Cross-language differences in the brain network subserving intelligible speech

Jianqiao Ge[a,b,1,2], Gang Peng[c,d,1,2], Bingjiang Lyu[a,b], Yi Wang[a,b], Yan Zhuo[e], Zhendong Niu[f], Li Hai Tan[g,h,i,2], Alexander P. Leff[j], and Jia-Hong Gao[a,b,k,2]

[a]Center for MRI Research, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China; [b]Beijing City Key Lab for Medical Physics and Engineering, Institution of Heavy Ion Physics, School of Physics, Peking University, Beijing 100871, China; [c]Joint Research Centre for Language and Human Complexity, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong, China; [d]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; [e]State Key Laboratory of Brain and Cognitive Science, Chinese Academy of Sciences, Beijing 100101, China; [f]School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China; [g]State Key Laboratory of Brain and Cognitive Sciences and [h]Department of Biomedical Engineering, School of Medicine, Shenzhen University, Shenzhen 518060, China; [i]Guangdong Key Laboratory of Biomedical Information Detection and Ultrasound Imaging, Shenzhen 518060 China; [j]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom; and [k]McGovern Institute for Brain Research, Peking University, Beijing 100871, China

Edited by Robert Desimone, Massachusetts Institute of Technology, Cambridge, MA, and approved January 22, 2015 (received for review August 21, 2014)

How is language processed in the brain by native speakers of different languages? Is there one brain system for all languages or are different languages subserved by different brain systems? The first view emphasizes commonality, whereas the second emphasizes specificity. We investigated the cortical dynamics involved in processing two very diverse languages: a tonal language (Chinese) and a nontonal language (English). We used functional MRI and dynamic causal modeling analysis to compute and compare brain network models exhaustively with all possible connections among nodes of language regions in temporal and frontal cortex and found that the information flow from the posterior to anterior portions of the temporal cortex was commonly shared by Chinese and English speakers during speech comprehension, whereas the inferior frontal gyrus received neural signals from the left posterior portion of the temporal cortex in English speakers and from the bilateral anterior portion of the temporal cortex in Chinese speakers. Our results revealed that, although speech processing is largely carried out in the common left hemisphere classical language areas (Broca's and Wernicke's areas) and anterior temporal cortex, speech comprehension across different language groups depends on how these brain regions interact with each other. Moreover, the right anterior temporal cortex, which is crucial for tone processing, is equally important as its left homolog, the left anterior temporal cortex, in modulating the cortical dynamics in tone language comprehension. The current study pinpoints the importance of the bilateral anterior temporal cortex in language comprehension that is downplayed or even ignored by popular contemporary models of speech comprehension.

speech perception | tonal language | functional MRI | cortical dynamics

The brain of a newborn discriminates the various phonemic contrasts used in different languages (1) by recruiting distributed cortical regions (2); by 6–10 mo, it is preferentially tuned to the phonemes in native speech that they have been exposed to (3, 4). In adult humans, the key neural nodes that subserve speech comprehension are located in the superior temporal cortex (5, 6) and the inferior frontal cortex (7). Do these regions interact in different ways depending on the type of language that is being processed? Little is known about how information flows among these critical language nodes in native speakers of different languages.

As one of the unique capacities of the human brain (8), the nature of compositional languages and their neural mechanisms have been the interests of scientific research for decades. There are more than 7,000 different spoken languages in the world today used for communication. By exploring the brain networks subserving universal properties across languages and specific differences within different languages, such research helps address the essential questions in neurolinguistics such as the constitution of knowledge of language, as well as how it is acquired (9). Although traditional universal grammar theory argues that functional components of linguistic ability are manifest without being taught (10), recent connectionist theory within a neural network approach emphasizes interactions among primary systems of neuronal processing units that support language acquisition and use, where the weights of connections among these units are gradually changed during learning and thus highly constrained by the unique feature of a given language (9, 11, 12). Related with connectionist views, the dual pathway model of language based on evidence from neuroimaging and anatomical studies was also framed to interpret the neural basis of language comprehension and production (13, 14), especially in speech comprehension.

Intelligible speech is processed hierarchically in human neocortex, with the anterior temporal (13, 15, 16) and frontal cortices (Broca's area) operating at a higher hierarchical level than the posterior region centered on superior temporal sulcus/gyrus (pSTS/pSTG, core of Wernicke's area), which itself receives inputs from primary and secondary auditory cortices (17). In the dual ventral-dorsal pathway model, the dorsal-stream pathway of speech processing assumes primary processing begins in the

Significance

Language processing is generally left hemisphere dominant. However, whether the interactions among the typical left hemispheric language regions differ across different languages is largely unknown. An ideal method to address this question is modeling cortical interactions across language groups, but this is usually constrained by the model space with the prior hypothesis due to massive computation demands. With cloud-computing, we used functional MRI dynamic causal modeling analysis to compare more than 4,000 models of cortical dynamics among critical language regions in the temporal and frontal cortex, established the bias-free information flow maps that were shared or specific for processing intelligible speech in Chinese and English, and revealed the neural dynamics between the left and right hemispheres in Chinese speech comprehension.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]J.G. and G.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jgao@pku.edu.cn, gejq@pku.edu.cn, tanlh@szu.edu.cn, or gpengjack@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416000112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1416000112

posterior region of the temporal cortex and is then sent through the dorsal part to the temporal-parietal cortex before finally reaching the frontal cortex for a sound-motor projection. The ventral-stream pathway assumes processing starts on the ventral side of the temporal lobe and then proceeds to the anterior regions, before reaching the frontal cortex for a sound-meaning mapping (7, 18).

