that unexpected speech would more strongly engage the ventral pathway than would expected speech, especially the anterior portion, and would require more intensive cross-stream integration to exert top-down modulation on the mediation among conflicting representations during sound-to-meaning mapping.

To test this hypothesis, fMRI was used to record the brain activity of native Mandarin Chinese speakers as they listen to expected, unexpected, and time-reversed phrases. The neural substrates involved in this sound-to-meaning mapping were delineated using both univariate and multivariate analyses and their functional and effective connectivity were further investigated to reveal how the interconnections among identified cortical regions are modulated by fulfilled or violated expectations.

Materials and Methods

Participants. Thirty right-handed native Mandarin Chinese speakers participated in this study (age 21–28 years, mean 24.2, 15 male) as paid volunteers, all with normal hearing and normal or corrected-to-normal vision. None of the participants reported having a history of mental disorders or language impairment. All participants provided written in-

formed consent and the study was conducted under the approval of the Institutional Review Board of Beijing MRI Center for Brain Research. One female subject with large head motions (>3 mm) during fMRI scanning was excluded from further analysis, so all of the following results (both univariate and multivariate) are based on the remaining 29 subjects.

Stimuli. Three types of auditory stimuli, expected phrases (EPs), unexpected phrases (UPs), and time-reversed phrases (TPs), were presented to the subjects. The EPs were Chinese idioms consisting of three to five characters, such as “耳边风” (er3 bian1 feng1, meaning unheeded advice; the letters represent the official Romanization “Pinyin” of Mandarin Chinese and the numbers represent the corresponding tones) and “班门弄斧” (ban1 men2 nong3 fu1, meaning showing one’s slight skill before an expert), whereas each character itself is a monosyllabic word (or morpheme). To investigate the impact of violated expectations on phonological–semantic mapping, auditory stimuli of UPs were created by keeping the first two characters of an idiom and replacing the last portion with character(s) from another irrelevant idiom. As a consequence, the UP stimuli consisted of words from two unrelated idioms; for example, “鹤立之鉴” (he4 li4 zhi1 jian4) was the rearranged combination of the first two characters of the idiom “鹤立鸡群” (he4 li4 ji1 qun2) and the last two characters of the idiom “前车之鉴” (qian2 che1 zhi1 jian4). Therefore, the UPs generally had the same phonological features as the expected ones, but violated the originally embedded phonological–semantic expectation. The EPs were idioms with specific meanings, whereas the UPs were meaningless as a phrase even though each syllable was still recognizable. Time-reversed phrases (TPs) were derived equally from both the EPs and UPs to create a low-level acoustic match because this manipulation preserved the acoustic spectrum and voice identity information while removing the intelligibility of the original speech. The stimuli of EPs and UPs were recorded digitally in a soundproof studio by one male and one female native Chinese speaker and the recorded sounds were then edited by Praat software (<http://www.praat.org>). There were 84 auditory stimuli (half by the male speaker and half by the female speaker) for each type of stimuli. All stimuli were edited for length (730–1007 ms) and quality and were amplified to ensure no significant differences between the speakers or among the different types of stimuli.

Procedure. The experimental procedure was adopted from a previous fMRI study on the cortical responses to intelligible speech (Leff et al., 2008). The experiment was organized in a block design with four sessions, with each session consisting of a 10 s dummy scan at the beginning and nine subsequent task blocks. Each task block started with a preparation cue lasting 3.15 s and ended with a blank screen with fixation for 9.45 s, and the whole block lasted for 40.95 s. In each block, seven stimuli of the same type (UPs, EPs, or TPs) with an alterable gender ratio (2:5, 3:4, 4:3, or 5:2) were played during the first 1180 ms, followed by a response cue and a fixation period lasted for 2420 and 450 ms respectively. Subjects were not informed of the content of the auditory stimuli in advance. They were asked to judge the gender of the speakers and to indicate their answers by pressing the corresponding keys (which were counterbalanced among subjects) after the presentation of response cues. The sequences of the task blocks and the speaker gender ratio were pseudorandomized across subjects and all auditory stimuli were played without repetition.

MRI data acquisition. MRI data were acquired using a Siemens Trio 3T scanner with a standard head coil. Auditory stimuli were presented binaurally using a pair of home-made MRI-compatible headphones. Given the noise during scanning, subjects were allowed to adjust the volume of the stimuli to a comfortable level during a short testing scan before the formal sessions. The auditory stimuli were presented initially at 80 dB/sound pressure level (SPL) and the headphones provided 25 dB/SPL attenuation of the scanner noise. Each session consisted of 182 whole-brain volumes. Thirty-five axial slices that covered the whole brain were acquired using a T2*-weighted gradient-echo EPI sequence with the following parameters: TR/TE/FA = 2.08 s/30 ms/90°, matrix size = 64 × 64, in-planar image resolution = 3 mm × 3 mm, slice thickness = 3 mm, slice interval = 0.75 mm. A high-resolution T1-weighted image was

