Cooperative cortical network for categorical processing of Chinese lexical tone

Xiaopeng Si[a], Wenjing Zhou[b], and Bo Hong[a,c,1]

[a]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China; [b]Epilepsy Center, Yuquan Hospital, Tsinghua University, Beijing 100084, China; and [c]McGovern Institute for Brain Research, Tsinghua University, Beijing 100084, China

Edited by Patricia K. Kuhl, University of Washington, Seattle, WA, and approved October 6, 2017 (received for review June 20, 2017)

In tonal languages such as Chinese, lexical tone with varying pitch contours serves as a key feature to provide contrast in word meaning. Similar to phoneme processing, behavioral studies have suggested that Chinese tone is categorically perceived. However, its underlying neural mechanism remains poorly understood. By conducting cortical surface recordings in surgical patients, we revealed a cooperative cortical network along with its dynamics responsible for this categorical perception. Based on an oddball paradigm, we found amplified neural dissimilarity between cross-category tone pairs, rather than between within-category tone pairs, over cortical sites covering both the ventral and dorsal streams of speech processing. The bilateral superior temporal gyrus (STG) and the middle temporal gyrus (MTG) exhibited increased response latencies and enlarged neural dissimilarity, suggesting a ventral hierarchy that gradually differentiates the acoustic features of lexical tones. In addition, the bilateral motor cortices were also found to be involved in categorical processing, interacting with both the STG and the MTG and exhibiting a response latency in between. Moreover, the motor cortex received enhanced Granger causal influence from the semantic hub, the anterior temporal lobe, in the right hemisphere. These unique data suggest that there exists a distributed cooperative cortical network supporting the categorical processing of lexical tone in tonal language speakers, not only encompassing a bilateral temporal hierarchy that is shared by categorical processing of phonemes but also involving intensive speech–motor interactions over the right hemisphere, which might be the unique machinery responsible for the reliable discrimination of tone identities.

lexical tone | Chinese | motor cortex | ECoG | high gamma

The ability to transform continuously varying stimuli into discrete meaningful categories is a fundamental cognitive process, called categorical perception (CP) (1). During categorical speech perception, listeners tend to perceive continuously varying acoustic signals as discrete phonetic categories that have been defined in languages (2–4). Stimuli changes within the same phonetic category are processed as invariances, whereas differences across categories are exaggerated (5). Phonemes, the basic unit of speech, are categorically perceived. For example, the equally spaced /ba/-/da/-/ga/ continuum generated by morphing the second formant transition is a classical CP example (6, 7). Neurolinguistics studies showed that the categorical perception of phonemes can be attributed to the neural representation at human superior temporal gyrus (STG) (8, 9). In addition to consonants and vowels, in tonal languages, the lexical tone (the pitch contour of a syllable) serves as a unique phonetic feature for distinguishing words (10, 11). In Mandarin Chinese, the meaning of a word cannot be determined without tonal information. For example, the syllable /i/ can be accented in four lexical tones (i.e., level tone T1, rising tone T2, dipping tone T3, and falling tone T4) to represent four distinct word meanings: medicine "医," aunt "姨," desk "椅," or difference "异," respectively. Behavioral studies have suggested that Mandarin tone is categorically perceived (12–14). However, the neural substrate supporting the categorical perception of lexical tone is not well understood.

Current theories postulate a hierarchical stream in the temporal cortex to map acoustic sensory signals into abstract linguistic objects such as phonemes and words (15–17). The STG, which receives primary auditory cortex input, is considered a hub for the spectrotemporal encoding of sublexical phonetic features (8, 15), whereas the MTG and the anterior temporal lobe (ATL) are responsible for the abstract representations of linguistic objects (18, 19). Lexical tone is a suprasegmental feature involving both acoustic and linguistic factors (20), posing more challenges on sound-meaning mapping than nontonal language. One possible strategy is to engage more neural resources from the higher-level linguistic areas. Behavioral study of lexical tone perception suggested a strong influence of higher-level linguistic information on the low-level acoustic processing (21). However, the neural evidence supporting this higher-level area involvement on lexical tone perception is scarce.

On the other hand, the pitch contour difference between lexical tone categories is very subtle, which poses another challenge for listener's auditory system in discrimination and identification. As postulated by the motor theory of speech perception, the repertoire of speech gestures is easier for the human brain to categorize than the extensive variability of acoustic speech sounds (2, 22). fMRI studies revealed that the motor cortex is involved in speech perception (23–26). Disrupting the speech–motor cortex by transcranial magnetic stimulation can impair phoneme categorization (25, 27). Given that lexical tones are generated via intricate articulatory vocal cord gestures (11), we further hypothesized that the motor cortex in the dorsal speech pathway is involved in lexical tone processing to facilitate the categorization.

Currently, the neural mechanism for lexical tone processing has been primarily studied by neuroimaging and noninvasive

Significance

The variation of pitch in speech not only creates the intonation for affective communication but also signals different meaning of a word in tonal languages, like Chinese. Due to its subtle and brisk pitch contour distinction between tone categories, the underlying neural processing mechanism is largely unknown. Using direct recordings of the human brain, we found categorical neural responses to lexical tones over a distributed cooperative network that included not only the auditory areas in the temporal cortex but also motor areas in the frontal cortex. Strong causal links from the temporal cortex to the motor cortex were discovered, which provides new evidence of top-down influence and sensory–motor interaction during speech perception.

Author contributions: B.H. designed research; X.S., W.Z., and B.H. performed research; X.S. and W.Z. contributed new reagents/analytic tools; X.S. and B.H. analyzed data; and X.S. and B.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Datasets for reproducing all analyses of this study, including ECoG data, MRI, CT images, and stimulus sound files can be accessed at https://doi.org/10.5281/zenodo.926082.