The cortical regions that process speech are likely to be common across languages, but how these regions interact with each other in the cortical pathway may depend on the distinctive phonetic-linguistic characteristics in different languages. Based on the connectionist approach, previous cross-language studies investigated the behavioral consequence of processing different languages with a distinction such as pitch accent processing (19). The current study aims to explore the dynamic neural networks of processing intelligible speech in two different languages with a featured phonological-semantic variant: Mandarin Chinese and English. These two languages are the most widely spoken languages in the world, but differ in several aspects such as the use of lexical tones. In tonal languages like Mandarin Chinese, suprasegmental features, i.e., different pitch patterns, serve to distinguish lexical meaning, whereas in nontonal languages, pitch changes are less complex and do not convey lexical information. Except the lexical tone, Mandarin Chinese includes more homophones than English, which also makes the sound-meaning mapping in Mandarin dependent more on tone and context information during speech and thus may place higher demands on the ventral pathway of the related neural network. To test this hypothesis, we examined the cortical dynamics underlying speech comprehension for two groups of native speakers of English and Mandarin languages.

We used functional MRI (fMRI) and dynamic causal modeling (DCM) (20) to first investigate the cortical dynamics among the left posterior region of superior temporal gyrus (pSTG), anterior region of superior temporal gyrus (aSTG), and inferior frontal gyrus (IFG) of native speakers in Chinese (a tonal language) compared with English (a nontonal language) with an identical experimental design. Thirty native Chinese speakers and 26 native English speakers with matched age, sex, and handedness (all right-handed) were scanned while presented with intelligible and unintelligible speech of their native languages (either Mandarin Chinese or English) in blocks, spoken by a male and a female. Subjects were instructed only to judge the gender of the speakers. The data of native English speakers were reanalyzed from a previous study (21). The brain activation of the intelligibility effect and the effective connectivity among the three left hemisphere brain regions were analyzed for both language groups under identical procedures and then put together for comparison.

Results

In both groups, the contrast of intelligible > unintelligible speech revealed significant neural activities in the left anterior temporal lobe, left posterior temporal gyrus, supplementary motor area, postcentral gyrus, and pars triangularis of the left inferior frontal gyrus (Fig. 1*A* and Table S1).

To establish the basic neural dynamic network of processing intelligible speech in Chinese and English, we first constructed dynamic models that consisted of the three shared left hemisphere brain regions that were engaged in processing speech in both languages: the left aSTG (region A), the left pSTG (region P), and the left IFG (region F). We computed an exhaustive series of models, varying input site (seven families) with all possible patterns of connectivity among the three nodes (63 models per family), which generated a total of 441 alternative models for each subject. These models were estimated, and the evidence for each was compared using a family-level random effect Bayesian model selection analysis (22). A Bayesian model average (BMA) analysis was then performed to provide average connectivity values for each connection across all possible

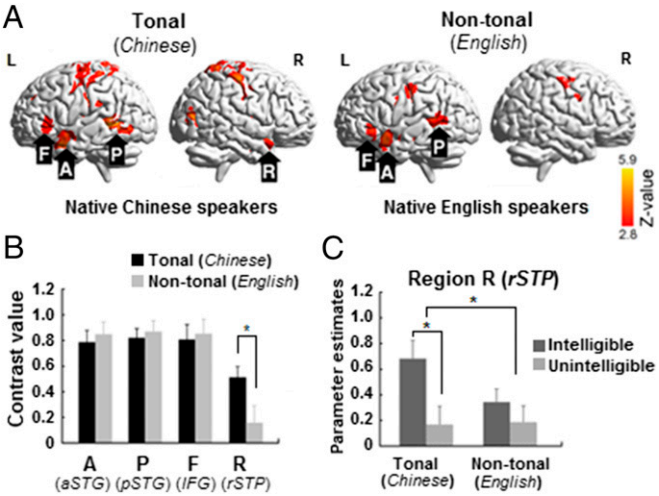

Fig. 1. Brain activations during the processing of intelligible speech. (*A*) Activations showing in the effect of intelligibility (intelligible speech > unintelligible speech, with threshold $P < 0.005$ voxel-level uncorrected and minimum cluster size 50 voxels) in native English speakers and native Chinese speakers. The brain areas that entered into DCM analysis were defined with a threshold of $P < 0.05$ FWE corrected, which are labeled with black arrows (left hemisphere: F, inferior frontal gyrus; A, anterior superior temporal gyrus; P, posterior middle/superior temporal gyrus; right hemisphere: R, superior temporal pole/gyrus). More information regarding the brain activations of these regions is detailed in Table S1. (*B*) ROI analysis for all four brain regions of interests, showing that the brain activity intensity in left pSTG ($t = -0.423$, $P = 0.67$), aSTG ($t = -0.462$, $P = 0.65$), and IFG ($t = -0.275$, $P = 0.78$) is compatible between the Chinese and English groups. (*C*) ROI analysis comparing the parameter estimates for signal intensity in region R (rSTP) between intelligible and unintelligible speech in native Chinese and native English speakers, showing a significant interaction effect ($P < 0.01$) for brain activation between the intelligibility of the speech and language group in rSTP.

models in the model space for each subject. These values were entered into both within- and between-group analysis for individual connection using one- and two-sample *t* tests with false discovery rate (FDR) correction. The results showed that for both groups the auditory signals entered the neural network through the pSTG node, which is, by definition, the lowest of the three nodes in the cortical hierarchy. In terms of interregional connections, hearing intelligible speech increased the strength of the ventral forward connection, pSTG-to-aSTG, in both groups. There were, however, clear group differences for other connections; specifically, the English speakers had a significantly stronger dorsal forward connection from the pSTG to IFG, whereas in the Chinese speakers, the two connections emanating from the aSTG (a backward connection to the pSTG and a lateral connection to the IFG) were stronger (Fig. 2*A* and Table 1).