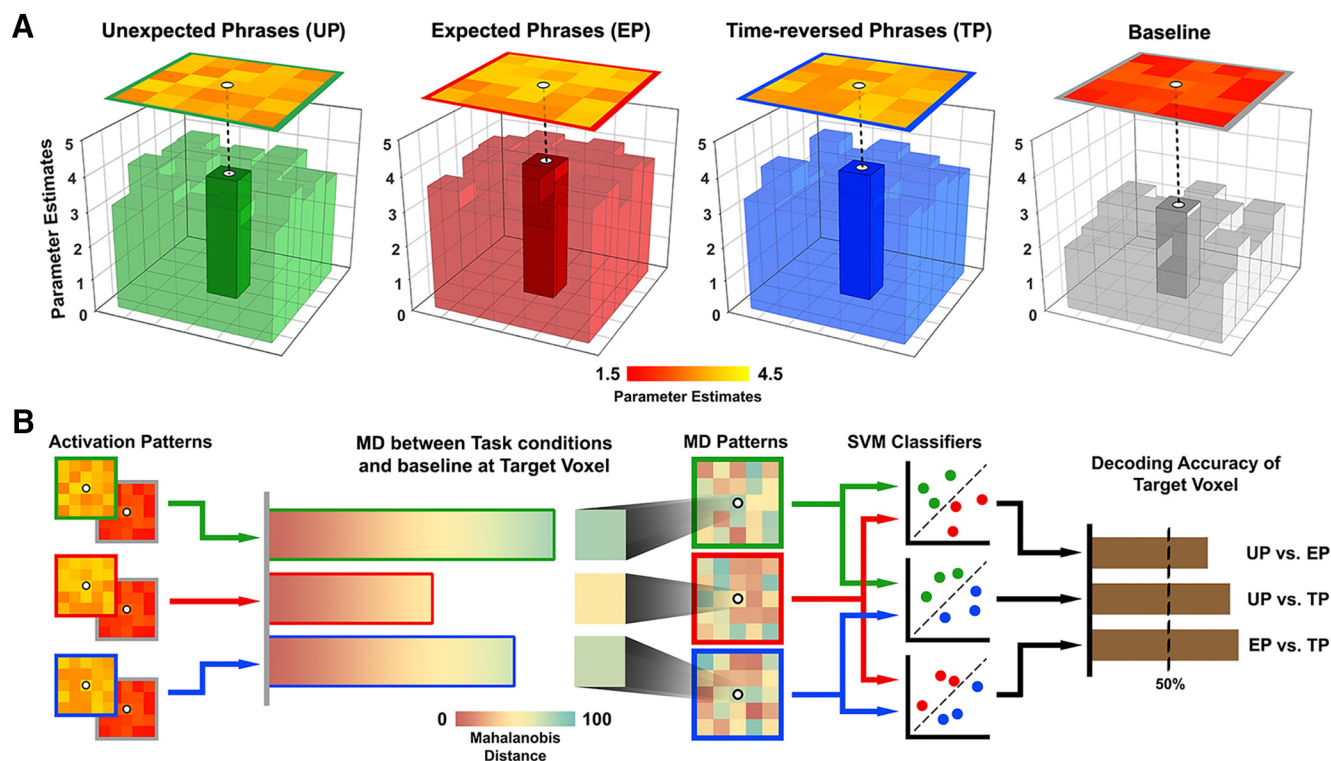


Figure 1. Schematic diagram showing complementary contributions of the univariate and multivariate analyses. While the univariate analysis reveals brain areas that show significantly different activation strengths among the different conditions (i.e., UPs, EPs, and TPs), the multivariate analysis reveals the differences in brain activity patterns even when the activation strengths are comparable. The values shown are arbitrary and are for illustrative purposes only. **A**, Local activation strength indicated by parameter estimates of the target voxel (i.e., solid colors in the center of each matrix) in response to UPs (green), EPs (red), TPs (blue), and at baseline (gray). Local activation patterns around the target voxel are shown in transparent colors. **B**, Procedure for the MVPA, with brain activity patterns indexed by the MD between the three task conditions and the baseline for each voxel, which is calculated based on the corresponding univariate parameter estimates (normalized before the calculation). The MD patterns for each task condition were then entered into the linear SVM classifiers to distinguish the activity patterns of different task conditions.

collected for anatomical details with isotropic 1 mm resolution using the MPRAGE sequence (TR/TE/FA = 2.6 s/3.02 ms/8°).

Postscan behavioral dictation test. Subjects of the fMRI experiment were contacted at least 4 months later and asked to participate in a surprise dictation test. Fifteen subjects (age 23–26 years, mean 24.3, 7 male) agreed to participate in this test. The same stimuli of EPs and UPs to which they had been exposed in the scanner were presented one at a time in a random sequence and the subjects were asked to write down the Chinese characters of the phrases they heard. The dictation results were classified into three categories: (1) correct; (2) phonologically correct (where the pronunciation of the characters was shown to be correct, but the characters were not correctly identified, i.e., the homophones); and (3) incorrect (both the word identity and the pronunciation were incorrect). Performance in the dictation test was measured based on the three categories above. In the following analyses, the dictation accuracy denotes the percentage of correct characters for all of the phonologically correct characters, which indicates successful sound-to-meaning mapping. An intersubject brain–behavior correlation analysis was conducted based on the subjects who participated in both the fMRI experiment and the postscan behavioral study.

Univariate analysis. To identify brain regions that showed significantly greater activities in response to different phrases, univariate analysis was performed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm>, RRID: SCR_007037). The first five images acquired during the dummy scan were discarded to avoid T1 saturation effects. Functional images were corrected for slice-timing and head-motion effects and were normalized to the Montreal Neurological Institute (MNI) standard space using individual high-resolution T1 anatomical images. The normalized images were then smoothed using a 6 mm isotropic Gaussian kernel. In the general linear model (GLM), the timings of different auditory stimuli were convolved with a canonical hemodynamic response function to

model their respective effects, with the six head motion parameters included as nuisance regressors. Whole-brain statistical parametric mappings were generated and contrasts were then defined to reveal brain areas specifically involved in the processing of each type of phrase. The individual statistical maps were further entered into a group-level random-effects analysis and the significant clusters were identified using a voxelwise threshold of $p < 0.001$ with cluster-level FWE correction at $p < 0.05$.

Multivariate pattern analysis. Multivariate pattern analysis (MVPA) was also conducted to identify the brain regions that consistently demonstrated a distinct spatial pattern of activity for one condition relative to another. Although the univariate analysis reveals brain areas that show significantly different activation strengths under different conditions, MVPA identifies the differences between brain activity patterns even when the activation strengths are comparable (Fig. 1A). Recently, a growing number of studies have adopted MVPA as a new way to investigate speech processing at the neuroanatomical level (Okada et al., 2010; McGettigan et al., 2012; Abrams et al., 2013; Boets et al., 2013; Evans et al., 2014). The combined use of univariate and multivariate methods will thus provide necessary complementary information (Poldrack and Farah, 2015) to improve the understanding of brain activity during the top-down modulation of speech processing.

In this study, MVPA was performed in a spherical searchlight manner to assess how well a voxel's local activity pattern could be used to distinguish between different conditions. The logic of such an analysis is that, if a classifier can distinguish among local activity patterns induced by different conditions with an accuracy significantly higher than chance level, then there should be information about these conditions embedded in such local patterns. Classifiers were trained using support vector machine (SVM) algorithms to differentiate local patterns induced by different conditions. The voxelwise Mahalanobis distances (MDs) between

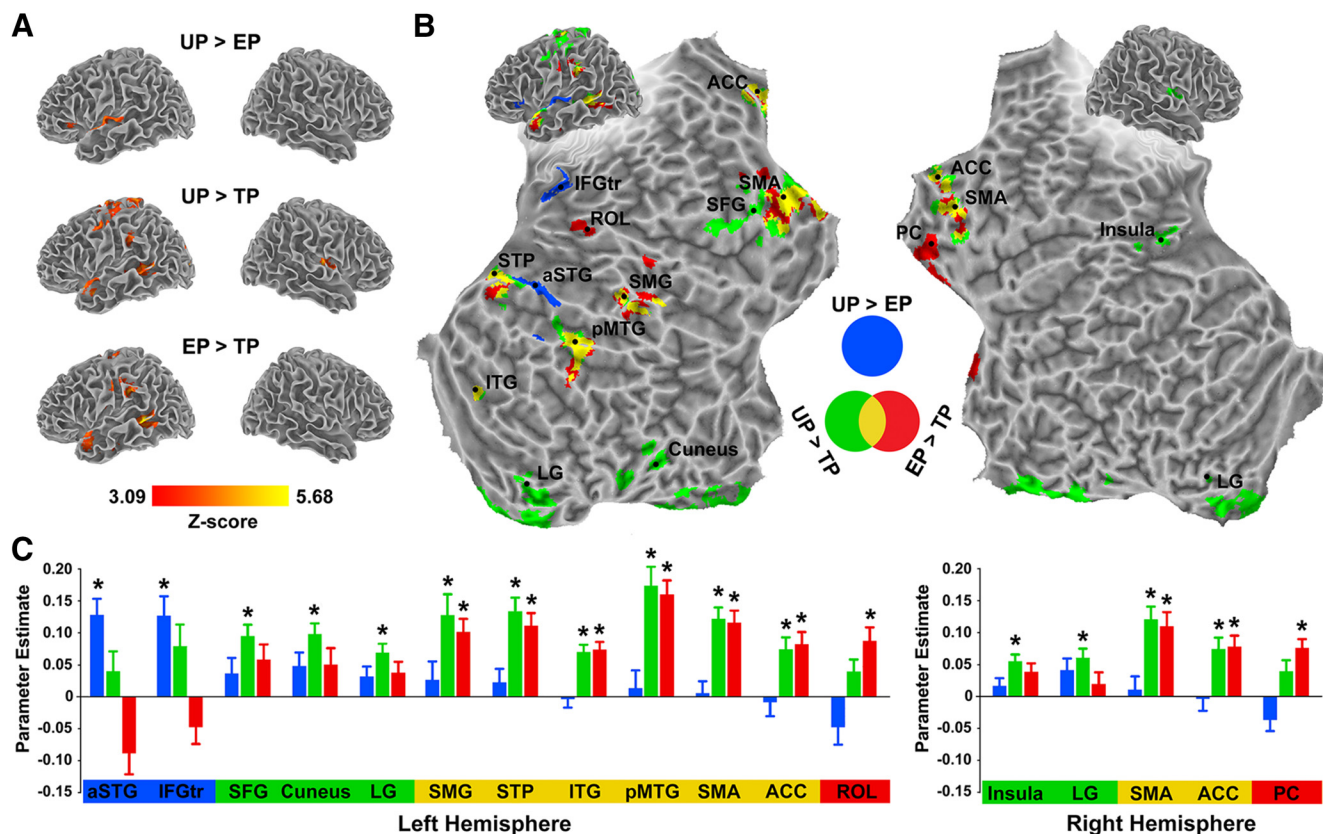


Figure 2. Univariate results showing brain activation strength in different contrasts. *A*, Univariate activations found in different contrasts. *B*, Overlap of brain regions with activations under different contrasts shown on an inflated cortical map. The results shown were thresholded at voxel-level $p < 0.001$ with a cluster-level FWE correction of $p < 0.05$. *C*, Parameter estimates were extracted from 6-mm-spherical ROIs centered on the foci in the overlapping regions that were obtained by averaging the coordinates of the nearest local maxima found in the respective contrasts with FDR correction of $p < 0.01$.