[1]To whom correspondence should be addressed. Email: hongbo@tsinghua.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1710752114/-/DCSupplemental.

electrophysiological techniques (28–36), which are not capable of simultaneously capturing the precise spatiotemporal dynamics of tone processing. Less affected by the skull, the electrocorticography (ECoG) directly recorded from the cortical surface in epilepsy patients provides a unique opportunity to acquire neural signals with both accurate spatial location (approximately millimeters) and high temporal resolution (approximately milliseconds) to explore the neural dynamics of speech processing (37–39). In the present study, ECoG recording coregistered with MRI cortical structure was employed to pinpoint the brain areas and to capture their dynamic interactions that are responsible for categorical encoding of lexical tone.

Results

Behavior tests on the synthesized tone continuum (Fig. 1*A* and Table S1) were first conducted to quantify the categorical perception of Chinese lexical tone and to determine the appropriate stimuli for subsequent ECoG experiments. T2 (rising tone) and T4 (falling tone) were selected as the representative of contour tone, and T1 was selected for level tone (11–13). The psychometric curve of the identification task on the rising–level–falling tone continuum displayed a logistic function, and its category boundary corresponded well with the peaks in the discrimination function (Fig. 1*B*). This result is in agreement with the behavioral model of categorical perception (6) and is consistent with previous studies on Chinese subjects (12, 13). A two-deviant oddball paradigm was adopted in the ECoG experiment (34, 40), in which stimulus token 5 (T1) in the continuum was selected as the frequently presented standard stimuli, whereas tokens 2 (T2) and 8 (T1) served as infrequently delivered deviants. These two deviant stimuli have the same physical distance but different perceptual tone identities with respect to the standard stimulus, forming a within-category tone pair (tokens 5 and 8) and a cross-category tone pair (tokens 2 and 5).

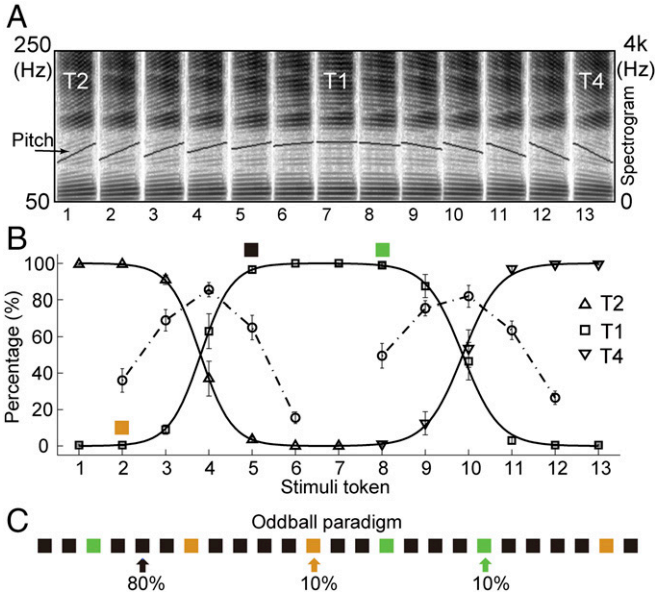

Fig. 1. Categorical behavior performance for the Mandarin tone continuum and the oddball paradigm for neural recordings. (*A*) Synthesized rising–level–falling tone continuum. Wideband spectrogram and pitch contour of the tone continuum synthesized with equal parametric changes in the pitch slope. These 13 tone tokens varied from rising tone (token 1) to level tone (token 7) and then to falling tone (token 13). (*B*) Psychometric functions derived from 10 native Mandarin Chinese speakers. Solid line represents the identification function with the *y* axis for the correct identification percentage in the 2AFC task. Dash-dotted line represents the discrimination function with the *y* axis for the correct discrimination percentage in the AX task (mean ± SEM). Tokens 2, 5, and 8 were selected as oddball stimuli. (*C*) The oddball paradigm for neural recordings. Black: standard stimuli (token 5, 80% trials); Orange, cross-category deviant (token 2, 10% trials); green, within-category deviant (token 8, 10% trials).

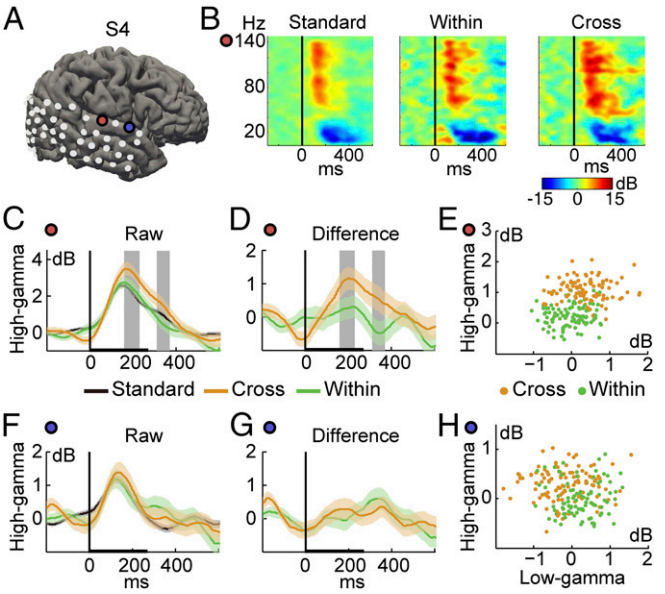

Fig. 2. Enlarged cross-category neural dissimilarity. (*A*) Electrode locations on subject S4's reconstructed cortical surface with examples of categorical (red circle) and noncategorical (blue circle) electrodes. (*B*) Event-related spectrograms for three stimuli in the oddball paradigm from the red electrode, averaged across trials and normalized to the baseline power. Black vertical lines indicate the onset of the auditory stimuli. (*C* and *F*) High-gamma responses for standard stimuli (black curve), cross-category deviant stimuli (orange), and within-category deviant stimuli (green). (*D* and *G*) Difference waveforms for cross-category contrast (orange) and within-category contrast (green). Gray area indicates significantly larger high-gamma responses for cross-category than for within-category stimuli (mean± SEM, Wilcoxon rank-sum test, $*P < 0.05$). (*E* and *H*) Neural responses dissimilarity in 2D space of high-gamma and low-gamma band power, for categorical electrode (*E*) and noncategorical electrode (*H*). Each dot represents an averaged bootstrap resample of 50% trials' mean response.