In addition to the three shared brain regions in the left hemisphere, the Chinese speakers had an additional activation in the right anterior temporal pole ($F_{1,51} = 8.141$, $P = 0.006$; Fig. 1 *B* and *C* and Fig. S1) during the processing of intelligible speech, consistent with previous findings that the anterior region of the right temporal cortex is functionally linked with pitch and tone processing (23). To investigate the comprehensive neural dynamics for the tonal language, we carried out a second analysis of the Chinese-only data that included this fourth region (right superior temporal pole/anterior region of superior temporal gyrus, rSTP/right aSTG). BMA analysis of an exhaustive set of 4,095 alternative models was conducted (input into the left pSTG, 12 connections systematically varied across models). The connections significantly modulated within the left hemisphere were revealed to be the same as those identified in the three-region analysis (Fig. 2*B*). This analysis also identified three interhemispheric connections significantly

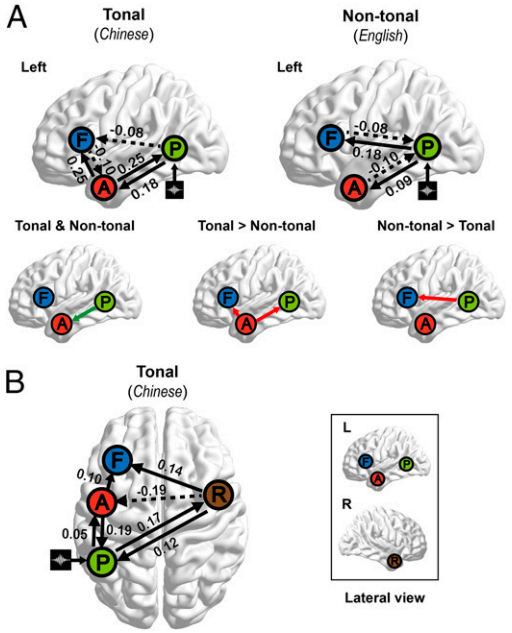

Fig. 2. Results of the DCM BMA analysis. (*A*) Three-region models (pSTG-aSTG-IFG, i.e., P-A-F) of processing intelligible speech in tonal language (i.e., Chinese) and nontonal language (i.e., English) are shown in *Left* and *Right Upper* panels, respectively. More information regarding the parameter estimates of the three-region models is detailed in Table 1. Green arrow is for the connection significantly modulated by intelligible speech in both languages (*Lower Left*); the red arrows for connections significantly activated in one language compared with the other (*Lower Center and Right*). (*B*) Four-region model (pSTG-aSTG-IFG-rSTP, i.e., P-A-F-R) of processing intelligible speech for Mandarin Chinese speakers. More information regarding the parameter estimates of the four-region model is detailed in Table S2. The auditory stimuli entered the neural system via pSTG (P) in all models, and the arrowed lines display the connections showed significantly enhanced (solid) or decreased (dashed) modulation of speech intelligibility with average modulatory parameter estimates (s^{-1}) shown alongside ($P < 0.05$, FDR corrected).

modulated by intelligibility: the bidirectional right aSTG-to-left pSTG connections (the same connectivity pattern as with the left aSTG-pSTG), and the right aSTG-to-left IFG connection (Table S2). Moreover, the modulation strength of intelligibility on the connection of the left aSTG-to-left pSTG is positively correlated with the strength of the right aSTG-to-left pSTG connection (Fig. 3*A* and Fig. S2).

Discussion

This study found that, during the processing of intelligible speech, the three regions in the left hemisphere, inferior frontal gyrus (Broca's area), posterior temporal gyrus (Wernicke's area), and anterior temporal gyrus, are shared by two language groups (both tonal and nontonal), whereas the interactions among those regions depend on the language. Processing intelligible speech in a tonal language engages the bilateral anterior temporal lobes, and its connections with classical language areas are much stronger than in a nontonal language, whereas the connection between the classical language areas in left hemisphere (from the posterior temporal lobe to inferior frontal lobe) is much stronger in a nontonal (English) than in a tonal language (Chinese).

Importantly, both forward and backward connections from the left and right aSTGs to the left pSTG are involved in the tonal-language network. In tonal languages such as Chinese, suprasegmental features (e.g., pitch changes) are used to signify the meaning of a word, resulting in much larger numbers of homophones in the daily vocabulary (24). The aSTG is considered to be a "semantic hub" that it is critical in supporting language function (25). The underlying cortical pathways of speech processing based on these ventral connections are especially important to accomplish a more complicated sound-meaning mapping in a tonal language. We identified increased backward connections, which convey information about prior expectations in hierarchical processing models (26), probably due to the lack of suprasegmental phonological information initially or the lack of the sentence structure to help resolve word identity because the auditory word pairs were heard in isolation. The involvement of both temporal poles in this task may be due to either these increased task demands (given that there are more homophones in Chinese than English), because the right hemisphere is more involved in processing pitch information in tonal languages (23, 27), or both. The increased backward modulation from the left anterior to the posterior parts of the temporal lobe was revealed to be significantly correlated with an increased modulation on the connection from the right anterior part to the left posterior part of the temporal lobe, suggesting that the top-down modulation for further semantic processing on the ventral pathway in a tonal language is supported by integrated processing based on full phonological information (including lexical tones) from bilateral temporal lobes.