task conditions (i.e., EPs, UPs, and TPs) and the baseline were used as features in the training of SVM classifiers. Here, the baseline refers to the constant item in the GLM analysis. The resting period (screen with only a fixation cross) was not explicitly modeled in the GLM analysis to avoid an overparameterized model (Pernet, 2014), so it was considered to be part of the baseline. The calculation of MDs takes the covariance structure into consideration and decreases the contribution from voxels that are noisy or strongly correlated with other voxels. Therefore, it accounts for the similarity between patterns, as well as the reliability and independence of each voxel in the pattern (Serences and Boynton, 2007). The MD maps were obtained using a sphere with 6 mm radius. This radius size was chosen because, in the original study (Kriegeskorte et al., 2006), the optimal detection performance was achieved using an intermediate-sized searchlight about twice the voxel size (i.e., 6 mm in the current study). Note that there was a 0.75 mm gap between slices, so there were actually 31 voxels in each 6-mm-radius sphere. The MD between condition A and B at voxel v_i is defined as follows:

$$MD(v_i) = [(\beta_B - \beta_A)\hat{\Sigma}^{-1}(\beta_B - \beta_A)^T]^{1/2} \quad (1)$$

where β_j is a 31-dimension vector consisting of the parameter estimates of condition j (β_j is z-normalized within each sphere to avoid the effects of the overall amplitude on the classification results), $\hat{\Sigma}$ is an estimate of the error covariance matrix Σ for the voxels within the sphere, and T refers to the transposition of vector. Σ was estimated using the R package “corpcor” (Schäfer and Strimmer, 2005), which realizes the optimal shrinkage toward the diagonalized sample covariance.

The framework of the current MVPA analysis is shown in Figure 1B. For each subject, the GLM was redone for each session using functional images that were only head-motion corrected (neither normalized nor smoothed). The voxelwise MDs between task conditions and baseline were calculated for each of the four sessions. Therefore, three voxelwise MD maps (i.e., EPs, UPs, and TPs) were generated for each session. For each voxel, feature vectors

containing the MD values of all the 31 voxels within the sphere were extracted from a 6-mm-radius sphere in corresponding MD maps. The searchlight MVPA was performed using SVM algorithms with a linear kernel and a default cost parameter of $C = 1$. At each voxel v_i , a classifier was trained to discriminate between two task conditions based on their respective feature vectors. Three pairwise classifications were performed: (1) UPs versus EPs, (2) UPs versus TPs, and (3) EPs versus TPs. SVM training was conducted using LIBSVM (Chang and Lin, 2011; RRID: SCR_010243) in a fourfold cross-validation manner. Each classifier was trained using data from three sessions and was tested with data from the withheld session. The average accuracy at each voxel was used to compose a whole-brain accuracy map, which was then normalized to MNI standard space for group analysis. The classification accuracy map instead of the MD map was normalized to avoid the influence of warp and interpolation on SVM classifier training. The classification accuracy value at each voxel was subtracted from 0.5 (i.e., chance level) and the zero-centered images were then submitted to one-sample t test. Clusters with accuracy significantly higher than the chance level were determined using a voxel-level threshold of $p < 0.001$ with cluster-level FWE correction at $p < 0.05$. It has been proposed that group-level MVPA results should be evaluated statistically using nonparametric permutation methods or the binomial distribution because the decoding accuracy does not exactly follow the normal distribution (Pereira and Botvinick, 2011; Stelzer et al., 2013). However, given that there were only four sessions in this study, the number of permutations is too small to obtain a distribution of the bootstrap classification accuracy. Instead, one-sample t test was used with a conservative threshold, which was also adopted in a recent study (Evans et al., 2014).

Due to the stimuli and methods we used, a direct interpretation of the specific representations (e.g., phonological or semantic) decoded by the MVPA is beyond the scope of the current study. Here, the main motivation behind the combination of both univariate and multivariate analyses was to provide a more comprehensive description of the neural

substrates underlying speech processing, as has been done in previous studies (McGettigan et al., 2012; Evans et al., 2014).

Psychophysiological interactions. After identifying brain areas involved in the processing of EPs or UPs, a psychophysiological interaction (PPI) analysis was performed (Friston et al., 1997) to examine their mutual functional connectivity. The PPI is a measurement of covariation that identifies task-dependent interaction between brain areas. We were particularly interested in the variation of functional connectivity within the left frontotemporal cortical areas during the processing of EPs or UPs. Seven seed areas were selected, including six identified in the univariate analysis, the anterior superior temporal gyrus (aSTG), pars triangularis of the IFG (IFGtr), superior temporal pole (STP), posterior middle temporal gyrus (pMTG), supplementary motor areas (SMA), and primary auditory cortex (PAC), and one revealed in the multivariate analysis, the pars opercularis of the IFG (IFGop). The individual time courses of each seed area were extracted within a 6-mm-radius sphere according to corresponding coordinates of group activation results. The PPI design matrix included the psychological regressor (i.e., UPs > EPs, UPs > TPs, or EPs > TPs), the physiological regressor (i.e., time course of the seed area), the PPI regressor (i.e., dot product of the first two regressors, which reflects the interaction of psychological variable and physiological reaction), and the six head motion parameters as nuisance regressors. The individual parameter estimates of the PPI regressor were further entered into a group-level random-effects analysis and the significant clusters were identified using a voxelwise threshold of $p < 0.005$ with cluster-level FWE correction at $p < 0.05$.

Dynamic causal modeling. Dynamic causal modeling (DCM) (Friston et al., 2003) analysis was performed to investigate how activity in one brain area is affected by the other areas through calculating the effective connectivity among them. The DCM analysis uses a directed model of how observed data were caused based on hidden neuronal and biophysical states. To be specific, it describes how the current state of one brain area causes the dynamic changes in another through endogenous connections and how these interactions change under the external influence of experimental manipulations (Friston, 2010). It models both the strength of intrinsic connection when no external inputs exist and the strength of the modulations of these connections caused by external inputs. A recent study demonstrated the modulation of top-down predictions in the framework of predictive coding (Tuennerhoff and Noppeney, 2016) by comparing different DCM families consisting of three nodes in the left temporal lobe. Here, we intended to consider simultaneously a greater number of nodes in the left frontotemporal cortex to elucidate how sound-to-meaning mapping is modulated by speech stimuli that fulfills or violates expectation. Seven brain areas in the previous PPI analysis were used as nodes to construct DCM models. For individual subjects, the location of each cortical node was defined by searching for the peak activation (uncorrected voxel-level $p < 0.05$) within a 6-mm-radius sphere centered at the coordinates of group activation results, which was also masked by the corresponding automated anatomical labeling templates to avoid contribution from adjacent areas. Then, the time series of each node were extracted as the principal eigenvalue of an activated cluster with at least 30 voxels. Because the specific location of activation might vary across subjects, this procedure guaranteed comparability among models by applying both functional and anatomical constraints in the extraction of time series (Stephan et al., 2007; Heim et al., 2009). Using these criteria, time series for all seven nodes in 22 of the 29 subjects were defined and extracted. The remaining subjects in whom at least one node did not meet the criteria were excluded from further DCM analysis.

