Neuroscience 359 (2017) 248–257

SCHIZOPHRENIA AFFECTS SPEECH-INDUCED FUNCTIONAL CONNECTIVITY OF THE SUPERIOR TEMPORAL GYRUS UNDER COCKTAIL-PARTY LISTENING CONDITIONS

JUANHUA LI,[a†] CHAO WU,[b,d†] YINGJUN ZHENG,[a†] RUIKENG LI,[a] XUANZI LI,[a] SHENGLIN SHE,[a] HAIBO WU,[a] HONGJUN PENG,[a] YUPING NING[a] AND LIANG LI[a,b,c*]

[a] *The Affiliated Brain Hospital of Guangzhou Medical University (Guangzhou Huiai Hospital), Guangzhou 510370, China*

[b] *School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100080, China*

[c] *Beijing Institute for Brain Disorder, Capital Medical University, Beijing, China*

[d] *School of Psychology, Beijing Normal University, Beijing 100875, China*

Abstract—The superior temporal gyrus (STG) is involved in speech recognition against informational masking under cocktail-party-listening conditions. Compared to healthy listeners, people with schizophrenia perform worse in speech recognition under informational speech-on-speech masking conditions. It is not clear whether the schizophrenia-related vulnerability to informational masking is associated with certain changes in FC of the STG with some critical brain regions. Using sparse-sampling fMRI design, this study investigated the differences between people with schizophrenia and healthy controls in FC of the STG for target-speech listening against informational speech-on-speech masking, when a listening condition with either perceived spatial separation (PSS, with a spatial release of informational masking) or perceived spatial co-location (PSC, without the spatial release) between target speech and masking speech was introduced. The results showed that in healthy participants, but not participants with schizophrenia, the contrast of either the PSS or PSC condition against the masker-only condition induced an enhancement of functional connectivity (FC) of the STG with the left superior parietal lobule and the right precuneus. Compared to healthy participants, participants with schizophrenia showed declined FC of the STG with the bilateral precuneus, right SPL, and right supplementary motor area. Thus, FC of the STG with the parietal areas is normally involved in speech listening against informational masking under either the PSS or PSC conditions, and declined FC of the STG in people with schizophrenia with the parietal areas may be associated with the increased vulnerability to informational masking.

Key words: schizophrenia, speech perception, precedence effect, functional connectivity, masking, superior temporal gyrus.

INTRODUCTION

Successful speech recognition under speech-on-speech masking (cocktail-party) conditions involves multiple perceptual/cognitive processes, including target detection, selective attention, sensory/working memory, and speech production. Thus, speech recognition against informational speech masking are based on activation of different brain regions with various perceptual/cognitive functions (e.g., Hill and Miller, 2010; Nakai et al., 2005; Scott and McGettigan, 2013; Scott et al., 2004; Wu et al., 2017a,b; Zheng et al., 2016). Among the brain regions that are related to speech recognition against informational masking, the superior temporal gyrus (STG) has been the most studied cortical structure, because speech-evoked activation of the STG can be enhanced by introducing a masking voice, suggesting that the STG is involved in overcoming informational-masking-induced difficulties in speech listening (e.g., Dole et al., 2014; Evans et al., 2016; Nakai et al., 2005; Scott et al., 2004, 2009). More specifically, the dorsolateral superior temporal lobes are consistently activated in speech perception within an informational masking context (speech-in-speech listening conditions compared to either rest or speech-in-noise listening conditions, for a review see Scott and McGettigan, 2013) and some MEG/EEG studies have demonstrated the critical role of the lateral STG in the selective representation of target speech in the presence of competing speech (Bidelman and Dexter, 2015; Ding and Simon, 2012, 2013). Particularly, the lateral STG is more activated by the introduction of informational masking conditions (speech-in-speech) than that of energetic masking conditions (speech-in-noise) (Scott et al., 2004; Scott and McGettigan, 2013). Thus, it is reasonable to hypothesize that the involvement of the STG in speech listening against informational masking may be largely dependent on functional connectivity (FC) of the STG with other brain regions that are critical to speech recognition against

*Correspondence to: L. Li, School of Psychological and Cognitive Sciences, Peking University, 52 Haidian Road, Beijing 100080, China.
E-mail address: liangli@pku.edu.cn (L. Li).

† The three authors contributed equally to this work and should be co-first authors.

http://dx.doi.org/10.1016/j.neuroscience.2017.06.043

informational speech masking. For other brain regions, such as the rostral and dorsal prefrontal cortices and the posterior parietal cortex whose activation is also vulnerable to masking sounds, greater activation occurs under energetic masking conditions (speech-in-noise) relative to informational masking conditions (speech-in-speech) (Scott et al., 2004; Scott and McGettigan, 2013), suggesting that these cortical regions may not be specific for overcoming informational masking.