With the grand averaged spectral pattern of ECoG response to all stimuli, we compared the power changes across major frequency bands: high-gamma (60–140 Hz), low-gamma (30–60 Hz), and beta (15–25 Hz) band (Fig. S1). High-gamma band exhibited the most prominent power change (Fig. S1 *A* and *B*), which is significantly larger than the low-gamma and beta band (Fig. S1*C*). Thus, our analysis will be mainly focused on the high-gamma frequency band. The neural dissimilarity of tone pairs was then measured by the difference of high-gamma response to the standard and to the deviant at each electrode. It is reasonable to postulate that the electrodes showing larger neural dissimilarity for cross-category pair than for within-category pair may contribute to the categorical perception of lexical tones. As an example, in one of our subjects with right hemisphere electrode coverage (Fig. 2 and Fig. S2) (another example with left hemisphere coverage is presented in Fig. S3), two STG electrodes showed distinct response patterns: a categorical response (Fig. 2 *C*–*E*) and a noncategorical response (Fig. 2 *F*–*H*). For the categorical response electrode, the event-related spectrogram exhibited an increased high-gamma response to cross-category tone stimulus (Fig. 2*B*), and the cross-category deviant stimulus had a significantly larger response power than the within-category deviant stimulus (Fig. 2*C*, $P < 0.05$). The difference signals between the high-gamma response to the standard (token 5) and to the deviant stimulus (token 2) also indicate that the cross-category neural dissimilarity was significantly larger than that of the within-category case (token 8) (Fig. 2*D*; $P < 0.05$). By contrast, for the noncategorical response electrode, although there existed a power increase for both deviant stimuli, the difference between the cross-category contrast and the within-category contrast was not significant (Fig. 2 *F* and *G*; $P > 0.05$). Neural response

clusters to different tone stimuli in the 2D features space of high gamma and low gamma also showed increased separability in the categorical electrode (Fig. 2*E*) than in the noncategorical electrode (Fig. 2*H*), which is consistent with our pilot study (41). Comparison of neural response separability between cross-category and within-category tones further confirmed the major contribution of high-gamma activity (Fig. S1*D*).

We examined the response patterns across all electrodes covering the temporal and motor cortices from six patients (Fig. 3*A* and Fig. S4). Electrodes from the bilateral STG showed the strongest auditory response (Fig. S5), which is consistent with previous ECoG findings (8). Among them, the categorical response electrodes for tone processing were identified as those that had significantly larger cross-category contrast than within-category contrast in the high-gamma response (*Materials and Methods*). The categorical response electrodes were distributed over the STG, the MTG, and the motor cortex bilaterally, which was shown on both the individual (Fig. 3*A*) and averaged cortical surfaces (Fig. 3*B*). We examined the categorical values of STG, MTG, and motor areas and found no significant difference between them (Kruskal–Wallis three-level one-way ANOVA test, $P = 0.51$). In addition, the averaged high-gamma peak power of right STG categorical electrodes is significantly larger than that of left STG (Fig. 3*C*; $P < 0.05$), whereas the categorical value of bilateral STG did not show any significant lateralization (Fig. 3*D*).

The spatial distribution of categorical response electrodes displayed a network composed of cortical areas from both ventral and dorsal pathways for speech processing. To further illustrate the dynamic information flow among these categorical sites, we examined the temporal latency of the high-gamma responses at the STG, MTG, and motor-related areas. The high-gamma peak latency increased from the STG and the motor area to the MTG area, showing a temporal propagation of cortical activations during lexical tone processing (Fig. 4*A*). Activation of the STG reached its peak earliest, with a mean value of 206 ms, which is significantly longer than the 110- to 150-ms latency that has been found for categorical

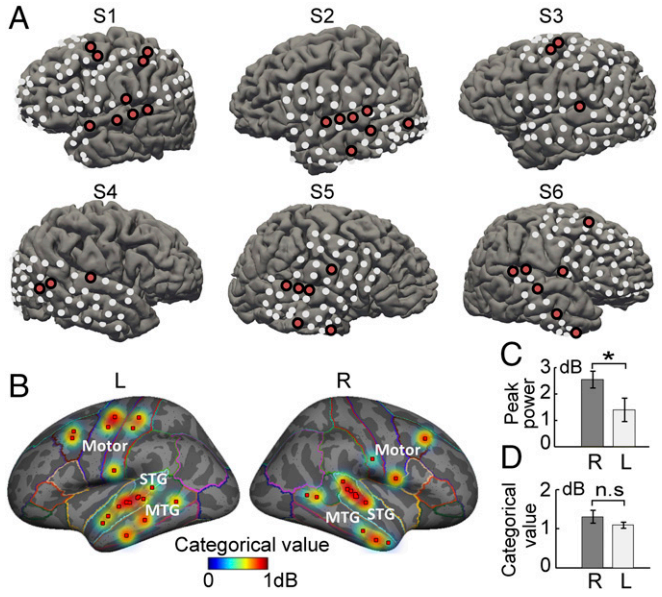

Fig. 3. Cortical sites with categorical responses to Chinese lexical tones. (*A*) The grid electrode coverage for the six subjects. Categorical response sites are colored in red on each individual subject's cortical surface. (*B*) Categorical responsive sites and the corresponding categorical values were interpolated and mapped onto the averaged inflated brain model. Categorical sites: bilateral superior temporal gyrus (STG, $n = 16$); bilateral middle temporal gyrus (MTG, $n = 8$); bilateral primary motor, somatosensory and premotor cortex (motor, $n = 10$). Comparison of (*C*) response peak power and (*D*) categorical value between STG of two hemisphere (mean ± SEM; Wilcoxon rank-sum test; $^*P < 0.05$; right STG, $n = 7$; left STG, $n = 9$).