Considering the massive computational demands for the immense number of plausible alternatives of such a large-scale model, a *post hoc* DCM approach was adopted because it provided an approximation of the model evidence by optimizing only the largest (full) model of the whole model set with all possible connections (Friston and Penny, 2011; Rosa et al., 2012). The recent developments in *post hoc* model optimization have expanded its use in data-driven analyses for bias-free network discovery (Friston et al., 2011; Seghier and Friston, 2013). In the full model, the endogenous connections and the modulations induced by auditory inputs (dashed lines in Fig. 6A) were hypothesized according to

Table 1. Brain activations identified in the univariate analysis

Regions	BA	MNI coordinates (mm)	Z-value	Size (voxels)
UP > EP				
L aSTG	22	−62, 0, −2	5.30	304
L IFGtr	45	−34, 30, 4	3.74	160
UP > TP				
L AND R SMA	6	−2, −8, 66	5.19	927
R Insula	13	38, −14, 8	5.09	226
L pMTG	22	−58, −46, 4	4.98	297
L STP	38	−52, 14, −18	4.94	186
L AND R Cuneus/Calcarine/Lingual	19	−12, −96, 24	4.66	1537
L Lingual	18	−18, −74, −6	4.09	346
R Lingual	18	18, −76, −4	4.26	151
L SFG	6	−24, −8, 74	4.39	134
L ITG/FF/Hippocampus	20	−42, −22, −18	4.33	207
L SMG	40	−62, −32, 30	3.96	111
EP > TP				
L pMTG	22	−60, −46, 2	5.68	325
L and R SMA	6	−4, −8, 64	5.16	975
L STP	38	−52, 12, −16	4.65	212
R Precuneus/Paracentral Lobule	31	12, −26, 48	4.68	286
L SMG	40	−60, −34, 28	4.52	289
L ITG/FF/Hippocampus	20	−38, −22, −22	4.36	225
L ROL	13	−48, −4, 12	4.03	112

Thresholded at voxel-level $p < 0.001$ with cluster-level FWE correction of $p < 0.05$. FF, fusiform; ITG, inferior temporal gyrus; ROL, rolandic operculum; SFG, superior frontal gyrus; SMG, supramarginal gyrus.

Table 2. Above-chance classification performance in the multivariate analysis

Regions	BA	MNI coordinates (mm)	Z-value	Size (voxels)
UP vs EP				
L pMTG	22	−60, −32, 2	5.30	304
UP vs TP				
L MTG/STG	22/21	−58, −22, −4	5.25	1920
L IFGtr/IFGob/IFGop	45	−50, 34, 10	4.59	847
L IFGob	47	−46, 30, −4	4.48	263
L IFGop	44	−58, 12, 14	4.26	79
L AND R SMA	6	−2, 4, 66	4.86	500
R STG/MTG	21/22	62, −24, 0	4.60	276
L SFG	10	−18, 54, 28	4.52	194
L MFG	9	−34, 20, 42	4.88	151
EP vs TP				
R STG/MTG/ROL/SMG	22/21/42/4	60, −26, 4	5.53	1764
L MTG/STG/STP	22/21/38	−56, −2, −6	5.04	1056

Thresholded at voxel-level $p < 0.001$ with cluster-level FWE correction of $p < 0.05$. MFG, middle frontal gyrus; ROL, rolandic operculum; SFG, superior frontal gyrus; SMG, supramarginal gyrus.

the dual-stream architecture. Sensory input (i.e., auditory stimuli) enters the model via the PAC and then is diverged into an anterodorsal stream including the aSTG, STP, and IFGtr and a posterodorsal stream consisting of the pMTG, SMA, and IFGop. The endogenous effects also include self-connections within the seven nodes. The *post hoc* DCM analysis was conducted using DCM12 and the full model for each subject was estimated and entered into the *post hoc* Bayesian model selection to find the one with the highest model evidence over possible reduced candidate models. The parameters in the winning model were computed via Bayesian parameter averaging (Kasess et al., 2010).

Results

Behavioral results

In the fMRI experiment, the gender judgment task was well performed by the subjects, which is demonstrated by the high accuracy for all types of phrases (UPs, $98.9 \pm 2.0\%$, mean \pm SD; EPs, $99.1 \pm 1.9\%$; and TPs, $98.0 \pm 2.4\%$). The accuracy for TPs was significantly lower than that of EPs and UPs (UPs vs TPs, $t_{(29)} =$

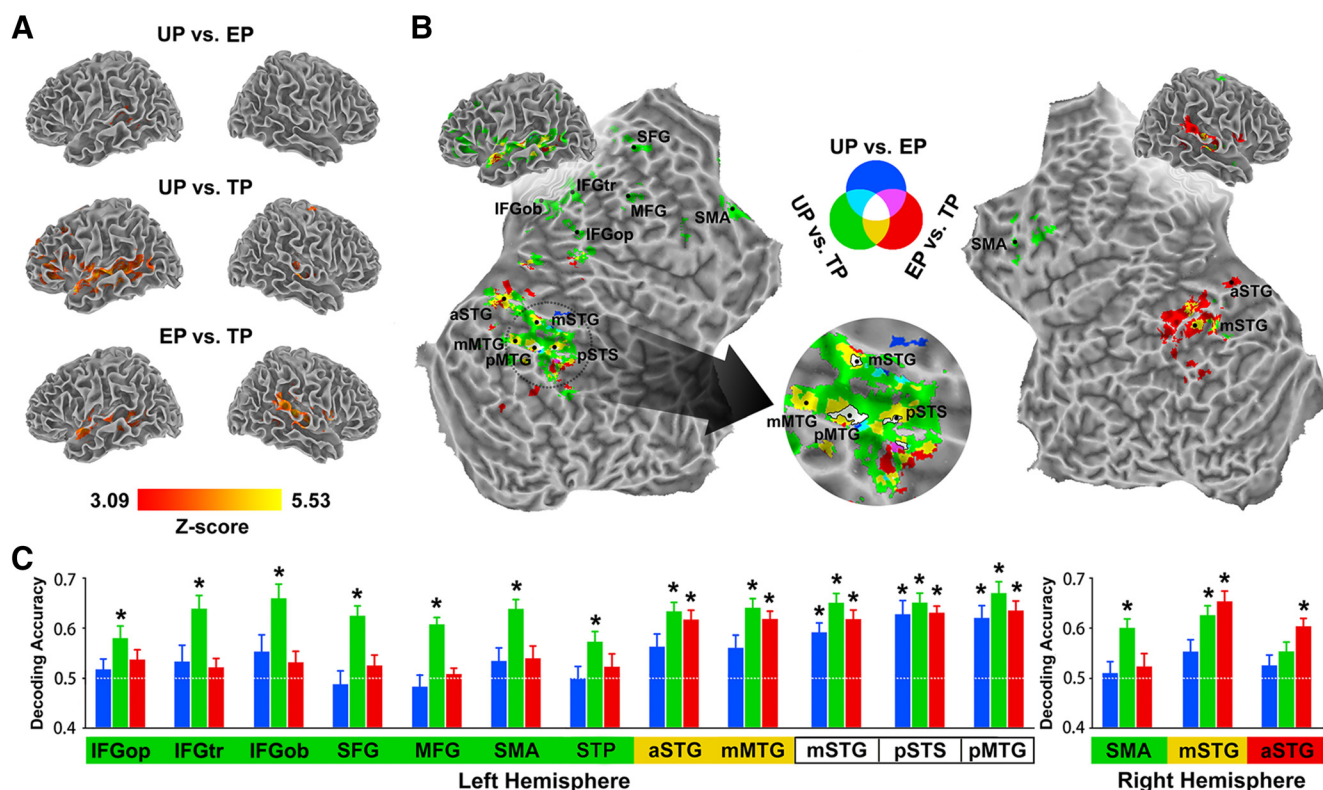


Figure 3. MVPA results. **A**, Informative regions identified via pattern analysis between task conditions. **B**, Overlap of informative brain regions shown on an inflated cortical map. The results were thresholded at voxel-level $p < 0.001$ with a cluster-level FWE correction of $p < 0.05$. **C**, Decoding accuracies were extracted from 6-mm-spherical ROIs centered on the foci in the overlapping regions that were obtained by averaging the coordinates of the nearest local maxima found in the respective classifications with FDR correction of $p < 0.01$.

