Cooperative cortical network for categorical processing of Chinese lexical tone

Xiaopeng Si, Wenjing Zhou, and Bo Hong*†

*Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China; †Epilepsy Center, Yuquan Hospital, Tsinghua University, Beijing 100084, China; and ‡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 perception, 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.

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 the 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 /si/ 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 repertoires 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

www.pnas.org/cgi/doi/10.1073/pnas.1710752114

PNAS | November 14, 2017 | vol. 114 | no. 46 | 12303–12308

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.

This is an open access article distributed under the PNAS license.

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.

1To 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 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. 1B). 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).

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. S1A and B), which is significantly larger than the low-gamma and beta band in the cross-category neural dissimilarity (gray). 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. 2B) and noncategorical response (Fig. 2H). For the categorical response electrode, the event-related spectrogram exhibited an increased high-gamma response to the cross-category tone stimulus (Fig. 2B), and the cross-category deviant stimulus had a significantly larger response power than the within-category case (token 8) (Fig. 2D; 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. 2D; 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. 2F and G; P > 0.05). 

 Neural response 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 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. 2B) and noncategorical response (Fig. 2H). For the categorical response electrode, the event-related spectrogram exhibited an increased high-gamma response to the cross-category tone stimulus (Fig. 2B), and the cross-category deviant stimulus had a significantly larger response power than the within-category case (token 8) (Fig. 2D; 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. 2D; 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. 2F 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. 2E) than in the noncategorical electrode (Fig. 2H), 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. 1D).

We examined the response patterns across all electrodes covering the temporal and motor cortices from six patients (Fig. 3A 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. 3A) and averaged cortical surfaces (Fig. 3B). 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. 3C; \( P < 0.05 \)), whereas the categorical value of bilateral STG did not show any significant lateralization (Fig. 3D).

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. 4A). Activation of the STG reached its peak earliest, with a mean value of 216 ms, which is significantly longer than the 110- to 150-ms latency that has been found for categorical processing of phonemes at the STG (8). Activation of the MTG reached its peak latest, with a mean value of 405 ms (Fig. 4B). 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. 4C–H). The Euclidean distances between neural responses to cross-category and within-category 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).
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. 5A and Fig. S7A and B) and for the cross-category deviant condition (Fig. 5F and Fig. S7C 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. 4A 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. 5B, 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 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. 2E 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 levels 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. 3A and B).

The functional hierarchy along the ventral pathway has been well established for the transformation from sound to meaning in tonal 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 feedforward 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. 4G), 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. 5B). 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. 5B). 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 tone (Fig. 1A 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 stimulus 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 3.0.0 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. 1B). 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%) (Fig. 1C). 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 epileptogenic activity or containing excessive noise or artifacts were eliminated. 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. 2B). 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. 2C and F and Figs. S2 and S3) 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 t 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 categorical response condition elicited 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. 3B).

Disimilarity 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 electrodes, in each brain region, the neural disimilarity was measured by the Euclidean distance (64, 65) between multilexicon electrode high-gamma responses in two conditions at each time point of 0–600 ms after stimulus presentation, resulting in a dissimilarity curve. To better visualize the relation 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 Causality 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 whiten 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 (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. 3A 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 hemisphere. In addition, there were 16 significant connections for the within condition (Fig. S7A) and 13 significant connections for the within condition (Fig. S7C). 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 Xiaoxing 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 the comments on behavior experiment design; Liberman AM, Harris KS, Hoffman HS, Griffin BC (1957) The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol 54:388–388.

Si et al. 10.1073/pnas.1710752114

Supporting Information

Materials and Methods

**ECoG Data Acquisition.** ECoG data were recorded using a g.USBamp amplifier/digitizer system (g.tec, Graz, Austria). The amplifier sampled data at 1,200 Hz with a 0.1 Hz cutoff frequency high-pass filter and a notch filter centered at 50 Hz harmonics to remove power line noise. Four inactive epidural electrodes were placed on the external surface of the skull with the contacts of the electrodes facing away from the skull and were employed as ground and reference (two as ground and two as reference, for redundancy).

**MRI and CT Acquisition.** Presurgical T1-weighted structural MRI images were acquired using a Philips Achieva 3.0T TX scanner with a magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence using the following imaging parameters: matrix size, 256 × 256; voxel size, 0.9 × 0.9 × 1 mm³; and 180 slices. CT images were obtained using a Siemens SOMATOM Sensation 64 CT.

**Anatomical Location of Electrodes.** A presurgical high-resolution T1-weighted structural MRI and a post-implantation CT scan were acquired for all subjects. To coregister brain anatomy and the electrode locations, the CT images were aligned to the presurgical MRI images using FReesurfer (69) (surfer.nmr.mgh.harvard.edu), which includes statistical parametric mapping (SPM) implementations of the mutual information-based transform algorithm (70), as previously described by Zhang et al. (71). The registration was verified visually and manually adjusted if necessary. The average registration error was within a few millimeters (<5 mm). Electrodes were classified according to the anatomical landmarks (STG, MTG, ITG, motor cortex, etc.) within each individual subject’s anatomical space (72) (Fig. S4). For the visualization of all subjects’ electrodes on the average brain surface, each individual subject’s MRI was coregistered to the average brain in Freeesurfer (Fig. 3B). All electrodes were then superimposed onto the inflated average brain.