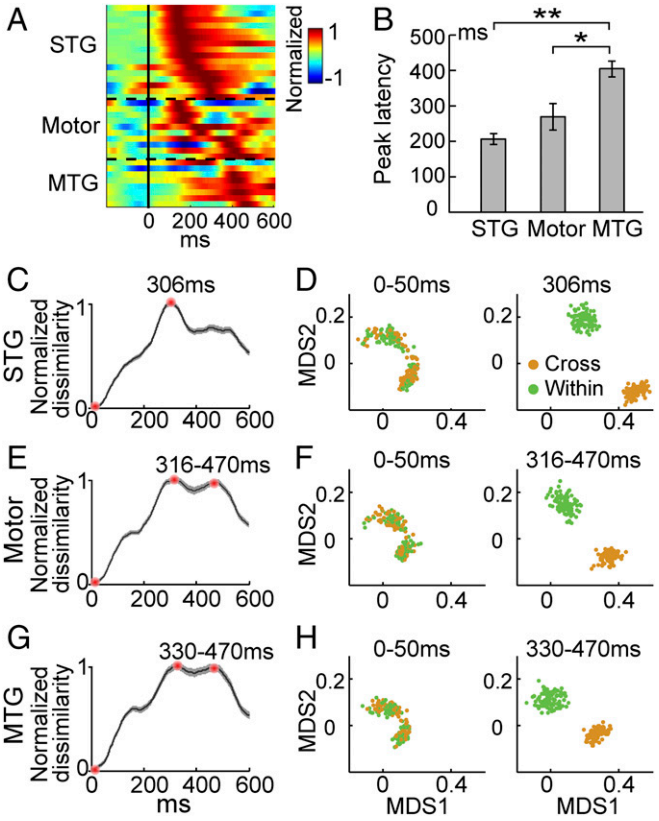

Fig. 4. Response latency comparison across all categorical electrodes (*A* and *B*) and the temporal dynamics of neural dissimilarity using multiple electrodes analysis (*C*–*H*). (*A*) Trial-averaged high-gamma responses of each electrode (STG, $n = 16$; MTG, $n = 8$; motor, $n = 10$). (*B*) Peak latency of the high-gamma response (mean ± SEM; Wilcoxon rank-sum test with the Bonferroni correction; $^*P < 0.05$, $^{**}P < 0.005$). (*C*) Normalized neural response dissimilarity function between cross-category (orange) and within-category (green) tone pairs for STG categorical electrodes (mean ± SEM; error bar was estimated using bootstrapping resampling methods with 100 times). (*D*) Relational organization of the onset time (0–50 ms) and peak stage's neural response dissimilarity using multidimensional scaling (MDS) for STG (peak at 306 ms). Each dot is a bootstrapping resampling sample. (*E* and *F*) Normalized dissimilarity function and relational organization for motor areas (peaked at 316 ms with the second largest peak at 470 ms). (*G* and *H*) Normalized dissimilarity function and relational organization for MTG (peaked at 330 ms with the second largest peak at 470 ms).

processing of phonemes at the STG (8). Activation of the MTG reached its peak latest, with a mean value of 405 ms (Fig. 4*B*). The motor cortex electrodes had diversified peak latencies, with a mean value of 270 ms, which was between the latencies of the STG and the MTG. Furthermore, we explored the temporal evolution of the neural dissimilarity by using multiple electrodes analysis (Fig. 4 *C*–*H*). The Euclidean distances between neural responses to cross-category and within-category tones from multiple electrodes within each region were calculated at each time point between 0 and 600 ms after stimulus onset. The neural dissimilarity curve of all three regions showed a sharp increase and peaked at around 300 ms but with different peak features (Fig. 4 *C*, *E*, and *G*). The temporal order of the dissimilarity peaks (STG 306 ms – motor 316 ms – MTG 330 ms) is in good accordance with the order of high-gamma peak latency of individual electrodes (Fig. 4*B*). There is a plateau of enlarged neural dissimilarity (around 300–470 ms) for both motor and MTG areas, whereas there is only a single peak around 300 ms in STG. This may suggest different neural coding mechanisms between early auditory processing (STG) and late perceptual processing (motor/MTG). Moreover, the visualization of the multielectrode neural dissimilarity with multidimensional scaling (Fig. 4 *D*, *F*, and *H*) further verified the finding of

NEUROSCIENCE

the enlarged neural dissimilarity in single electrode during categorical tone perception.

To reveal the neural interaction among major nodes in the network, Granger causality (GC) analysis was used to explore the directional information flow between electrode pairs (Fig. S6). GC influences were estimated for the within-category deviant condition (Fig. 5*A* and Fig. S7 *A* and *B*) and for the cross-category deviant condition (Fig. 5*B* and Fig. S7 *C* and *D*). In both conditions, electrodes over the motor cortex were found to interact with the STG and the MTG during lexical tone processing. Although we were not able to pinpoint the exact timing of the interaction, this dual-way interplay may explain the diversified response latency of motor sites during the time window of 200–400 ms (Fig. 4 *A* and *B*). Moreover, we found both enhanced and emerged GC connections under the cross-category condition compared with the within-category condition, especially in the right hemisphere. The right ATL had feedback influences to the right motor cortex and received feed-forward connections from the right STG (Fig. 5*B*, *Right*). In addition, the right STG received feedback information from the posterior MTG (pMTG).