2.75, $p = 0.020$; EPs vs TPs, $t_{(29)} = 4.09$, $p < 0.001$), which might be due to the unfamiliarity of the time-reversed speech. No significant difference was found between UPs and EPs (UPs vs EPs, $t_{(29)} = -0.78$, $p = 0.44$). The reaction time was 560.2 ± 188.9 ms for UPs, 533.6 ± 176.7 ms for EPs, and 555.9 ± 175.6 ms for TPs. A significant difference in reaction times was only found between UPs and EPs (UPs vs EPs, $t_{(29)} = 2.86$, $p = 0.008$), which suggests a higher processing demand for the novel UPs stimuli. To ensure that the task performance was not biased by gender effects, a 2×2 repeated-measures ANOVA with the genders of the subjects and speakers as factors was performed and no significant gender effects were found.

In the postexperiment dictation test, subjects showed significantly lower accuracies for UPs than for EPs in both the “correct” and “phonologically correct” categories (correct: UPs $70.1 \pm 11.1\%$ vs EPs $92.4 \pm 3.0\%$, $t_{(12)} = -8.20$, $p < 0.001$; phonologically correct: UPs $93.7 \pm 3.0\%$ vs EPs $99.0 \pm 0.9\%$, $t_{(12)} = -9.39$, $p < 0.001$), which suggests that it is more difficult to identify the words and the corresponding phonological features when the phonological–semantic expectation is violated. A 2×2 repeated-measures ANOVA also showed a significant interaction between phrases types and correct categories ($F(1,14) = 69.75$, $p < 0.001$), which further indicates that, compared with phonological processing, the identification of words is more difficult for the UPs.

Neural substrates revealed via univariate analysis

Compared with the effect of EPs, the UPs elicited stronger activations in the left aSTG and IFGtr (Fig. 2A, first row), whereas no significant activation was found for the opposite contrast (EPs > UPs). Relative to the TPs, both EPs and UPs gave rise to similar activations in the left temporal lobe (Fig. 2A, second and third

rows). The overlapping areas of activations for these two contrasts were mainly located in the left STP, pMTG, inferior temporal gyrus, and bilateral SMA (Fig. 2B, yellow regions). The contrast values for the parameter estimates were extracted and averaged from each subject as a 6-mm-radius spherical ROI for all of the activated clusters (Fig. 2C, Table 1). Although the left aSTG and IFGtr exhibited negative parameters for EPs relative to TPs, both EPs and UPs themselves elicited significant positive activations in these areas.

Neural substrates revealed via MVPA

In the whole-brain searchlight MVPA, voxelwise linear SVM classifiers were trained for the following condition pairs: UPs versus EPs, UPs versus TPs, and EPs versus TPs (see details in Table 2). In the classification of UPs versus EPs, a cluster centered in the left pMTG was identified with significantly higher accuracy than chance level (Fig. 3A, first row). As for the classifications of UPs versus TPs and EPs versus TPs, above-chance accuracies were found within the bilateral STG/MTG (Fig. 3A, second and third rows). In the left hemisphere, informative regions were identified along the superior temporal sulcus (STS), with a more extended range observed for UPs versus TPs compared with EPs versus TPs. In contrast, a more extensive informative cluster in the right STG/MTG was found for EPs versus TPs than for UPs versus TPs. The activity in the left aSTG was distinguishable for both UPs and EPs versus TPs, whereas the left STP only showed above-chance performance for UPs. In the frontal cortex, all the three subregions of the left IFG showed significant above-chance performance for the classification of UPs versus TPs, as did the left superior frontal gyrus, middle frontal gyrus, and bilateral SMA. The overlap of informative areas for UPs versus TPs and EPs

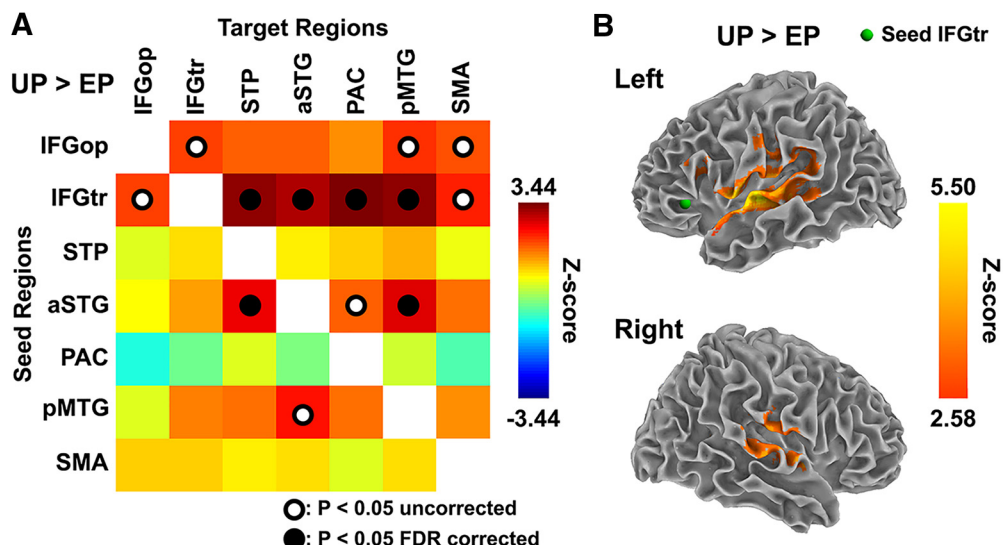


Figure 4. Results of the PPI analysis. **A**, PPI connectivity for UPs > EPs among the seven seed areas within the left frontotemporal cortex. Warm colors indicate a strengthened connection for UP relative to EPs and cold colors represent the opposite effect. **B**, Enhanced PPI connectivity with the IFGtr as the seed area induced by UPs relative to that induced by EP. The results were thresholded at voxel-level $p < 0.005$ with a cluster-level FWE correction of $p < 0.05$.

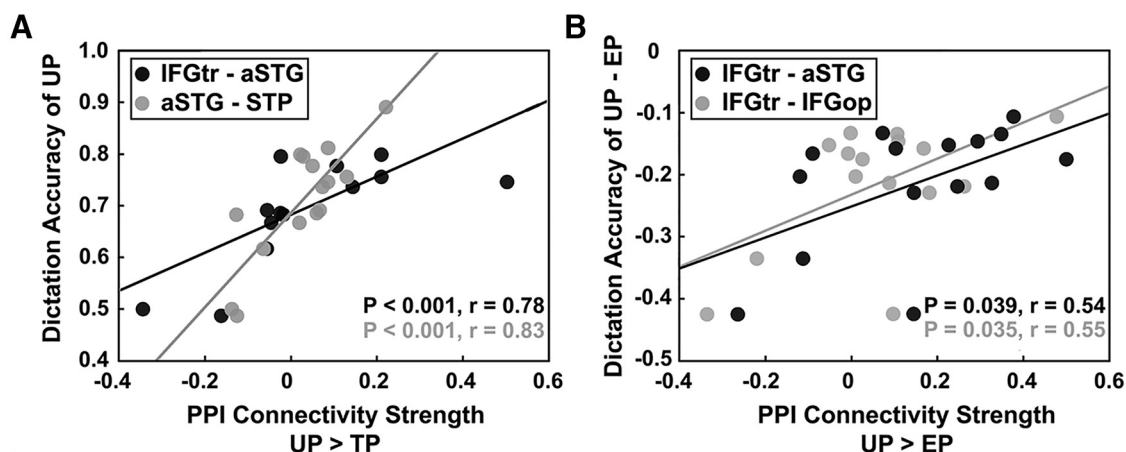


Figure 5. Intersubject brain-behavior correlation results. **A**, **B**, The dictation accuracy denotes the percentage of correct characters identified of all the phonologically correct characters during the postscan dictation test, which indicates successful sound-to-meaning mapping. These results survived multiple-comparisons correction (FDR $p < 0.05$).