People with schizophrenia experience more difficulties in filtering distracting signals to prevent information overflow that causes numerous cognitive dysfunctions (Gottesman and Gould, 2003; Braff and Light, 2005). Particularly, speech recognition in people with schizophrenia is markedly vulnerable to masking, especially informational masking (Lee et al., 2004; Ross et al., 2007; Wu et al., 2012, 2013, 2017a,b; Zheng et al., 2016). For example, both first-episode patients and chronic patients with schizophrenia perform worse than their matched healthy controls in recognizing target speech when a masker, particularly a two-talker-speech masker is presented (Wu et al., 2012, 2013, 2017a,b; Zheng et al., 2016). Up to date, the brain substrates underlying the schizophrenia-related augmentation of the vulnerability of speech recognition to informational speech masking largely remain unknown. Thus, it is important to investigate whether the schizophrenia-related vulnerability to informational masking is related to alterations in speech-listening-induced FC of the STG under speech-on-speech masking.

To improve speech recognition under cocktail-party listening condition with multiple talkers, listeners can use various perceptual/cognitive cues available to facilitate perceptual segregation between speech sources and enhance selective attention to the target speech. One of the cues is the precedence effect-induced perceived spatial separation (PSS) between target speech and masking speech (Freyman et al., 1999; Li et al., 2004; Wu et al., 2005; Rakerd et al., 2006; Huang et al., 2008; Zheng et al., 2016). Zheng et al. (2016) have recently reported that compared to the perceived spatial co-location (PSC) listening condition (where target speech and masking speech are perceived from same location on the basis of the auditory precedence effect), introducing the listening condition with perceived spatial separation (between target speech and masking speech), which releases target speech from informational speech masking, induces enhanced activation in the superior parietal lobule (SPL), precuneus, anterior cingulate cortex, lateral middle frontal gyrus, and triangular inferior frontal gyrus. It is not clear whether FC of the STG with these cortical regions is involved in speech listening against informational masking.

This study aimed to explore differences in speech-listening-induced FC of the STG under informational speech masking conditions between healthy listeners and listeners with schizophrenia. More specifically, general psychophysiological interaction (gPPI) analyses (McLaren et al., 2012) were used to identify FC of the STG associated with target-speech listening against speech masking in healthy listeners and listeners with schizophrenia, when the listening condition with either PSS or PSC between the target speech and masking speech was introduced (Zheng et al., 2016).

EXPERIMENTAL PROCEDURES

Participants

Participants with schizophrenia (whose first language was Mandarin Chinese), diagnosed with the Structured Clinical Interview for DSM-IV (SCID-DSM-IV, First et al., 1997), were recruited in the Affiliated Brain Hospital of Guangzhou Medical University (Guangzhou Huiai Hospital). Some patient participants were excluded from this study if they had comorbid diagnoses, substance dependence, and/or other conditions that affected experimental tests (e.g., hearing loss, a treatment of the electroconvulsive therapy (ECT) within the past three months, a treatment of trihexyphenidyl hydrochloride with a dose of more than 6 mg/day, and/or an age younger than 18 or older than 59) (Zheng et al., 2016).

Demographically matched healthy participants (i.e., healthy controls) were recruited from the communities around the hospital with the recruiting criteria used previously (Wu et al., 2012, 2013, 2017a,b; Zheng et al., 2016). They were telephone interviewed first and then those who passed the telephone interview were screened with the SCID-DSM-IV as used for patient participants. None of the selected healthy controls had either a history of Axis I psychiatric disorder as defined by the DSM-IV.

Twenty-four patients and 18 healthy controls participated in the study. Two patient participants and 1 healthy-control participant were excluded from data analyses due to their excessive head movements (more than 3 mm in translation and/or 3° in rotation). Two healthy participants and 4 patient participants were excluded due to failure in following the instructions to button-press. The remaining 20 patients (9 females and 11 males, aged 32 ± 9.8 years) and 16 controls (8 females and 8 males, aged 30.3 ± 9.1 years) were included in fMRI data analyses (Table 1). All participants were right-handed with normal pure-tone hearing thresholds at each ear (< 30 dB Hearing Level) at frequencies between 125 and 8000 Hz. All the participants had Mandarin Chinese as their first language. All the patient participants were clinically stable during their participation, and received antipsychotic medications during this study with the average chlorpromazine equivalent of 605 mg/day based on the conversion factors described by Woods (2003). Some of the patient participants received benzodiazepines based on doctors' advice for the purpose of improving sleeping.

The locally validated version of the Positive and Negative Syndrome Scale (PANSS) tests (Si et al., 2004) was conducted on the day of fMRI scanning for all participants. Patients, patients' guarantees (for their consent of patients' participation in the study) and healthy participants gave their written informed consent for participation in this study. The procedures of this study were approved by the Independent Ethics Committee (IEC) of the Guangzhou Huiai Hospital.