Discussion

In contrast to previous findings of localized areas for lexical tone processing (31, 32), our results revealed a distributed network involving both the ventral and dorsal streams of speech processing. The bilateral STG is responsible for the initial stage of categorical processing of lexical tone, corresponding to the earliest peak latency (~200 ms). The bilateral MTG is responsible for the higher level of categorical processing, with the latest peak response (~400 ms), which may be responsible for lexical processing of tones. Surprisingly, the bilateral motor cortex was found to be involved in categorical lexical tone processing, which exhibited interactions with both the STG and the MTG. In the cross-category condition, there was enhanced Granger influence in the right hemisphere, in which the anterior part of the temporal lobe not only is influenced by the STG but also has causal influence on the motor cortex. Taking together, in high spatial and temporal resolutions, we report that there exists a cooperative cortical network with recurrent connections that supports the categorical processing of lexical tone in tonal language

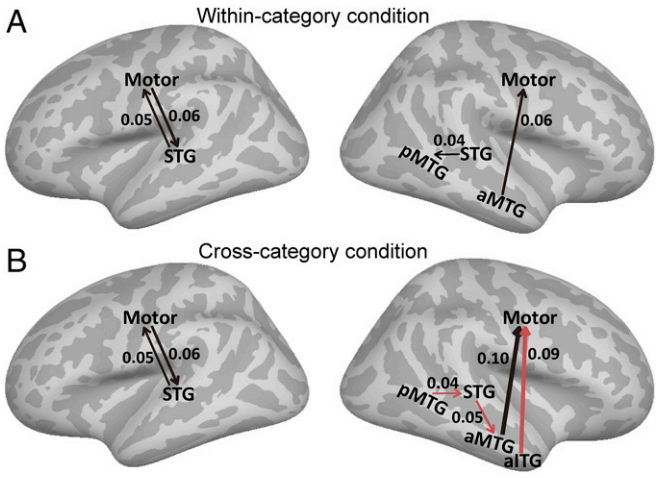

Fig. 5. Granger causality (GC) analysis across all categorically responsive electrodes. (*A*) Significant Granger causality influence under the within-category deviant tone condition (permutation test, $P < 0.001$). (*B*) Significant GC influence under the cross-category deviant tone condition (permutation test, $P < 0.001$). Red line indicates the unique connection of the cross-category condition compared with the within-category condition. The magnitude of the GC value is indicated by the line width [posterior middle temporal gyrus (pMTG), anterior middle temporal gyrus (aMTG), and anterior inferior temporal gyrus (aITG)].

speakers, encompassing a bilateral temporal hierarchy and involving enhanced sensory–motor interactions.

Previous studies have shown that the high-gamma response is a robust neural feature for cortical functional processing (42–44), tightly correlated with neuronal firing (45, 46), whereas low-gamma and beta band activity is usually considered as a neural oscillation generated by certain cortical networks (47, 48). In this study, we found that high-gamma activity in multiple cortical areas showed not only a reliable strong response power but also a better separability for tone stimuli from different categories. For the enlargement of neural dissimilarity in categorical perception, the low-gamma and beta band power contributed much less than high-gamma (Fig. 2 *E* and *H* and Fig. S1). These observations further support the role of high gamma activity in reflecting local neuronal processing. Meanwhile, the causal links between cortical sites occurred in low-frequency band (Fig. S6), which suggested the unique role of low-frequency band activity in remote functional connections (49).

In the Oddball paradigm we used, the physical distance between the standard tone and the cross-category tone is the same as that with the within-category tone. However, the neural response dissimilarity between the cross-category tone pairs is enlarged, whereas that for the within-category tone pairs is not. This finding provides direct neural substrate supporting the behavioral studies that have postulated categorical perception of Chinese lexical tone (12–14). The selective neural dissimilarity enlargement represents the nonlinear neural mechanism of the categorical perception (8, 50). In our data, multiple cortical sites exhibited this nonlinear amplification effect, which supplements previous findings of phoneme categorical representation in STG (8) and our early observation of lexical tone processing in STG/MTG (41). There might be multiple sources contributing to the categorical perception of Chinese tone, including the acoustic stimulus complexity at the bottom, the long-term phonetic representation, and semantic dictionary on the top (13, 21). The neural network and its dynamics we observed here may correspond to these multiple level of nonlinear transformation. Our data also indicate that this categorical processing occurs not only in auditory modality, originating from the bilateral STG, but also with contributions from high-level semantic hub and even motor cortex (Fig. 3 *A* and *B*).

The functional hierarchy along the ventral pathway has been well established for the transformation from sound to meaning in nontonal languages (16, 17). In our study, the temporal order of processing stages was captured by the peak of response power and dissimilarity function, which supported the same role of this feed-forward stream in Chinese lexical tone processing (Fig. 4). The middle temporal gyrus (MTG) was found to be involved in categorical phonemic tone processing in both hemispheres. Given the latest response latency and unique plateau period of neural dissimilarity curve (Fig. 4*G*), we argue that the MTG may store the lexical knowledge of tones and is the lexical interface between phonetic and semantic representations (15, 51). The posterior-to-anterior Granger information flow we observed in the temporal cortex further supported the existence of a processing hierarchy (Fig. 5*B*). Besides, it has been proposed that ATL acts as a semantic hub for phoneme representations at higher level (18). We found that the right ATL was recruited not only with information flow from STG but also with causal influence on motor cortex (Fig. 5*B*). This finding is in line with a structural MRI study that showed the right ATL is a neuroanatomical marker for Chinese speakers (52). A recent fMRI connectivity study also implicated the right ATL as a unique hub for Chinese speech perception (53). Our results specifically support the functional role of the ATL in Chinese lexical tone processing. In a broad sense, our findings provided neural substrates for the dual-process model of speech categorical perception in general (4, 13, 54), with the STG–MTG hierarchy processing the continuous auditory features (bottom-up acoustic processing) and the ATL serving as the semantic hub to facilitate cross-category discrimination (top-down linguistic influence). The prevalent effect of cross-category exaggeration across many cortical sites, including auditory, sensorimotor, and semantic areas, may

explain the dominant influence from linguistic domain on lexical tone perception for native Chinese speakers (21).