versus TPs was found along the left STG/MTG and in the right middle STG/MTG (Fig. 3B, yellow regions). A cluster centered at the left pMTG, which also extended into the left middle STG and posterior STS, was found to be distinguishable for all three types of phrases (Fig. 3B, white regions in the magnified circle). For all of the classifications, the average decoding accuracy in a 6-mm-radius sphere of all the informative areas is detailed in Figure 3C.

PPIs within the frontotemporal cortex

The PPI analysis found that the left IFGtr tended to exhibit stronger functional connectivity with other seed areas along both streams for UPs than for EPs (Fig. 4A,B), which agrees with its role in cross-stream integration. More evidence for this function is provided by the brain-behavior analysis. We found that the strengths of two connections with IFGtr serving as seed area were both positively correlated with subject performance in the dictation test, one along the ventral stream (between IFGtr and aSTG; $r = 0.78$, $p < 0.001$; $r = 0.54$, $p = 0.039$; Fig. 5A,B), and the other within the dorsal stream (between IFGtr and IFGop; $r = 0.55$, $p = 0.035$; Fig. 5B). The

strengthened connections between IFGtr and these two processing streams predicted better dictation performance (i.e., smaller differences between the dictation accuracy for UPs and EPs). Moreover, the strengthened connectivity between aSTG and STP for processing UPs relative to TPs also showed a positive correlation with the dictation accuracy for UPs ($r = 0.83$, $p < 0.001$; Fig. 5A). These results suggest that, especially when expectations are violated, top-down constraints are likely to be exerted on the ventral stream, facilitating the interconnections within anterior temporal regions.

Dynamic causality within the dual-stream model

A *post hoc* DCM analysis was conducted to investigate how sound-to-meaning mapping is modulated by speech that fulfills or violates expectations within the dorsal-ventral streams architecture. Our full model included seven nodes within the left frontotemporal cortex: the PAC from which auditory input enters the model, the aSTG, STP, IFGtr along the ventral stream and the pMTG, SMA, and IFGop along the dorsal stream (Fig. 6A). The winning model preserved all of the hy-

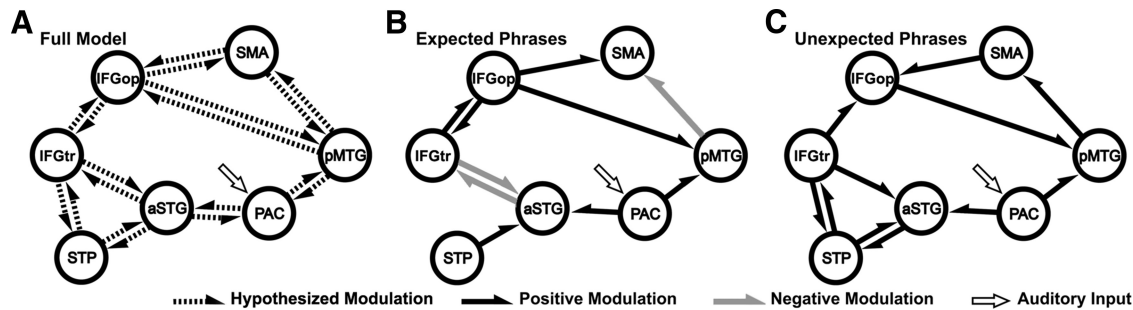


Figure 6. DCM specification and modulation results. **A**, Full model specified according to the dual-stream model. The directed dashed lines indicate the hypothesized connections. The hollow arrows represent auditory inputs. Significant modulatory effects of EPs and UPs are shown separately in **B** and **C**. The black and gray arrows indicate positive and negative modulation, respectively (detailed parameters are listed in Table 3).

Table 3. Posterior estimates of significant modulatory effects and auditory inputs in the winning model (mean ± SD)

Connections	Expected phrases		Unexpected phrases	
	Strength (Hz)	<i>p</i> -value	Strength (Hz)	<i>p</i> -value
IFGtr → IFGop	0.43 ± 0.05	0.021*	0.31 ± 0.02	0.002*
IFGtr → STP	0.04 ± 0.03	0.386	0.46 ± 0.03	<0.001*
IFGtr → aSTG	−0.32 ± 0.04	0.027*	0.40 ± 0.04	0.014*
IFGop → IFGtr	0.38 ± 0.04	0.017*	0.32 ± 0.09	0.229
IFGop → SMA	0.23 ± 0.02	0.011*	0.06 ± 0.05	0.407
IFGop → pMTG	0.52 ± 0.05	0.019*	0.49 ± 0.04	0.008*
SMA → IFGop	0.07 ± 0.07	0.418	0.29 ± 0.02	0.005*
SMA → pMTG	−0.32 ± 0.08	0.205	−0.10 ± 0.07	0.383
STP → IFGtr	0.12 ± 0.03	0.202	0.48 ± 0.06	0.035
STP → aSTG	0.69 ± 0.02	<0.001*	0.61 ± 0.02	<0.001*
aSTG → IFGtr	−0.35 ± 0.02	<0.001*	−0.10 ± 0.09	0.406
aSTG → PAC	0.34 ± 0.07	0.159	−0.002 ± 0.13	0.499
aSTG → STP	0.07 ± 0.01	0.157	0.18 ± 0.02	0.009*
PAC → aSTG	0.20 ± 0.02	0.012*	0.36 ± 0.02	<0.001*
PAC → pMTG	0.41 ± 0.02	<0.001*	0.57 ± 0.07	0.035
pMTG → IFGop	−0.06 ± 0.04	0.384	−0.03 ± 0.03	0.412
pMTG → SMA	−0.23 ± 0.02	0.022*	0.56 ± 0.07	0.045
pMTG → PAC	0.51 ± 0.09	0.108	0.55 ± 0.12	0.160
Auditory input				
Connections	Strength (Hz)	<i>p</i> -value		
Input → PAC	0.43 ± 0.03	<0.001		

Asterisks denote FDR *p* < 0.05 corrected.

hypothesized endogenous and modulatory connections and survived the *post hoc* Bayesian selection with an overwhelming posterior probability *p* = 0.98, which also confirmed the rationality of the full model specified according to the dual-stream model. The significant modulation induced by EPs and UPs are shown separately in Figure 6, *B* and *C* (see details in Table 3). For both EPs and UPs, two information streams come from PAC and are directed toward both the anterior and posterior temporal regions. In the ventral stream, the external input of UPs caused, not only positive modulation from IFGtr to aSTG, but also reciprocal and indirect positive modulation between IFGtr and aSTG via STP, whereas EPs elicited reciprocal negative modulation between IFGtr and aSTG. In the framework of DCM, a positive modulation from node A to node B suggests that, under the effect of this modulatory input, node A causes an increase in the rate of change at node B and vice versa (Friston et al., 2003). Therefore, the positive modulation from the left IFGtr to aSTG agrees well with their increased activations shown by the univariate analysis in the UPs > EPs contrast. In the dorsal stream, both EPs and UPs elicited feedback modulation from IFGop to pMTG. In re-

sponse to EPs, SMA received positive modulation from IFGop and negative modulation from pMTG, whereas the feedforward flow from pMTG to IFGop via SMA was highlighted when expectations were violated. For cross-stream integration, reciprocal positive modulation between IFGtr and IFGop were found in EPs, but only positive modulation from IFGtr to IFGop was found in response to UPs.