Table 1. Characteristics of participants with schizophrenia and healthy controls

Characteristic	Schizophrenia (n = 20)	Healthy Control (n = 16)
Age (years ± SD)	32.0 (9.8)	30.3 (9.1)
Male%	55.00 (11)	50.00 (8)
Education (years ± SD)	13.05 (3.07)	14.56 (2.80)
MID (years ± SD)	7.75 (6.39)	NA
PANSS	53.65 (6.51)	NA
P-scale	14.50 (5.28)	NA
N-scale	11.25 (4.30)	NA
G-scale	27.75 (3.84)	NA
Diagnostic subtype	N	
Paranoid	9	
Non-paranoid	11	
Typical	10	
Atypical	17	
Typical/atypical*	7	
Chlorpromazine equivalent	Mean:605.38 SD:365.00 Range:200–1600	

SD: standard deviation; PANSS: positive and negative syndrome scale; MID: mean illness duration; NA: not applicable. *Note that 7 patients received 2 different antipsychotics.

Speech stimuli

There were two types of speech stimuli: target speech and masking speech. Target-speech stimuli were Chinese nonsense sentences with 6 words and each word contained 2 syllables. The target-speech sentences were spoken by a female talker (Talker A). They were syntactically ordinary but not semantically meaningful (Yang et al., 2007; Zheng et al., 2016). The speech masker was a 47-s loop of digitally combined continuous recordings for Chinese nonsense sentences (whose keywords did not appear in target sentences), which were spoken by two other young female talkers (female Talkers B and C, Zheng et al., 2016).

All the speech signals were digitally processed with head-related transfer functions (HRTFs) to generate virtual sound images that appeared to occur under free-field listening conditions. The speech signals for a single voice were filtered with the HRTFs to simulate source locations at 90 degree left and 90 degree right to the listener in the azimuth, respectively (for details see Zheng et al., 2016). Both the PSS and PSC perceptual effects were based on the auditory preference effect (Fig. 1; also see Li et al., 2004). Under the PSC condition, the image of target speech and that of masking speech were perceived as coming from the same loudspeaker positions and the target speech was considerably masked by the masking speech. Under the PSS condition, both the image of target speech and that of masking speech were perceived as coming from different loudspeaker position, leading to that the target speech was released from informational masking (Fig. 1).

Acoustic speech stimuli were presented through a magnetic resonance-compatible pneumatic headphone system (SAMRTEC, Guangzhou, China) driven by

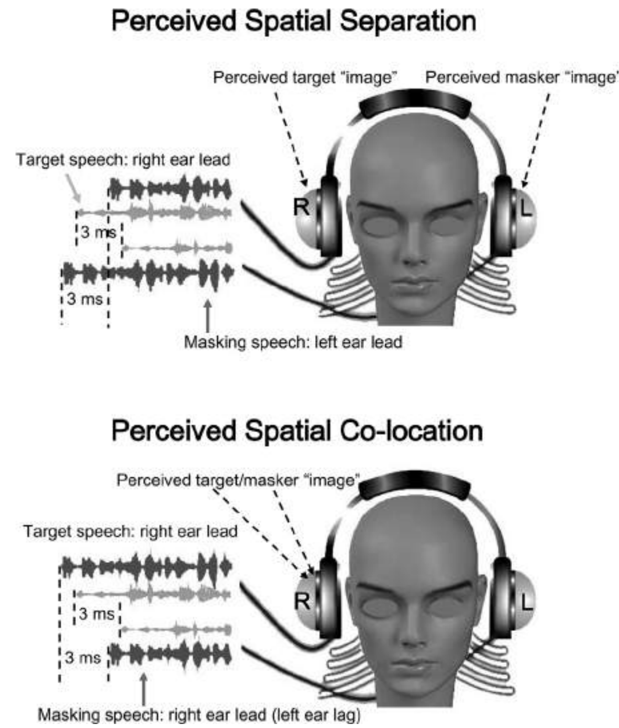

Fig. 1. Based on both the auditory precedence-effect paradigm and the head-related transfer function (HRTF), the target speech and masking speech were simulated as being presented by each of the two spatially separated "loudspeakers" in the frontal field with the inter-source interval of 3 ms. For example, under the perceived spatial separation (PSS) condition (upper panel), when the onset of the target sound presented from the left headphone led that from the right headphone by 3 ms, and the onset of the masker sound presented from the left headphone lagged behind that from the right headphone by 3 ms, due to the precedence effect, the perceptually fused target image was perceived as coming from the left location and the perceptually fused masker image was perceived as coming from the right location. Also, under the perceived spatial co-location (PSC) condition (lower panel), both the onset of the target sound and that of the masker sound presented from the left headphone either led or lagged behind those from the right headphone by 3 ms, leading to a perceptually fused target sound "image" and a perceptually fused masker "image" as coming from the same location.