Current views suggest that the dorsal language stream is utilized in sensory–motor transformations during listening and speaking (43, 55, 56). The motor theory argues that articulatory gestures are less variable than speech sounds and suggests that speech perception is the perception of speech motor gestures (2, 22). In the current study, during a passive listening task, the motor cortex in the dorsal speech stream was found to be involved in categorical lexical tone processing, which adds a third neural resource to the ventral network information flows. This result is in line with previous ECoG findings on English phoneme, which showed robust high-gamma responses of the motor cortex under pure listening conditions (43). Because different Chinese lexical tones are produced by intricate control of the tension and thickness of vocal cords (11), it is likely that the motor cortex, which contains the tonal articulatory representation (43, 57), facilitates the categorization of lexical tone. The bidirectional influence between motor and STG (Fig. 5) may underlie this facilitation (43). Furthermore, the motor cortex was found to receive significant Granger influence from the higher linguistic area ATL, which suggests that the perceptual processing of speech by the motor cortex may require the guidance of top-down feedback.

Materials and Methods

Subjects. The subjects were medically intractable epilepsy patients who underwent electrode implantation for localizing the epileptic seizure foci to guide neurosurgical treatment. Six patients (S1–S6) with surface electrode coverage participated in this study (Fig. S4 and Table S2). Electrode placement was determined solely by clinical need. No seizure had been observed 1 h before or after the tests in all patients. Written informed consent was obtained from the patients, and this study was approved by the Ethics Committees of the Yuquan Hospital, Tsinghua University.

Tone Continuum. Behavior testing of the categorical perception of Mandarin Chinese tone was conducted to select the appropriate stimuli for the oddball paradigm. A synthesized T2–T1–T4 (rising–level–falling) tone continuum of Mandarin monosyllables /i/ with equal pitch distance change from the neighboring token (Fig. 1*A* and Table S1) was utilized as stimuli in the behavioral study. The equal pitch distance was measured via equivalent rectangular bandwidth (ERB), an objective parameter commonly used in hearing studies (13, 58, 59). The tone continuum was synthesized by a pitch-synchronous overlap/add method (60) implemented in Praat software (61). The original syllable, a level tone /i/, was retrieved from the Mandarin monosyllabic speech corpora of the Chinese Academy of Social Sciences–Institute of Linguistics.

Behavior Task. Ten subjects, all native speakers of Mandarin Chinese, were recruited for behavior testing (five male, five female, 20–30 y). No subject reported any hearing or vision difficulty. All subjects provided written informed consent, and this study was approved by the Ethics Committees of Medical School of Tsinghua University. The identification task was a two-alternative forced choice (2AFC) task during which the subjects were asked to identify each stimulus identity by pressing a button corresponding to the correct identity. In this session, each stimulus was presented in 20 trials. The AX discrimination task required subjects to judge whether the presented stimuli pairs were the same or different. Stimuli pairs were delivered in two-step intervals, and each pair was used in 10 trials. The experiment was conducted in a double-walled, soundproof chamber (Industrial Acoustics), and stimuli were randomly presented using Psychophysics Toolbox 3.0 extensions (62) implemented in MATLAB (The MathWorks Inc.).

Oddball Paradigm. Based on the psychometric function derived from the behavior tests, stimuli tokens 2, 5, and 8 were chosen as stimuli for the passive listening oddball paradigm for ECoG recording (Fig. 1*B*). Stimuli token 2 was used for standard trials (80% trials), token 5 was used for cross-category deviant trials (10% trials), and token 8 was used for within-category deviant trials (10% trials) (Fig. 1*C*). Relative to the standard stimulus, the two deviant stimuli had the same physical distance but different category labels. The oddball paradigm contained 500 trials for all subjects (except S6, who underwent 250 trials due to clinical considerations). The interstimulus interval (onset–onset) was 1,100 ms with 5% jitter to avoid the subject's expectation effect. The subjects were asked to watch a silent movie during the experiment.

Analysis of High-Gamma ECoG Responses. All data processing was implemented in MATLAB. Each electrode was visually checked, and electrodes showing epileptiform activity or containing excessive noise were removed. All remaining electrodes that covered the temporal lobe, the sensorimotor cortex, and the premotor cortex were selected for analysis. The baseline period was defined as 0–300 ms before stimulus onset. Event-related spectrograms were calculated using the log-transformed power as previously reported (55, 63) and were derived by normalizing each frequency power band to the baseline mean power using a dB unit. Power was calculated via short-time Fourier transform with a 200-ms Hamming-tapered, 95% overlapping moving window (Fig. 2*B*). After a comparison of response power and stimulus discriminability across beta, low-gamma, and high-gamma frequency bands (Fig. S1), we focused our analysis on the high-gamma response (60–140 Hz), which provided the most robust spectral measure of cortical activation (42, 63). The time-varying high-gamma power envelopes (Fig. 2 *C* and *F* and Figs. S2 and S3 *C* and *F*) were processed using the following steps: (*i*) raw ECoG data were band-pass filtered to 60–140 Hz with an FIR filter; (*ii*) the filtered data were then translated into a power envelope by taking the absolute amplitude of the analytic signals passed through a Hilbert transform; (*iii*) to calculate the event-related power changes, the power envelopes were baseline corrected by dividing by the baseline mean power; and (*iv*) finally, the high-gamma power envelopes were log-transformed into dB units.