Discussion

This study investigated how the human brain achieves sound-to-meaning mapping in a predictive manner within the framework of the dual-stream model of speech processing (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). We found that the left aSTG showed significantly stronger activation in response to the unexpected speech (relative to the expected speech) when expectations about the upcoming words were violated. A recent study on temporal predictive coding for single spoken words revealed that the difference between lexical predictions and heard speech is mainly represented in the left aSTG (Gagnepain et al., 2012). Multimodal neuroimaging studies using a semantic priming paradigm have also found that the processing of higher (vs lower) predictive validity for semantically predictable words is facilitated in the left aSTG (Lau et al., 2013; Lau et al., 2016). These findings consistently suggest a critical role of the aSTG in the phonological–semantic prediction of spoken words. Interestingly, our MPPA results show that the local activity patterns elicited by both EPs and UPs in the aSTG could be distinguished from those elicited by the TPs, which indicates that information encoded in this region is compatible for these two types of speech. We also found that the stronger functional connectivity between the left IFGtr and aSTG predicted better dictation performance on word identification when expectations were violated (Fig. 5*A,B*). Together, these findings suggest the involvement of the left aSTG in supporting and facilitating rapid sound-to-meaning mapping with top-down constraints.

In addition, EPs and UPs elicited similar activations in the left STP, which is more anterior than the aSTG. This activated cluster in the left STP also includes a large portion of the anterior STS, the activation of which has been consistently reported in studies comparing intelligible speech with complex acoustical match (Scott et al., 2000; Crinion et al., 2003; Thierry et al., 2003; Scott et al., 2006; Leff et al., 2008; Friederici et al., 2010). It has been proposed that the left STP is associated with the primitive combination that underlies semantic and syntactic structure building during the processing of phrases and sentences (Friederici, 2011; Poeppel, 2014). Although lacking a clear meaning as a whole phrase, unexpected speech might still require the recombination of words in the left STP during top-down processing. Therefore, the left STP may play a role in combining words into higher-level structures in an automatic manner even

when predictions are not met. The anterior temporal activations revealed in this study are also consistent with the finding that the recognition of auditory objects of different lengths (e.g., phonemes, syllables, words, and phrases) is hierarchically organized along the superior temporal lobe, with the processing of words and phrases extending from the left aSTG to the anterior STS (DeWitt and Rauschecker, 2012).

The classical language area in the left IFG (Broca's area) is commonly considered to operate at a higher hierarchical level in speech processing (Poldrack et al., 1999; Friederici, 2011; Uddén and Bahlmann, 2012) and thus has been hypothesized to be engaged in the top-down modulation that is crucial for achieving rapid sound-to-meaning mapping. Here, we found that the ventral part of the left IFG (i.e., the IFGtr or BA45) exhibited stronger activation when expectations were violated, which agrees with the previous finding that the ventral IFG is essential for multimodal semantic integration and retrieval (Dapretto and Bookheimer, 1999; Price, 2012), especially when semantic ambiguity emerges (Rodd et al., 2005). A recent electrocorticography study also found that frontal high-gamma band activity signals unpredictable deviants selectively, which indicates that the frontal cortex tracks the expected input and shows a response when predictions are violated (Dürschmid et al., 2016). Moreover, this area is also related to the convergence of current category-invariant knowledge that could further interact with lower-level representation (Rauschecker and Scott, 2009; Sohoglu et al., 2012). Importantly, functional connectivity between the left IFG and cortical regions within the temporal cortex was enhanced during the processing of UPs, especially within the IFGtr and the anterior temporal regions (i.e., aSTG and STP). This may be due to the increased demands of dealing with violations of prior expectations on the forthcoming words (Weber et al., 2016). Furthermore, the cortical dynamics observed indicate enhanced modulation on IFGtr-to-aSTG, aSTG-to-STP, and reciprocal IFGtr-to-STP connections in response to unexpected speech (Fig. 6C). The feedback connections, which convey prior expectations in hierarchical models (Friston, 2005), were enhanced from the IFGtr to both the aSTG and STP. This is likely due to the additional requirements of top-down constraints to help determine the word identity or the lack of higher-level information to accomplish semantic integration and retrieval for UPs. Consistently, we found that the individual subjects' functional connectivity of IFGtr-aSTG and aSTG-STP strongly predicted their performance on word recognition for the UPs (Fig. 5A). Compared with the functional connectivity for EPs, the more the functional connectivity between the IFGtr and aSTG was increased by the UPs, the better the word recognition performance was for these phrases (Fig. 5B). In addition, the extra requirements caused by violated expectations were also engaged by bottom-up processing with enhanced forward connections (i.e., STP-to-IFGtr and aSTG-to-STP; Fig. 6C). In contrast, when the expectation was fulfilled, as was the case for the normal idiomatic phrases, word recognition was easily achieved with much higher accuracy and the specific semantic meaning was accessed successfully, which was reflected by the reciprocal reduced modulations between the left IFGtr and aSTG (Fig. 6B).

Interestingly, we also found that the dorsal stream underlying auditory-motor mapping involved and depended on contextual information. Although the left pMTG showed stronger activation for both expected and unexpected speech, its local activity pattern was significantly distinctive for each type of auditory stimuli (i.e., UPs, EPs, and TP). In speech processing, the posterior portion of the STG is commonly regarded as being involved at a lower hierarchical level and previous studies have found that it is involved in acoustic-phonetic processing to support speech perception

(Okada et al., 2010; Evans et al., 2014). Our pattern analysis results suggest that the pMTG might also encode information at a lower level for further top-down modulation of speech processing. In the higher hierarchical level within the dorsal stream, the dorsal part of the IFG (IFGop, BA44) has been suggested to address the increased demands induced by mismatches between predictions and what is heard through the selection of the forthcoming lexical candidates (Lau et al., 2008; Price, 2010), which could further modulate the bottom-up processing initiated in the posterior temporal region. Consistent with this perspective, we identified an enhanced feedback connection from the left IFGop to pMTG and an enhanced forward connection from the pMTG to IFGop via SMA during the processing of unexpected speech, which further suggests a strong interaction between the selection of context-dependent action programs (in the prefrontal and premotor cortex) and lower-level information to alleviate the difficulty encountered in the articulatory coding (Rauschecker and Scott, 2009) of novel combinations in UPs.

Recently, the prefrontal cortex, especially the IFG, was generally proposed as a region subserving the integration of dorsal and ventral streams in speech processing (Bornkessel-Schlesewsky et al., 2015). We found that, during the processing of UPs, the left ventral and dorsal IFG (i.e., IFGtr and IFGop) showed greater connections with the ventral and dorsal streams, respectively, and we identified strengthened functional connectivity between the left IFGtr and IFGop. This finding indicates that the IFG is engaged for greater cross-stream integration during the processing of speech that violates expectations. Moreover, enhanced information flow from IFGtr to IFGop was discovered for both EPs and UPs, which may be attributed to a transformation from category-invariant representations into motor-articulatory representations for intelligible speech (Rauschecker and Scott, 2009). A model of hierarchical control in the Broca's regions also suggests that the IFGtr is involved in organizing superordinate actions through top-down interactions that initiate and terminate the successive selections of components in the IFGop (Koechlin and Jubault, 2006). However, future work involving the explicit manipulation of articulation will be required to depict fully the integrations in such a cross-stream interaction for processing expected and unexpected speech.