Moreover, the stronger forward connections between the anterior temporal poles and Broca's area may be due to further semantic processing that is included in word identification through phonological information in Chinese. Subjects with greater difficulties in identifying the word (Fig. 3*B*) but intact performance in identifying the pronunciation (Fig. S3) showed greater modulated connectivity on the bilateral connections from the aSTG to Broca's area (i.e., the IFG). This preliminary result for Chinese speakers suggests an integrated forward processing of mapping the phonological information to the semantic-related representation from both hemispheres in this tonal language.

The only connection that was stronger in the English speakers was the forward connection from the pSTG to IFG. This result likely represents a greater reliance of nontonal languages on the dorsal stream, which is implicated in tasks that stress phonological (elemental speech sound) processing of speech (28), where initial phonological features are informative enough to identify words in a nontonal language.

The right temporal lobe activation has been widely observed during the phonological processing for Chinese subjects (29, 30), whereas the right anterior temporal pole was anatomically related to speakers of Chinese (23). The activation of the rSTP in Chinese data of our research was consistent with these previous findings. The fact that the English subjects showed no activation in the right anterior temporal pole was probably because of the absence of explicit demands on pitch-related processing in the experiment. In tasks involving pitch or intonation processing for

Table 1. Parameter estimates (s^{-1}) of modulation of speech intelligibility on connections in the three-region models (P-A-F) of the Chinese and English groups

	Tonal (Chinese)		Nontonal (English)		Chinese vs.
Connection	Mean (SEM)	*t*	Mean (SEM)	*t*	English (*t*)
From pSTG to					
aSTG	0.182 (0.021)	8.94*	0.093 (0.026)	3.63*	0.84
IFG	−0.083 (0.019)	−4.33*	0.177 (0.026)	6.78*	−2.66*
From aSTG to					
pSTG	0.254 (0.026)	9.79*	−0.097 (0.027)	−3.59*	2.18†
IFG	0.247 (0.018)	13.62*	−0.025 (0.025)	−1.01	2.76*
From IFG to					
pSTG	0.026 (0.023)	1.13	−0.078 (0.027)	−2.87*	1.57
aSTG	−0.100 (0.017)	−5.90*	0.008 (0.026)	0.31	−0.96

*$P < 0.05$, FDR corrected.
†$P < 0.05$, uncorrected.

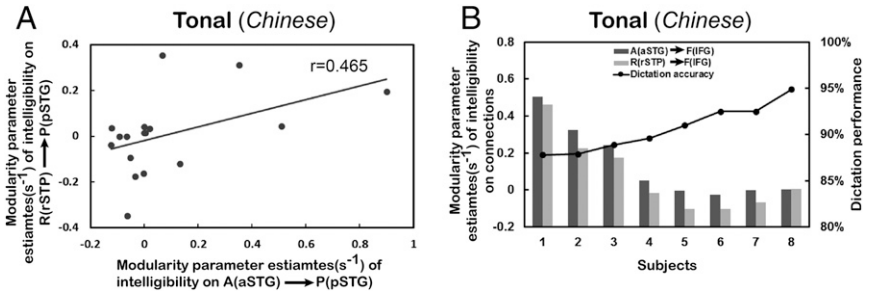

Fig. 3. Correlations of modulation strength of brain connections in tonal-language (Chinese). (*A*) Positive correlation was found between the modulation strengths on connections of left aSTG-to-pSTG and right aSTP-to-left pSTG. ($r = 0.896$, $P < 0.001$, Fig. S2; after one outlier removed, $r = 0.465$, $P = 0.06$) across individuals in Chinese group; no significant correlation of modulation strength was found in the cortical dynamics of English (all $P > 0.5$). (*B*) Modulation strength on connections of both left and right aSTG/STP to left inferior frontal gyrus predicted the individual subject's behavior performance of idiom dictations on the percentage proportion of correctly identified word (left aSTG-to-IFG: three-region DCM analysis $r = -0.811$, $P = 0.015$ shown in figure, four-region DCM analysis $r = -0.703$, $P = 0.052$; right aSTG/STP-to-IFG: $r = -0.719$, $P = 0.045$).

English, significant activations could be detected in the right superior temporal area (31). In the current study, however, subjects were asked to make a gender judgment on the auditory stimuli without any explicit requirement to understand the speech or make a tone judgment, which thus required minimal demands of pitch-related processing circuitry. Therefore, the results mostly suggested an automatic cortical engagement, where the right temporal lobe underlying the processing of tonal information is involved in cortical dynamics for perceiving intelligible speech in a tonal language.

Our research has revealed for the first time, to our knowledge, that, particularly in tonal languages, classical left hemisphere language areas such as Wernicke's and Broca's areas interact with the semantic system in anterior temporal lobes from both hemispheres when perceiving intelligible speech. At least two popular contemporary models of speech comprehension either downplay the importance (32) of these regions or ignore them altogether (7). The dynamic maps we describe here only reached to the subsentence level of speech comprehension, and further investigation is required to consider types of sentence processing most frequently encountered in daily communication. As for the comparison between different language groups, the cross-center MRI data acquisition may induce potential confounding on images such as distortion differences on the temporal lobe. Therefore, future cross-language comparison research on brain networks should consider these factors with extensive control on the data acquisition.