Electrode Classification. Electrodes without auditory responses to any of the three oddball stimuli were excluded from the analysis. An electrode was identified as auditory responsive if it had a significantly larger high-gamma response than baseline for a period lasting at least 50 ms (paired test according to Wilcoxon signed-rank test, $P < 0.05$) (Fig. S5). An electrode was identified as categorically responsive if it met the following criteria: (*i*) the electrode showed an auditory response to the cross-category stimuli, (*ii*) the cross-category condition evoked a significantly larger high-gamma response than the within-category condition, and (*iii*) the significance period lasted continuously for at least 50 ms (two-sample test by Wilcoxon rank-sum test, $P < 0.05$). The auditory responsive electrodes that did not meet the above criteria were classified as noncategorical electrodes.

Categorical Value. To quantify the strength of an electrode's categorical response, we defined the categorical value as the peak value of the difference signal between the cross-category high-gamma response and the within-category high-gamma response. In the case of categorical response, this value should be bigger than 0. For visualization, the categorical value of each electrode was color-coded on the inflated brain (Fig. 3*B*).

Dissimilarity Measurement and Multidimensional Scaling Analysis. To examine the temporal evolution of distance between neural representation of lexical tones, we constructed a multidimensional space by using the high-gamma power of all categorically responsive electrodes in three regions (STG, $n = 16$; MTG, $n = 8$; motor, $n = 10$). To quantify the overall spatial activation differences between cross and within category tones, in each brain region, the neural dissimilarity was measured by the Euclidean distance (64, 65) between multielectrode high-gamma responses in two conditions at each time point of 0–600 ms after stimulus onset, resulting in a dissimilarity curve. To better illustrate the dynamic change across time, the dissimilarity curve was normalized to 0–1 by the maximum and minimum distance values (Fig. 4 *C*, *E*, and *G*). To further visualize the relational organization of the neural responses to different lexical tones, the unsupervised multidimensional scaling (MDS) was used to project the high-dimensional neural space onto a 2D plane (8, 43). A 100-times bootstrapping resampling method was used to estimate the mean and variance of the neural representation in the multidimensional neural space (Fig. 4 *D*, *F*, and *H*).

Granger Causality Analysis. To investigate the directional information flows between category areas, the Granger Causal Connectivity Analysis (GCCA) Toolbox (66) was used. Because Granger causality (G-causality) requires the covariance stationarity of each time series, we applied a Box–Jenkins autoregressive integrative moving average model (67, 68) to prewhiten the ECoG data. Stationarity was confirmed by a Kwiatkowski Phillips Schmidt Shin (KPSS) test (66). The spectral G-causality analysis (GCA) (Fig. S6) was conducted using a multivariate autoregressive model included in the GCCA toolbox. For the model, we used a rank of 75 ms according to our corresponding estimates for cortical-to-cortical high-gamma signal propagation as obtained from the previous peak latency analysis. We used a 500-times permutation resampling method (the electrode pairs' corresponding trials were shuffled randomly) to determine the significant threshold value of spectral G-causality. A G-causality analysis was performed on each individual subject's poststimulus 0.3- to 0.8-s ECoG data, which prevented evoked potential influences. All categorical responsive electrodes shown in Fig. 3*A* were used for GCA calculation. The total

number of sites for GCA is 35 (STG, $n = 16$; MTG, $n = 8$; motor, $n = 10$; ITG, $n = 1$). The GCA analysis was conducted between all possible pairs of above electrodes within each subject's hemisphere. In total, there were 16 significant connections for the cross condition (Fig. S7*A*) and 13 significant connections for the within condition (Fig. S7*C*). The mean GC values between cortical areas were also calculated and reported (Fig. S7 *B* and *D*).

ACKNOWLEDGMENTS. We thank Xiaofang Yang, Juan Huang, and Xiaoqin Wang for their comments on the behavior experiment design; Chen Song and Yang Zhang for the comments on neural data analysis; Hao Han and Le He for MRI data collection; and Rami Saab for language modification. We thank the reviewers/editors for critical reading of the manuscript. We appreciate the time and dedication of the patients and staff at Epilepsy Center, Yuquan Hospital, Tsinghua University, Beijing. This work was supported by the National Science Foundation of China (NSFC) and the German Research Foundation (DFG) in project Crossmodal Learning, NSFC 61621136008/DFG TRR-169 (to B.H.), NSFC 61473169 (to B.H.), and National Key R&D Program of China 2017YFA0205904 (to B.H.).