---

**Fig. S1.** Comparisons of different frequency bands’ responses. (A and B) Averaged event-related spectrograms and power spectral densities (PSDs) computed across all subjects for all electrodes \((n = 205)\) and tone category-sensitive electrodes \((n = 35)\). (Left) Event-related spectrograms. The rectangular box indicates the high-gamma response. (Right) PSDs for the poststimulus power changes \((\text{mean} \pm \text{SEM})\). Pink area indicates high-gamma \((60–140 \text{ Hz})\), orange area indicates low-gamma \((30–60 \text{ Hz})\), and blue area indicates beta \((15–25 \text{ Hz})\) band. A notch filter centered at 50 Hz and its harmonics was used to remove the 50 Hz power line noise. (C) The comparison of response power for high-gamma, low-gamma, and beta band. Averaged across all subjects’ category-sensitive electrodes \((\text{mean} \pm \text{SEM}, \text{Wilcoxon rank-sum test with the FDR correction, } * P < 0.05, n = 35)\). (D) The comparison of the separability index \(d’\) (the ratio of mean distance between within-category and cross-category response over their averaged SDs: \(d’ = (\mu_{\text{cross}} - \mu_{\text{within}}) / \sqrt{\sigma_{\text{cross}}^2 + \sigma_{\text{within}}^2 / 2}\) (mean \(\pm \) SEM, Wilcoxon rank-sum test with the FDR correction, \(** P < 0.005, n = 35)\).
Fig. S2. Example of high-gamma responses across the electrodes of subject S4. (A) The subject’s brain model. (B) High-gamma responses of a categorical electrode (in red) under three conditions (mean ± SEM). (C) Zoomed-in depiction of the example high-gamma responses of electrodes located within the rectangular zone covering the STG and MTG areas.

Fig. S3. Enlarged cross-category neural dissimilarity in subject S3. (A) Electrode locations on subject S3’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.
Fig. S4. Reconstruction of the electrode locations in six subjects. Cortical parcellations were based on anatomical landmarks and were implemented in FreeSurfer. The motor-related areas include primary motor (blue), somatosensory (yellow), and premotor cortex (brown).

Fig. S5. Auditory responsive sites (in red with scale). Auditory response strength was measured by the high-gamma (60–140 Hz) power.

Fig. S6. Examples of significant GC results under the cross-category condition. (A) Significant GC influence from the pMTG to the STG. (B) Significant GC influence from the aMTG to the motor. The significance threshold was determined by permutation test and is indicated by the dashed line. Significant spectral GC results (permutation test, $P < 0.001$) are marked by arrows. Spectral values are shown as the mean ± STD based on a bootstrapping test.
Fig. S7. Matrices of significant Granger causality influences. (A and C) Significant GC value matrices of all electrode pairs in the left and right hemispheres for the within-category (A) and cross-category (C) conditions (permutation test, $P < 0.001$). (B and D) Mean GC values between different cortical areas.

Table S1. Parameters of the tone continuum

<table>
<thead>
<tr>
<th>Step</th>
<th>F0 onset (Hz)</th>
<th>F0 offset (Hz)</th>
<th>Step size $\Delta f$ (Hz)</th>
<th>Step size $\Delta E$ (ERB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102.08</td>
<td>130.00</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>106.55</td>
<td>130.00</td>
<td>4.47</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>111.08</td>
<td>130.00</td>
<td>4.53</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>115.70</td>
<td>130.00</td>
<td>4.62</td>
<td>0.12</td>
</tr>
<tr>
<td>5</td>
<td>120.38</td>
<td>130.00</td>
<td>4.68</td>
<td>0.12</td>
</tr>
<tr>
<td>6</td>
<td>125.15</td>
<td>130.00</td>
<td>4.77</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>130.00</td>
<td>130.00</td>
<td>4.85</td>
<td>0.12</td>
</tr>
<tr>
<td>8</td>
<td>130.00</td>
<td>125.15</td>
<td>4.85</td>
<td>0.12</td>
</tr>
<tr>
<td>9</td>
<td>130.00</td>
<td>120.38</td>
<td>4.77</td>
<td>0.12</td>
</tr>
<tr>
<td>10</td>
<td>130.00</td>
<td>115.70</td>
<td>4.68</td>
<td>0.12</td>
</tr>
<tr>
<td>11</td>
<td>130.00</td>
<td>111.08</td>
<td>4.62</td>
<td>0.12</td>
</tr>
<tr>
<td>12</td>
<td>130.00</td>
<td>106.55</td>
<td>4.53</td>
<td>0.12</td>
</tr>
<tr>
<td>13</td>
<td>130.00</td>
<td>102.08</td>
<td>4.47</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table S2. Clinical profiles of the subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (y)</th>
<th>Sex</th>
<th>Hand</th>
<th>Seizure focus</th>
<th>Grid locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>32</td>
<td>M</td>
<td>L</td>
<td>Lf</td>
<td>Lt, Lf, Lp</td>
</tr>
<tr>
<td>S2</td>
<td>26</td>
<td>F</td>
<td>R</td>
<td>Lt</td>
<td>Lt, Lf, Lo</td>
</tr>
<tr>
<td>S3</td>
<td>22</td>
<td>M</td>
<td>R</td>
<td>Lo</td>
<td>Lt, Lp, Lo</td>
</tr>
<tr>
<td>S4</td>
<td>48</td>
<td>F</td>
<td>R</td>
<td>Ro</td>
<td>Rt, Ro</td>
</tr>
<tr>
<td>S5</td>
<td>31</td>
<td>M</td>
<td>R</td>
<td>Rt</td>
<td>Rt, Rf, Rp</td>
</tr>
<tr>
<td>S6</td>
<td>23</td>
<td>M</td>
<td>R</td>
<td>Rf</td>
<td>Rt, Rf, Rp</td>
</tr>
</tbody>
</table>

All of the subjects had normal cognitive capacity and were functionally independent. F, female; Lf, left frontal; Lo, left occipital; Lt, left temporal; M, male; Rf, right frontal; Ro, right occipital; Rp, right parietal; Rt, right temporal.