Presentation software (Version 0.70). The target sound-pressure level was 90 dB SPL (before attenuation by earplugs) and the signal-to-masker ratio (SMR) was set at −4 dB.

Imaging acquisition

fMRI scanning was performed using a 3.0-Tesla Philips Achieva MRI scanner (Veenpluis 4–6, 5680 DA Best, Netherlands) at the Guangzhou Brain Hospital MRI Facility. Blood-oxygen-level-dependent (BOLD) gradient echo-planar images ($64 \times 64 \times 33$ matrix with $3.44 \times 3.44 \times 4.6\ mm^3$ spatial resolution, acquisition time = 2000 ms, time to repeat = 9000 ms, echo time = 30 ms, flip angle = 90°, field of view = $211 \times 211\ mm^2$) were first acquired. A T1-weighted structural images ($256 \times 256 \times 188$ matrix with the spatial resolution of $1 \times 1 \times 1\ mm^3$, repetition time = 8.2 ms, echo time = 3.8 ms, flip angle = 7°) was subsequently obtained.

Design and procedures

The whole scanning course consisted of an 8-min run for localization of the auditory cortex, 2 identical 10-min functional scanning runs for target speech identification against the masking speech and an 8-min structure-scanning run. An event-related fMRI design was used for the functional run.

A total of 61 volumes were acquired from each participant over the first scanning run for localization of the auditory cortex. The target speech with 0 ms of interaural time delay and the silence (rest) were presented alternately 500 ms after the scanning phase. In a scanning trial with the target-speech condition, the speech stimulus was presented in quiet 800 ms after the last scanning trial. The duration of the speech sound was 3200 ms (Fig. 2). In the auditory-cortex-localizing task, thirty images were collected for the speech condition and 30 images were collected for the silence condition, plus 1 single dummy image at the beginning of this run (which was discarded from analyses).

There were 61 scanning trials for each of the 2 functional runs with a single dummy image obtained at the beginning (not included in data analyses) of each run and with 60 experimental trials (20 trials for each of the 3 conditions: PSS, PSC, and baseline stimulation) (Fig. 2). The baseline-stimulation condition contained the masking speech only. For an individual participant, the 60 trials across the 3 conditions were presented with a random order. Across the 2 functional scanning runs, 120 volumes in total were acquired and included in data analyses for each participant. Forty images were collected for each condition.

The sparse-imaging technique (Hall et al., 1999) was used to avoid the effect of machine noise on image data collection: Speech stimuli were presented only during the silent period of the scanner between successive scans (Fig. 2). In each trial the midpoint of the speech stimulus was presented 4100 ms prior to the onset of the next scanning, ensuring that the hemodynamic responses evoked by the speech stimulus peaked within the scanning period (Wild et al., 2012).

In a scanning trial with either the PSS or PSC condition (Fig. 2), the two-talker masker was presented in quiet 800 ms after the last scanning trial. About 1 s later, the target speech was presented. Then the target speech terminated with the masker. In a scanning trial with the baseline condition, only the masker (without target presentation) was presented 800 ms after the last scanning trial with a duration of 4200 ms. To maintain participants' attention to target speech, participants were instructed either to press the left button on a response box with their right index finger if they heard a target speech (with a relative lower sound volume) or to press the right button of the response box if they did not.

All participants were screened for MR safety prior to scanning. A brief training was conducted to ensure that participants understood the instruction and knew how to conduct their button-press responses. Speech sentences used in training were different from those in experimental scanning.

fMRI data preprocessing

All fMRI data were processed and analyzed using the Statistical Parametric Mapping (SPM8, the Wellcome Trust Centre for Neuroimaging, London, UK). Functional images were preprocessed including realignment (correction for head movements), co-registration to the anatomical image, warping into standard Montreal Neurological Institute (MNI) space (re-sampling to a voxel size of $3.0 \times 3.0 \times 4.0\ mm^3$), and spatial smoothing with a Gaussian kernel with 8-mm full-width at half maximum (FWHM). No slice timing was used due to the long TR of this sparse-imaging paradigm.

A model with two levels was used in statistical analyses in SPM8. At the first level, the onsets and durations of each session were modeled using a General Linear Model (GLM) according to the condition types. For the first run for localizing the auditory cortex, two conditions (speech presentation, silence) were included in the model using the canonical hemodynamic response function (HRF). At the second level, contrast images of 'speech > silence' from the first-level analysis in each participant were entered into the second-level one-sample t test in the healthy control group and the patient group separately, and then in all participants (controls and patients were pooled together). The peak signals that were statistically significant at the p value less than 0.05 [voxel-based family-wise error (FWE) corrected, with the activation size larger than 10 contiguous voxels]. The second and third runs were modeled as one run within the design matrix, and three conditions (separation, co-location, and masker only) were included in the model. Six realignment parameters were included to account for residual movement-related effects, and the frame-wise displacement cut-off of 0.5 mm was used (Power et al., 2012).