Regardless of these limitations, our results suggest both language-common and language-specific cortical dynamics coexist for speech comprehension and emphasize the importance of the bilateral anterior temporal lobes and their connections with Wernicke's and Broca's areas in speech perception, particularly for tonal languages such as Mandarin Chinese.

Materials and Methods

Subjects. Thirty native Chinese speakers and 26 native English speakers participated in the current research. The Chinese and English groups were matched on subject sex, age, and handedness. Native Chinese speakers participated in this study as paid volunteers (15 males and 15 females; aged between 21 and 28 y; mean age, 24.2 y). Native English speakers participated in an experiment that was reported in a previous study by Leff et al. (21), and the brain imaging data were reanalyzed in the current study. All participants were right-handed, with normal hearing and normal or corrected-to-normal vision, with Mandarin Chinese or English as their first language, and had no neurological or psychiatric history. Written informed consent was obtained from each participant before scanning, and the study was conducted under the approval of the Institutional Review Board of Beijing MRI Center for Brain Research.

Stimuli. The experimental paradigm was adopted from a previous study about the cortical dynamics of intelligible English speech (21), whereas both intelligible and unintelligible auditory stimuli were presented to subjects to make a gender judgment of the speakers. Half of the intelligible stimuli in English were idiomatic word pairs such as "cloud nine," and the other half were reordered word pairs of idioms (e.g., "mint nine"). To create their unintelligible counterparts, the stimuli of intelligible stimuli were time-reversed because this method removed the intelligibility of the forward speech while preserving the acoustic and voice identity information (5, 21). The Chinese stimuli were designed with consideration of matching the acoustic and psycholinguistic characteristics of the idiomatic word pairs between Chinese and English. Words in the English stimuli were mainly disyllabic. We selected Chinese intelligible stimuli with a three-, four-, or five-syllable length that matched in duration with the English word pairs. Half of the intelligible stimuli in Chinese were idiomatic words with three to five characters such as 和事佬 (he2 shi4 lao3, means "peacemaker," the letters represent the official Romanization of standard Chinese, that is, Pinyin, whereas the number indicates the corresponding tone), 画龙点睛 (hua4 long2 dian3 jing1, means "finishing touch"), and the other half consisted of words from two unrelated idioms [e.g., 鸿门户 (hong2 men2 hu4), 恶贯好龙 (e4 guan4 hao4 long2); 恶贯好龙 was the combination reordered from the first two words of the idiom 恶贯满盈, (e4 guan4 mang3 ying2) and the last two words of the idiom 叶公好龙 (ye4 gong1 hao4 long2)]. All intelligible stimuli were recorded digitally in a soundproof studio using Adobe Audition CS4 software. A male and a female speaker (both native speakers) produced all of the stimuli twice. As for Chinese unintelligible stimuli, the unintelligible counterparts of Chinese stimuli were also time-reversed of the intelligible stimuli that removed the intelligibility but preserved the acoustic and voice identity information. There were 84 auditory stimuli (half by the male speaker and half by the female speaker) for each of the three stimulus types, and no stimulus was repeated. All stimuli were edited for quality and length and amplified so that there was no difference of loudness between speakers or between the idioms and reordered idioms.

Comparison and Psycholinguistic Analysis of Stimuli Between Groups. We first calculated the duration of all auditory stimuli in the English and Chinese groups (English: 677–1,080 ms; Chinese: 730–1,007 ms). A two-sample *t* test revealed no significant difference on the duration between English and Chinese groups. A similar procedure was also conducted to examine the stimulus loudness across languages and types of stimuli, and no significant difference was found. Both the Chinese and English idioms were phrases established by use and referred to a certain meaning. The reordered idioms were created by combining two unrelated idiom. Therefore, the reordered idioms in both English and Chinese were meaningless but still recognizable syllable by syllable.

Because there are no comprehensive databases recording key psycholinguistic variables for either English or Chinese idioms, two psycholinguists who use English or Chinese as their native language independently rated the idiomatic stimuli in their native language (either English or Chinese) on a binary scale, guided by the Cronk criteria (33), splitting the stimuli into the following four uneven groups: (*i*) high familiarity, high literalness; (*ii*) high familiarity, low literalness; (*iii*) low familiarity, high literalness; and (*iv*) low familiarity, low literalness. Percentage proportions of stimuli that fell into each group were then calculated (Chinese: English): (*i*) 35%:33%; (*ii*) 39%:31%; (*iii*) 8%:11%; and (*iv*) 18%:35%. Nonparametric statistical tests of all four groups (independent samples Mann–Whitney *U* test) revealed no significant differences across the two languages (all $P = 1.0$).

Procedure. In the experiment, subjects were required to make a gender judgment of the speaker for each stimulus, and they were not informed of the types and contents of the stimuli in advance. Therefore, the task was orthogonal to the effect of interest. The auditory stimuli were arranged in blocks by stimulus types with an alterable ratio of male and female speakers (2:5, 3:4, 4:3, or 5:2). All auditory stimuli only were presented once. There

were 12 blocks for each type of stimuli, and they were evenly distributed in four scanning sessions. Each block was preceded by a preparation cue for 3,150 ms, which was followed by seven trials. On each trial, an auditory stimulus was presented for 1,180 ms, followed by a response cue for 2,420 ms and then a fixation cross for 450 ms. The blocks were separated by a symbol "~" of 9,450 ms presented at the center of the screen. Participants pressed one of the two buttons with their right index or middle finger to indicate their gender judgment of the speaker after the response cue immediately. The order of the blocks and the assignment of response buttons were counterbalanced across participants.