In conclusion, our research addressed comprehensively how prior expectations about words influence sound-to-meaning mapping at the neuroanatomical level. It highlights the role of the anterior frontal-temporal network consisting of the left aSTG, STP, and IFG in both top-down and bottom-up modulations and the critical role of the IFG in the cross-stream interaction of the ventral and dorsal pathways. These results suggest that the human brain relies on adjacent cortical areas and their interconnections for efficient back-and-forth processing of local and contextual information, which facilitates daily communication via spoken language.

References

- Abrams DA, Ryali S, Chen T, Balaban E, Levitin DJ, Menon V (2013) Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's Areas. *Cereb Cortex* 23:1703–1714. [CrossRef Medline](#)
- Boets B, Op de Beeck HP, Vandermosten M, Scott SK, Gillebert CR, Mantini D, Bulthé J, Sunaert S, Wouters J, Ghesquière P (2013) Intact but less accessible phonetic representations in adults with dyslexia. *Science* 342:1251–1254. [CrossRef Medline](#)
- Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP (2015) Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn Sci* 19:142–150. [CrossRef Medline](#)
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27.
- Crinion JT, Lambon-Ralph MA, Warburton EA, Howard D, Wise RJ (2003)

- Temporal lobe regions engaged during normal speech comprehension. *Brain* 126:1193–1201. [CrossRef Medline](#)
- Dapretto M, Bookheimer SY (1999) Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24:427–432. [CrossRef Medline](#)
- DeLong KA, Urbach TP, Kutas M (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci* 8:1117–1121. [CrossRef Medline](#)
- DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci U S A* 109:E505–E514. [CrossRef Medline](#)
- Dürschmid S, Edwards E, Reichert C, Dewar C, Hinrichs H, Heinze HJ, Kirsch HE, Dalal SS, Deouell LY, Knight RT (2016) Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proc Natl Acad Sci U S A* 113:6755–6760. [CrossRef Medline](#)
- Evans S, Kyong JS, Rosen S, Golestani N, Warren JE, McGettigan C, Mourão-Miranda J, Wise RJ, Scott SK (2014) The pathways for intelligible speech: Multivariate and univariate perspectives. *Cereb Cortex* 24:2350–2361. [Medline](#)
- Friederici AD (2011) The brain basis of language processing: From structure to function. *Physiol Rev* 91:1357–1392. [CrossRef Medline](#)
- Friederici AD, Kotz SA, Scott SK, Obleser J (2010) Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp* 31:448–457. [CrossRef Medline](#)
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836. [CrossRef Medline](#)
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138. [CrossRef Medline](#)
- Friston K, Penny W (2011) Post hoc Bayesian model selection. *Neuroimage* 56:2089–2099. [CrossRef Medline](#)
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229. [CrossRef Medline](#)
- Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19:1273–1302. [CrossRef Medline](#)
- Friston KJ, Li B, Daunizeau J, Stephan KE (2011) Network discovery with DCM. *Neuroimage* 56:1202–1221. [CrossRef Medline](#)
- Gagnepain P, Henson RN, Davis MH (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr Biol* 22:615–621. [CrossRef Medline](#)
- Heim S, Eickhoff SB, Ischebeck AK, Friederici AD, Stephan KE, Amunts K (2009) Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. *Hum Brain Mapp* 30:392–402. [CrossRef Medline](#)
- Herrmann B, Maess B, Hahne A, Schröger E, Friederici AD (2011) Syntactic and auditory spatial processing in the human temporal cortex: An MEG study. *Neuroimage* 57:624–633. [CrossRef Medline](#)
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402. [CrossRef Medline](#)
- Kasess CH, Stephan KE, Weissenbacher A, Pezawas L, Moser E, Windischberger C (2010) Multi-subject analyses with dynamic causal modeling. *Neuroimage* 49:3065–3074. [CrossRef Medline](#)
- Koechlin E, Jubault T (2006) Broca's area and the hierarchical organization of human behavior. *Neuron* 50:963–974. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci* 9:920–933. [CrossRef Medline](#)
- Lau EF, Gramfort A, Hämäläinen MS, Kuperberg GR (2013) Automatic semantic facilitation in anterior temporal cortex revealed through multi-modal neuroimaging. *J Neurosci* 33:17174–17181. [CrossRef Medline](#)
- Lau EF, Weber K, Gramfort A, Hämäläinen MS, Kuperberg GR (2016) Spatiotemporal signatures of lexical-semantic prediction. *Cereb Cortex* 26:1377–1387. [CrossRef Medline](#)
- Leff AP, Schofield TM, Stephan KE, Crinion JT, Friston KJ, Price CJ (2008) The cortical dynamics of intelligible speech. *J Neurosci* 28:13209–13215. [CrossRef Medline](#)
- MacGregor LJ, Pulvermüller F, van Casteren M, Shtyrov Y (2012) Ultra-rapid access to words in the brain. *Nat Commun* 3:711. [CrossRef Medline](#)
- Marslen-Wilson WD, Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10:29–63. [CrossRef](#)
- McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK (2012) An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *J Cogn Neurosci* 24:636–652. [CrossRef Medline](#)
- Obleser J, Kotz SA (2010) Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex* 20:633–640. [CrossRef Medline](#)
- Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G (2010) Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex* 20:2486–2495. [CrossRef Medline](#)
- Pereira F, Botvinick M (2011) Information mapping with pattern classifiers: a comparative study. *Neuroimage* 56:476–496. [CrossRef Medline](#)
- Pernet CR (2014) Misconceptions in the use of the general linear model applied to functional MRI: a tutorial for junior neuro-imagers. *Front Neurosci* 8:1–12. [CrossRef Medline](#)
- Poeppel D (2014) The neuroanatomic and neurophysiological infrastructure for speech and language. *Curr Opin Neurobiol* 28:142–149. [CrossRef Medline](#)
- Poldrack RA, Farah MJ (2015) Progress and challenges in probing the human brain. *Nature* 526:371–379. [CrossRef Medline](#)
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15–35. [CrossRef Medline](#)
- Price CJ (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann NY Acad Sci* 1191:62–88. [CrossRef Medline](#)
- 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* 62:816–847. [CrossRef Medline](#)
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87. [CrossRef Medline](#)
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724. [CrossRef Medline](#)
- Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15:1261–1269. [Medline](#)
- Rosa MJ, Friston K, Penny W (2012) Post-hoc selection of dynamic causal models. *J Neurosci Methods* 208:66–78. [CrossRef Medline](#)
- Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* 4:Article32. [Medline](#)
- Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406. [CrossRef Medline](#)
- Scott SK, Rosen S, Lang H, Wise RJ (2006) Neural correlates of intelligibility in speech investigated with noise vocoded speech-A positron emission tomography study. *J Acoust Soc Am* 120:1075–1083. [CrossRef Medline](#)
- Seghier ML, Friston KJ (2013) Network discovery with large DCMs. *Neuroimage* 68:181–191. [CrossRef Medline](#)
- Serences JT, Boynton GM (2007) The representation of behavioral choice for motion in human visual cortex. *J Neurosci* 27:12893–12899. [CrossRef Medline](#)
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. *J Neurosci* 32:8443–8453. [CrossRef Medline](#)
- Stelzer J, Chen Y, Turner R (2013) Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *Neuroimage* 65:69–82. [CrossRef Medline](#)
- Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38:387–401. [CrossRef Medline](#)
- Thierry G, Giraud AL, Price C (2003) Hemispheric dissociation in access to the human semantic system. *Neuron* 38:499–506. [CrossRef Medline](#)
- Tuenerhoff J, Noppeney U (2016) When sentences live up to your expectations. *Neuroimage* 124:641–653. [CrossRef Medline](#)
- Udden J, Bahlmann J (2012) A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philos Trans R Soc Lond B Biol Sci* 367:2023–2032. [CrossRef Medline](#)
- Weber K, Lau EF, Stillerman B, Kuperberg GR (2016) The yin and the yang of prediction: An fMRI study of semantic predictive processing. *PLoS One* 11:e0148637. [CrossRef Medline](#)