Generalized psychophysiological interaction (gPPI) analyses

The Psychophysiological interaction (PPI) analysis (Friston et al., 1997) is a method to investigate FC between a seed region and the rest of the brain during a modulation of psychological variables (e.g. task conditions). Generalized PPI (gPPI, http://brainmap.wisc.edu/PPI), which is configured to automatically accommodate more than two task conditions in the same PPI model by spanning the entire experimental space, increases flexibility of statistical modeling and improves fit of model compared to the conventional PPI model (McLaren et al., 2012).

In this study, gPPI analyses were performed to identify which brain regions showing significant FC with the activity of the seed region (i.e., STG) related to (1) the PSS condition versus the PSC condition, (2) the PSS condition versus the masker-only condition, and (3) the PSC condition versus the masker-only condition, respectively. Blood-oxygen-level-dependent (BOLD) signals were extracted from the seed region and deconvolved. The gPPI variable was created by

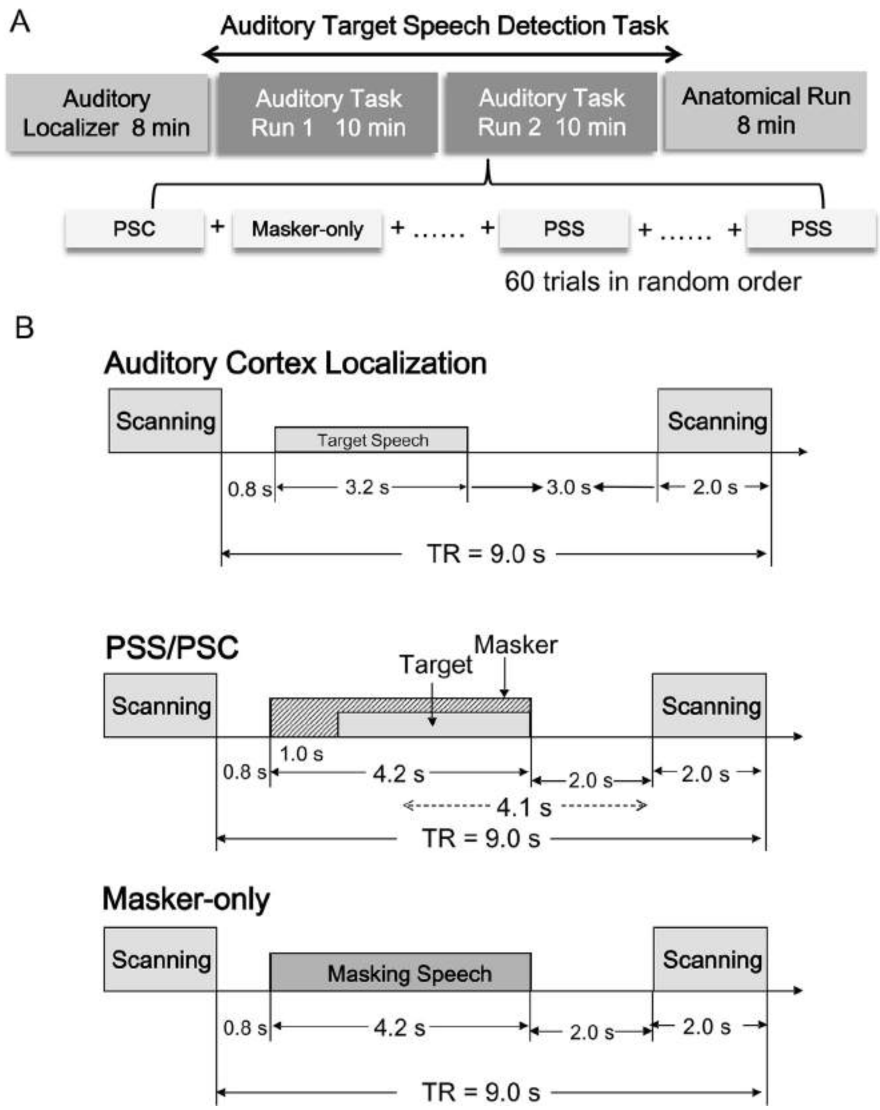

Fig. 2. Illustrations of the fMRI experimental procedures. (A) Both the second experimental run and the third experimental run comprised 20 trials for each of the three listening conditions (PSS, PSC, and baseline (masker-only)) that were presented in random order for a participant. (B) For a scanning trial with either the PSS or PSC condition, the masking-speech and target-speech stimuli were presented 800 ms and 1800 ms after the end of the previous scanning, respectively. The target and the masker terminated at the same time. The midpoint of the auditory stimulus was presented 4.1 s prior to scanning. TR = Time to Repeat; TA = Acquisition Time.

multiplying the signal of the seed with the task condition regressors (PSS, PSC, masker-only).