1. Harnad SR (1987) *Categorical Perception: The Groundwork of Cognition* (Cambridge Univ Press, Cambridge, UK).
2. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461.
3. Fry DB, Abramson AS, Eimas PD, Liberman AM (1962) The identification and discrimination of synthetic vowels. *Lang Speech* 5:171–189.
4. Pisoni DB (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept Psychophys* 13:253–260.
5. Perkell JS, Klatt DH (1986) *Invariance and Variability in Speech Processes* (Lawrence Erlbaum Associates, Hillsdale, NJ).
6. Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 54:358–368.
7. Liberman AM, Harris KS, Kinney JAS, Lane H (1961) The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J Exp Psychol* 61:379–388.
8. Chang EF, et al. (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13:1428–1432.
9. Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343:1006–1010.
10. Howie JM (1976) *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge Univ Press, Cambridge, UK).
11. Duanmu S (2000) *The Phonology of Standard Chinese* (Oxford Univ Press, Oxford).
12. Wang WS (1976) Language change. *Ann N Y Acad Sci* 280:61–72.
13. Xu Y, Gandour JT, Francis AL (2006) Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J Acoust Soc Am* 120:1063–1074.
14. Peng G, et al. (2010) The influence of language experience on categorical perception of pitch contours. *J Phonetics* 38:616–624.
15. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
16. DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci USA* 109:E505–E514.
17. Leonard MK, Chang EF (2014) Dynamic speech representations in the human temporal lobe. *Trends Cogn Sci* 18:472–479.
18. Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8:976–987.
19. Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *J Neurosci* 30:7604–7612.
20. Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 363:1087–1104.
21. Zhao TC, Kuhl PK (2015) Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing. *J Acoust Soc Am* 138:EL133–EL137.
22. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36.
23. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702.
24. Pulvermüller F, et al. (2006) Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci USA* 103:7865–7870.
25. Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. *Curr Biol* 17:1692–1696.
26. Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013) Automatic phoneme category selectivity in the dorsal auditory stream. *J Neurosci* 33:5208–5215.
27. Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci* 29:9819–9825.
28. Klein D, Zatorre RJ, Milner B, Zhao V (2001) A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage* 13:646–653.
29. Hsieh L, Gandour J, Wong D, Hutchins GD (2001) Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang* 76:227–252.
30. Gandour J, et al. (2004) Hemispheric roles in the perception of speech prosody. *Neuroimage* 23:344–357.
31. Wong PC, Parsons LM, Martinez M, Diehl RL (2004) The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *J Neurosci* 24:9153–9160.
32. Xu Y, et al. (2006) Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum Brain Mapp* 27:173–183.
33. Luo H, et al. (2006) Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc Natl Acad Sci USA* 103:19558–19563.
34. Xi J, Zhang L, Shu H, Zhang Y, Li P (2010) Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience* 170:223–231.
35. Zhang L, et al. (2011) Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS One* 6:e20963.
36. Bidelman GM, Lee C-C (2015) Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage* 120:191–200.
37. Pei X, et al. (2011) Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *Neuroimage* 54:2960–2972.
38. Pasley BN, et al. (2012) Reconstructing speech from human auditory cortex. *PLoS Biol* 10:e1001251.
39. Dastjerdi M, Ozker M, Foster BL, Rangarajan V, Parvizi J (2013) Numerical processing in the human parietal cortex during experimental and natural conditions. *Nat Commun* 4:2528.
40. Näätänen R, et al. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
41. Si X, Zhou W, Hong B (2014) Neural distance amplification of lexical tone in human auditory cortex. *Conf Proc IEEE Eng Med Biol Soc* 2014:4001–4004.
42. Crone NE, Sinai A, Korzeniewska A (2006) High-frequency gamma oscillations and human brain mapping with electrocorticography. *Prog Brain Res* 159:275–295.
43. Cheung C, Hamiton LS, Johnson K, Chang EF (2016) The auditory representation of speech sounds in human motor cortex. *Elife* 5:e12577.
44. Edwards E, et al. (2009) Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J Neurophysiol* 102:377–386.
45. Mukamel R, et al. (2005) Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 309:951–954.
46. Nir Y, et al. (2007) Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr Biol* 17:1275–1285.
47. Ray S, Maunsell JHR (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9:e1000610.
48. Kopell N, Ermentrout GB, Whittington MA, Traub RD (2000) Gamma rhythms and beta rhythms have different synchronization properties. *Proc Natl Acad Sci USA* 97:1867–1872.
49. Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat Commun* 5:4694–4694.
50. Raizada RDS, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740.
51. Gow DW, Jr (2012) The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain Lang* 121:273–288.
52. Crinion JT, et al. (2009) Neuroanatomical markers of speaking Chinese. *Hum Brain Mapp* 30:4108–4115.
53. Ge J, et al. (2015) Cross-language differences in the brain network subserving intelligible speech. *Proc Natl Acad Sci USA* 112:2972–2977.
54. Fujisaki H, Takako K (1969) On the modes and mechanisms of speech perception. *Annu Rep Eng Res Inst* 28:67–73.
55. Cogan GB, et al. (2014) Sensory-motor transformations for speech occur bilaterally. *Nature* 507:94–98.
56. Sammler D, Grosbras MH, Anwander A, Bestelmeyer PEG, Belin P (2015) Dorsal and ventral pathways for prosody. *Curr Biol* 25:3079–3085.
57. Correia JM, Jansma BMB, Bonte M (2015) Decoding articulatory features from fMRI responses in dorsal speech regions. *J Neurosci* 35:15015–15025.
58. Greenwood DD (1961) Critical bandwidth and the frequency coordinates of the basilar membrane. *J Acoust Soc Am* 33:1344–1356.
59. Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S (2011) Pitch perception beyond the traditional existence region of pitch. *Proc Natl Acad Sci USA* 108:7629–7634.
60. Moulines E, Laroche J (1995) Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Commun* 16:175–205.
61. Boersma P (2002) Praat, a system for doing phonetics by computer. *Glot Int* 5:341–345.
62. Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
63. Flinker A, et al. (2015) Redefining the role of Broca's area in speech. *Proc Natl Acad Sci USA* 112:2871–2875.
64. Samuelsen CL, Gardner MPH, Fontanini A (2012) Effects of cue-triggered expectation on cortical processing of taste. *Neuron* 74:410–422.
65. Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. *Annu Rev Neurosci* 37:435–456.
66. Seth AK (2010) A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Methods* 186:262–273.
67. Leuthold AC, Langheim FJP, Lewis SM, Georgopoulos AP (2005) Time series analysis of magnetoencephalographic data during copying. *Exp Brain Res* 164:411–422.
68. Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. *Science* 344:424–427.
69. Fischl B (2012) FreeSurfer. *Neuroimage* 62:774–781.
70. Wells WM, 3rd, Viola P, Atsumi H, Nakajima S, Kikinis R (1996) Multi-modal volume registration by maximization of mutual information. *Med Image Anal* 1:35–51.
71. Zhang D, et al. (2013) Toward a minimally invasive brain-computer interface using a single subdural channel: A visual speller study. *Neuroimage* 71:30–41.
72. Desikan RS, et al. (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31:968–980.