MRI Data Acquisition. Scanning of native Chinese speakers was performed on a 3-T Siemens MRI scanner in our laboratory in Beijing, China. Scanning of native English speakers was performed by one of our coauthors in London, UK, and was reported in a previous study (21). In both scannings, functional images were acquired using a gradient-echo echo-planar imaging (EPI) pulse sequence (TR/TE/θ = 2.08 s/30 ms/90°, 64 × 64 × 35 matrix with 3 × 3 × 3-mm^3 spatial resolution). The visual displays were presented through a Sinorad LCD projector (Shenzhen Sinorad Medical Electronics) onto a rear-projection screen located over the subject's head, viewed with an angled mirror positioning on the head coil. The auditory stimuli were presented binaurally using a pair of home-made MRI compatible headphones that provided 25–30 dB/SPL attenuation of the scanner noise. In consideration of the scanner noise generated by EPI sequence, participants were required to adjust sound volume of the sample auditory stimuli to a clear-and-comfortable level during a pilot EPI scan before the experiment. After the volume adjustment, participants were instructed to pay attention to the experimental task. Four sessions of functional task scanning were acquired while participants performed the gender judgment of the auditory stimuli. Each session started with a blank screen for 10 s and was followed by nine blocks of auditory stimuli, which lasted 378.56 s in total. After the functional scanning, T1-weighted 3D structural images were also obtained (TR/TE = 2.6 s/3.02 ms, 1 × 1 × 1-mm^3 spatial resolution).

Behavioral Data Analysis. Response accuracy and reaction time were recorded for each type of stimuli. Because our interests in this research were focused on the processing of speech intelligibility, paired *t* tests were conducted to compare the differences in behavioral performance between intelligible speech (averaging across idioms and rearranged idioms) and unintelligible speech (time-reversed idioms). The results of behavioral data in English participant were reported in a previous paper (21). The mean response accuracy of Chinese participants for gender judgment across all auditory stimuli was 98.5%. A paired *t* test of response accuracy showed that subjects made significantly more accurate gender judgments for intelligible speech than for time-reversed speech ($t = 3.612$, $P < 0.01$; intelligible: 99.0 ± 1.8%; time-reversed: 98.0 ± 2.3%), whereas there was no significant difference in reaction time between the two speech types (overall reaction time, 1,731 ± 176 ms).

fMRI Data Analysis. Statistical parametric mapping software (SPM8; Wellcome Trust Centre for Neuroimaging) was used for imaging data processing and analysis. Native English speakers participated in the experiment in a previous study, and the fMRI data were reanalyzed in the current research with an identical procedure. The EPI images were realigned to the first scan to correct head motion. Then, the mean image produced during the process of realignment, and the realigned images were coregistered to the high-resolution T1 anatomical image. All images were spatially normalized to standard MNI space. The normalized EPI images were then spatially smoothed using an isotropic Gaussian kernel with a full-width at half maximum (FWHM) parameter of 8 mm. The functional imaging data were modeled using a boxcar function with head motion parameters as unrelated regressors. Parameter estimates for each condition (three types of stimuli) were calculated from a general linear model (GLM) based on the hemodynamic response function with overall grand mean scaling. Whole-brain statistical parametric mapping analyses were performed, and contrasts were then defined to reveal brain areas specifically involved in processing intelligible stimuli (idioms and reordered idioms) and that of unintelligible stimuli (time-reversed idioms). The *t*-contrast images were generated for comparison at each voxel. Statistical tests were first assessed in individual subjects, and random effect analyses were then conducted based on statistical parameter maps from each individual subject to allow population inference. A one-sample *t* test was applied to determine group-level activation for intelligibility effect. Moreover, parameter estimates of signal intensity for processing intelligible and unintelligible speech were extracted from regions of interests (ROIs) and compared between the Chinese group and the English group using ANOVA. The ROIs in the left anterior temporal cortex (region A), left posterior temporal cortex (region P), and left inferior frontal cortex (region F) were defined as spheres with 6 mm diameter centered at the local maxima voxel around the center of ROI landmarks (group-level peak voxel) observed in the intelligibility effect in both groups. Because no significant activation could be found in the right anterior temporal cortex for the English group even with a more liberal threshold ($P < 0.01$ voxel-level uncorrected), the landmark ROI definition of this region (R) in the English group was identical to the Chinese group.

DCM Analysis. After we identified the involvement of several brain regions in the processing of speech intelligibility, we conducted a DCM analysis (20) to examine the effective connectivity among these brain regions. In DCM analysis, differential equations, $d\mathbf{x}/dt = (\mathbf{A} + \mathbf{uB})\mathbf{x} + \mathbf{Cu}$, were used to model the cortical dynamics of the neuronal populations in brain ROIs, which describe how the current state of one neuronal populations causes dynamics in another through synaptic connections that are intrinsic and fixed and how these interactions change under the external influence of experimental manipulations or the influence of endogenous brain activity (34). Here, **x** is a vector representing the neural state of all brain regions that are in consideration, and **u** is a vector representing all external input. Matrix **A** represents the strength of fixed connection between the brain regions that are in consideration; in other words, the strength of connection when no external input exists. Because it has been shown that an anatomical connection exists between each pair of the three brain regions (7, 18), it was assumed that a reciprocal fixed connection between each pair of the brain regions existed: in other words, all parameters in matrix **A** would be set to a nonzero value (34). Matrix **B** represents the strength of modulation of the connections by external inputs, i.e., the experimental manipulation (in the current research, this was the processing of the speech varying in intelligibility). Matrix **C** represents the direct influence to the brain regions by the external inputs that were generally considered to be sensory stimuli, such as the auditory input in this experiment. Furthermore, it is also widely assumed that modulatory stimuli, such as intelligibility of speech in this experiment, can only indirectly influence brain regions, i.e., only some related parameters in matrix **B** would be nonzero, whereas all relative parameters in matrix **C** would be zero. In our model-space design, the supposition above was applied and all nonzero parameters in the matrixes were assumed Gaussian.