More in detail, the centers of the seeds were defined based on peak coordinates of bilateral STG in the auditory cortex-localizing task: (1) left STG (x, y, $z = -42$, -25, 6 with $T = 10.96$ for healthy participants; x, y, $z = -54$, -1, -6 with $T = 11.98$ for patient participants; x, y, $z = -48$, -16, 2 with $T = 14.31$ for all participants); and (2) right STG (x, y, $z = 54$, -13, -2 with $T = 9.34$ for healthy participants; x, y, $z = 60$, -10, -2 with $T = 11.68$ for patient participants; x, y, $z = 54$, -13, -2 with $T = 14.48$ for all participants (Fig. 3). A seed region in each participant was defined as a sphere with 5-mm radius centered at the peak voxel. In this study, the peak coordinates of seed regions of STG were obtained from pooled data across participants, ensuring that the same seed regions were used for each participant and the results were comparable between the two participant groups.

For the first-level analyses, for each participant, a separate gPPI model was estimated for each seed. The gPPI generated seven regressors: three for the task conditions, one for the seed, and three for the seed × condition interaction. The obtained interaction variable was convolved with the canonical hemodynamic response function (HRF) to associate it with the BOLD level (Zhang et al., 2016). One additional regressor for correct or incorrect button-press response was included in the GLM for each participant because people with schizophrenia had lower percent correct-response in target speech detection than healthy controls (75.6% for healthy controls and 65.8% for people with schizophrenia, Zheng et al., 2016). The contrast PSS > PSC was created for each gPPI model corresponding to a seed by subtracting the gPPI interaction regressor of the PSC condition from the interaction regressor of the PSS condition. Similarly, the contrast of PSS > masker-only was obtained for each gPPI model corresponding to a seed by subtracting the gPPI interaction regressor of the masker-only condition from the interaction regressor of the PSS condition, and the contrast of PSC > masker-only was obtained for each gPPI model corresponding to a seed by subtracting the gPPI interaction regressor of the masker condition from the interaction regressor of the PSC condition.

For the second-level analyses, the individual contrast images, which reflected the effects of PPI between the seed regions and other brain areas, were subsequently subjected to the one-sample t tests in each of the participant groups to identify the brain areas showing co-variation with the activity of the seed regions in analyses of the (1) PSS condition versus the PSC condition, (2) PSS condition versus the masker-only condition, and (3) PSC condition versus the masker-only condition, respectively. Then individual participants' contrast images were entered into the second-level two-sample t tests for group comparisons. A threshold was set to $p < 0.05$ FWE corrected on cluster level with a cluster-defining threshold (CDT) of $p < 0.001$, uncorrected.

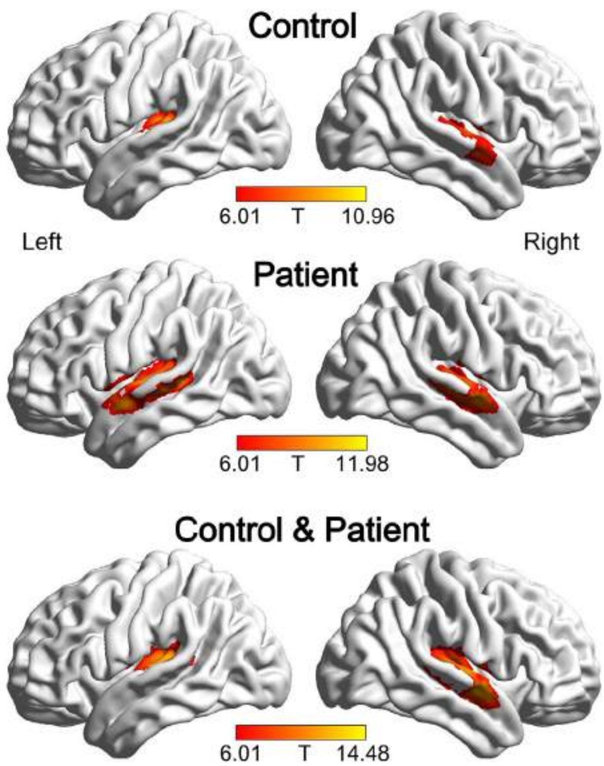

Fig. 3. Brain regions activated by the contrast of "target speech > rest" in the healthy control group (upper panel) and those in the patient group (lower panel). The activation map was thresholded at $p < 0.05$ (FWE corrected with an extend threshold of more than 20 voxels) and overlaid on the template from BrainNet Viewer (http://www.nitrc.org/projects/bnv/). Peak MNI coordinates of (1) left STG ($x, y, z = -42, -25, 6$ with $T = 10.96$ for healthy participants; $x, y, z = -54, -1, -6$ with $T = 11.98$ for patient participants; $x, y, z = -48, -16, 2$ with $T = 14.31$ for all participants); and (2) right STG ($x, y, z = 54, -13, -2$ with $T = 9.34$ for healthy participants; $x, y, z = 60, -10, -2$ with $T = 11.68$ for patient participants; $x, y, z = 54, -13, -2$ with $T = 14.48$ for all participants) for each group were localized.

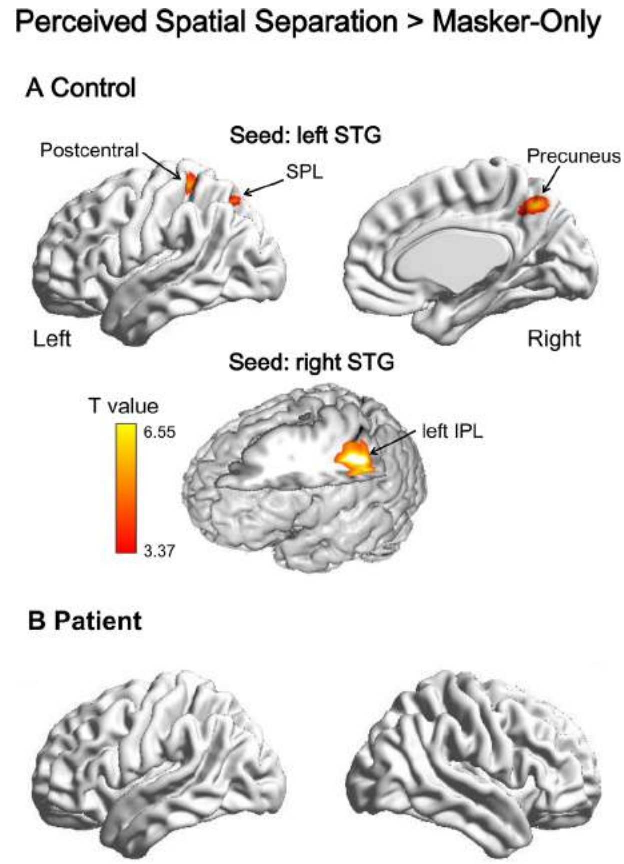

Fig. 4. Psychophysiological interaction (PPI) analyses of functional connectivity of the superior temporal gyrus (STG) associated with the "perceived spatial separation > masker-only" contrast in healthy controls (panel A) and people with schizophrenia (panel B). A cluster-defining threshold (CDT) of $p = 0.001$ ($T = 3.37$ for control group) and a cluster based FWE-corrected threshold of $p = 0.05$ was used. The map was overlaid on the template from BrainNet Viewer (http://www.nitrc.org/projects/bnv/) and Mango (http://rii.uthscsa.edu/mango//index.html). IPL = inferior parietal lobule; SPL = superior parietal lobule; STG = superior temporal gyrus.

Note that data related to the unmasking effect of PSS on target-speech recognition from the same participants have been reported elsewhere (Zheng et al., 2016). The current study had its focus only on FC of the STG during target-speech listening against informational speech masking.

RESULTS

Functional connectivity of the STG in healthy participants and participants with schizophrenia

For the contrast of the PSS condition against the masker-only condition, enhanced FC of the left STG in healthy controls was observed with the left postcentral ($x, y, z = -30, -37, 62$ with $T = 6.06$), right precuneus ($x, y, z = 6, -64, 46$ with $T = 5.69$) and left superior parietal lobule ($x, y, z = -18, -70, 50$ with $T = 5.04$) (upper panel of Fig. 4A). Enhanced FC of the right STG was observed with the left inferior parietal lobule (IPL; $x, y, z = -30, -40, 30$ with $T = 6.55$) (lower panel of Fig. 4A). In participants with schizophrenia, however, no significant changes in FC of the left STG or right STG were observed at the threshold of $p < 0.05$ (cluster-wise FWE corrected).

For the contrast of the PSC condition against the masker-only condition, enhanced FC of the left STG in healthy participants was observed with the left superior parietal lobule ($x, y, z = -18, -64, 42$ with $T = 8.55$) and right precuneus ($x, y, z = 18, -61, 42$ with $T = 7.73$) (Fig. 5A). No significantly enhanced FC of the right STG was observed. In participants with schizophrenia, significantly enhanced FC of the left STG was observed with the left middle frontal gyrus (MFG; $x, y, z = -30, 2, 34$ with $T = 5.41$) (upper panel of Fig. 5B), and enhanced FC of the right STG was observed with the left MCC ($x, y, z = -12, 11, 42$ with $T = 7.56$), right opercularis of inferior frontal gyrus (OperIFG; $x, y, z = 48, 14, 14$ with $T = 6.92$), and right paracentral lobule ($x, y, z = 9, -28, 70$ with $T = 6.61$) (lower panel of Fig. 5B).