ROIs Selection and Time Series Extraction. In the current research, we were particularly interested in the brain regions in the temporal and frontal cortex that showed significant involvement in the processing of speech intelligibility. The coordinates of the peak voxel in the clusters identified in the group-level random effect analysis of the intelligibility effect (comparing the neural activity during listening to intelligible speech with unintelligible speech, i.e., the time-reversed speech) with $P < 0.05$ family-wise error (FWE) corrected threshold were used to serve as a landmark for the individual ROIs. For each Chinese subject, ROIs were defined as 6-mm spherical volumes centered at the peak activation voxel of intelligibility effect in the regions of our interests by searching voxels that survived at least the $P < 0.05$ threshold around the landmarks (revealed in group-level analysis) within the 8-mm radius distance and within the same anatomical regions. Because the exact location of activation varied for each subject, this procedure ensured comparability of models and the extracted time series across subjects by applying both functional and anatomical constraints (35, 36). Given these criteria, we were able to define ROIs and extract time series for a three-region model (P-A-F) in 22 of 30 Chinese subjects and for a four-region model (P-A-F-R) in 18 of 30 Chinese subjects, and the remaining subjects were excluded from the respective DCM analysis in whom the activation of at least one brain region failed to meet the criteria. An identical procedure was performed on English subjects and was reported in the previous study (21). For each ROI in an individual subject, the time series was extracted and computed as the first eigenvector across all suprathreshold voxels.

DCM Specification. We formed a specific model space that contained the whole set of alternative models that were anatomically and functionally plausible and used Bayesian methods to estimate the model parameters. In model selection, considering the large number of the models to compare, we used the family-level inference and Bayesian model averaging within families on the model space (22) instead of the "single best model" selection strategy. This procedure provided inference about model parameters that could minimize the assumption bias on the model structure.

To establish the basic neural dynamic network of processing intelligible speech in Chinese and English, we first specified the model space that consisted of three shared left hemisphere brain regions: A, P, and F. For input "auditory", treated as a sensory input, seven alternate ways of how input auditory could enter the system were contained into the model space: via pSTG only, via aSTG only, via pSTG and aSTG, via IFG only, via pSTG and IFG, via aSTG and IFG, and via all three

brain regions. Intelligibility was treated as a modulatory input; it may modulate any combination of six directed connections among the three regions, resulting in 63 (i.e., $2^6 - 1$) different model structures with the null model excluded from the analysis. Therefore, 63 different structures of modulatory input intelligibility crossed with the 7 different structures of sensory input auditory resulted in a total of 441 models to be compared in the three-region modeling analysis.

To investigate the specific neural dynamic for the tonal language (Chinese), we then specified the four-region model space that included the right STP into the DCM analysis together with the previously identified left pSTG, aSTG, and IFG. Based on our prior results, we assumed that inputs were into the left pSTG only. Modulatory input intelligibility was assumed to modulate any combination of the 12 directed connections among the four regions, with the null model excluded from the analysis. Thus, a total of 4,095 (i.e., $2^{12} - 1$) different models were estimated and compared.

Model Estimation and BMA. After the model space including all of the candidate models was specified, all candidate models of all subjects were estimated using expectation maximization algorithm, calculating the parameters in each model and the free energy *F* as a good estimation of the log-evidence of each model. Then the model estimation was compared among families with different input sites or different modulated connections.

Our first question was where sensory inputs entering the network. For the three-region model space, seven families with different auditory inputs were compared by the random effect method (RFX) to determine the most possible input of the network, and for the four-region model, the input was assumed to be the same as that in three-region model. After the comparison of the model-input families, we then tested the existence of modulation on each connection. For modulation on each connection, all the candidate models could be categorized into two families according to whether such connection is included (i.e., modulation parameter is assumed to be nonzero) in this model. Then the paired families were compared by the RFX method on each connection. The winning families were determined according to the exceedance probability of the RFX result with high confidence and entered into the BMA to generate a model summary that combined the likely parameters values for each family of good model fitness (22). The comparison for seven families of input sites for three-region model space showed that the input auditory entered the system most likely via only pSTG (with the highest exceedance possibility of 0.53), thus indicating the auditory input likely entered the neural system in the Chinese group exclusively from the posterior part of left temporal cortex, which was similar to the data from the English speakers (21, 22). The family-wise comparison analysis for the four-region model space in the Chinese group showed that the modulations from pSTG to IFG, from IFG to aSTG, from IFG to rSTP, and from aSTG to rSTP were probably nonexistent (all with exceedance possibility < 0.01). Thus, there were 255 models that survived the above restrictions, and these were entered into BMA to calculate the group-level parameters (means and SEs).

Second-Level Analysis of Model Parameters. A one-sample *t* test was used to examine the statistical significance of modulation (nonzero parameter in matrix B) across each group of subjects based on individual BMA results with a threshold of $P < 0.05$ (FDR corrected). Two-sample *t* tests were then used to compare the modulation parameters on these connections between the Chinese and English groups.

Behavioral Dictation in Chinese. Chinese subjects who participated in the MRI experiment and were eligible in the DCM analysis were contacted 4 mo after their brain scanning and asked if they would participate in a surprise dictation test. Eight subjects (4 males and 4 females; mean age, 24.0 y) agreed to participate in the dictation study. They were presented the same intelligible stimuli of Chinese idioms they had been exposed to in the scanner and were required to write down the Chinese characters in the idioms they heard. The dictation results were classified into three categories: (*i*) correct ($90.6 \pm 2.5\%$, mean $\pm$ SD); (*ii*) phonologically correct ($8.7 \pm 2.3\%$, where pronunciation of the words were shown to be correct, but the words were not correctly identified, either the Chinese characters chosen were homophones or Pinyin transcription were given, rather than the correct Chinese character); and (*iii*) wrong ($0.7 \pm 0.5\%$, both the word identity and the pronunciation were wrong). The performance of the dictation was calculated as the percentage of correctness of their writing according to the above three categories. We then conducted correlation analysis to investigate the correlations between the dictation results and the modulation strength of brain network connections from the DCM analysis (Fig. 3*B* and Fig. S3).

ACKNOWLEDGMENTS. This work was supported by China's National Strategic Basic Research Program (973) Grant 2012CB720700 and National Natural Science Foundation of China Grants 31200761, 31421003, 81227003, 81430037, and 61135003.

1. Streeter LA (1976) Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259(5538):39–41.
2. Grossmann T, Oberecker R, Koch SP, Friederici AD (2010) The developmental origins of voice processing in the human brain. *Neuron* 65(6):852–858.
3. Kuhl P, Rivera-Gaxiola M (2008) Neural substrates of language acquisition. *Annu Rev Neurosci* 31:511–534.
4. Stager CL, Werker JF (1997) Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388(6640):381–382.
5. Binder JR, et al. (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10(5):512–528.
6. Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343(6174):1006–1010.
7. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8(5):393–402.
8. Roth G, Dicke U (2005) Evolution of the brain and intelligence. *Trends Cogn Sci* 9(5):250–257.
9. Seidenberg MS (1997) Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275(5306):1599–1603.
10. Cook VJ, Newson M (2007) *Chomsky's Universal Grammar: An Introduction* (Blackwell Publishers, Malden, MA), 3rd Ed.
11. Ueno T, Saito S, Rogers TT, Lambon Ralph MA (2011) Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* 72(2):385–396.
12. Ueno T, Lambon Ralph MA (2013) The roles of the "ventral" semantic and "dorsal" pathways in conduite d'approche: a neuroanatomically-constrained computational modeling investigation. *Front Hum Neurosci* 7:422.
13. Scott SK, Blank CC, Rosen S, Wise RJS (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123(Pt 12):2400–2406.
14. 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.
15. Evans S, et al. (2014) The pathways for intelligible speech: Multivariate and univariate perspectives. *Cereb Cortex* 24(9):2350–2361.
16. Binder JR, et al. (2011) Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *Neuroimage* 54(2):1465–1475.
17. Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 97(22):11793–11799.
18. Friederici AD (2011) The brain basis of language processing: From structure to function. *Physiol Rev* 91(4):1357–1392.
19. Ueno T, et al. (2014) Not lost in translation: Generalization of the primary systems hypothesis to Japanese-specific language processes. *J Cogn Neurosci* 26(2):433–446.
20. Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19(4):1273–1302.
21. Leff AP, et al. (2008) The cortical dynamics of intelligible speech. *J Neurosci* 28(49):13209–13215.
22. Penny WD, et al. (2010) Comparing families of dynamic causal models. *PLOS Comput Biol* 6(3):e1000709.
23. Crinion JT, et al. (2009) Neuroanatomical markers of speaking Chinese. *Hum Brain Mapp* 30(12):4108–4115.
24. Hannas WC (1996) *Asia's Orthographic Dilemma* (Univ of Hawaii Press, Honolulu), p 181.
25. Visser M, Lambon Ralph MA (2011) Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *J Cogn Neurosci* 23(10):3121–3131.
26. Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360(1456):815–836.
27. Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 363(1493):1087–1104.
28. Obleser J, Wise RJ, Dresner MA, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci* 27(9):2283–2289.
29. Gandour J, et al. (2004) Hemispheric roles in the perception of speech prosody. *Neuroimage* 23(1):344–357.
30. Yue Q, Zhang L, Xu G, Shu H, Li P (2013) Task-modulated activation and functional connectivity of the temporal and frontal areas during speech comprehension. *Neuroscience* 237:87–95.
31. Gandour J, et al. (2003) Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain Lang* 84(3):318–336.
32. Binder JR, Desai RH (2011) The neurobiology of semantic memory. *Trends Cogn Sci* 15(11):527–536.
33. Cronk BC, Schweigert WA (1992) The comprehension of idioms: The effects of familiarity, literalness, and usage. *Appl Psycholinguist* 13(2):131–146.
34. Stephan KE, et al. (2010) Ten simple rules for dynamic causal modeling. *Neuroimage* 49(4):3099–3109.
35. Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38(3):387–401.
36. Heim S, et al. (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(2):392–402.