For the contrast of the PSS condition against the PSC condition, no significantly enhanced or reduced FC of either the left or right STG was observed in healthy

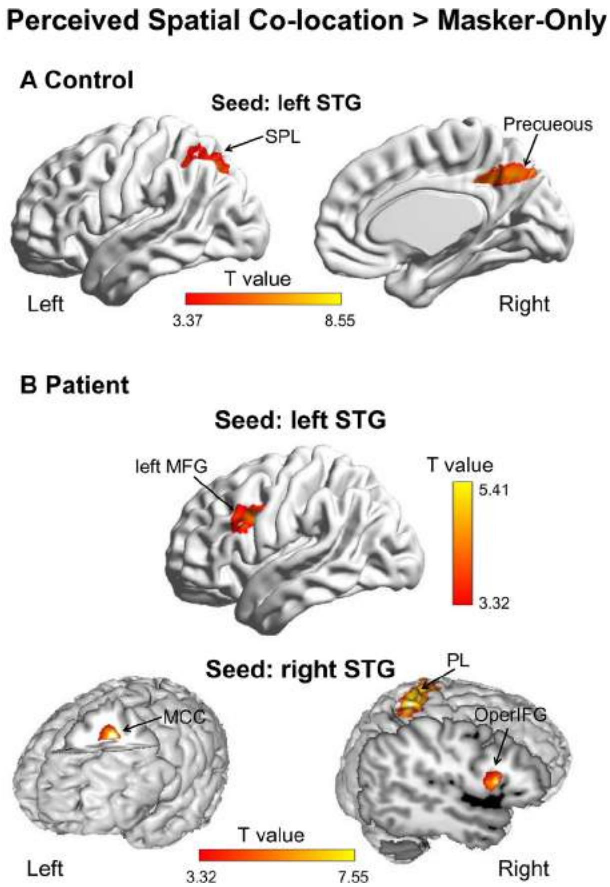

Fig. 5. Psychophysiological interaction (PPI) analyses of functional connectivity of the superior temporal gyrus (STG) associated with the "perceived spatial co-location > masker-only" contrast in healthy controls (panel A) and people with schizophrenia (panel B). A cluster-defining threshold (CDT) of p = 0.001 (T = 3.37 for control group and T = 3.31 for patient group) and a FWE-corrected threshold of p = 0.05 was used. The map was overlaid on the template from BrainNet Viewer and Mango Software. MFG = middle frontal gyrus; OperIFG = opercularis of inferior frontal gyrus; PL = paracentral lobule; SPL = superior parietal lobule; STG = superior temporal gyrus.

participants and participants with schizophrenia (at the threshold of $p < 0.001$, uncorrected), suggesting that FC of the STG was not involved in the unmasking effect of perceptual spatial separation on target-speech listening.

Difference in functional connectivity of the STG between healthy participants and participants with schizophrenia

The results of direct comparisons between healthy controls and participants with schizophrenia showed that (1) for the contrast of the PSS condition against the masker-only condition, the left STG exhibited reduced FC with the bilateral precuneus (left precuneus: x, y, z = −9, −64, 46 with T = 4.36; right precuneus: x, y, z = 6, −64, 46 with T = 4.40), right SPL (x, y, z = 15, −58, 66 with T = 4.32), and right supplementary motor area (SMA; x, y, z = 6, −28, 54 with T = 4.18) in participants with schizophrenia than that in healthy participants; (2) for the contrast of the PSC condition against the masker-only condition, the left STG exhibited reduced FC with the bilateral precuneus (left precuneus: x, y, z = −9, −67, 42 with T = 4.61; right precuneus: x, y, z = 12, −58, 42 with T = 4.77) and right SPL (x, y, z = 21, −61, 54 with T = 4.58) in participants with schizophrenia than that in healthy participants. No enhanced or reduced FC of the right STG was observed (Fig. 6).

DISCUSSION

Previous studies have shown that speech-evoked activation of the STG is enhanced by introducing masking speech, suggesting that the STG is involved in overcoming informational masking-induced difficulties in speech listening (Scott et al., 2004, 2009; Nakai et al., 2005; Scott and McGettigan, 2013; Dole et al., 2014; Evans et al., 2016). It has been known that the speech-recognition performance under cocktail-party listening conditions (with multiple talkers) is poorer in listeners with schizophrenia than in healthy listeners (Wu et al., 2012, 2013, 2017a,b; Zheng et al., 2016). This study for the first time investigated whether FC of the STG for target-speech listening under informational-masking conditions alters in listeners with schizophrenia.

The results of this study showed that although PPI analyses disclosed that neither healthy listeners nor listeners with schizophrenia exhibited significant changes in FC of the STG for the contrast of the PSS condition against the PSC condition (suggesting that FC of the STG is not involved in the unmasking effect on speech recognition when the listening condition is shifted from the PSC condition to the PSS condition), the contrast of either the PSS condition or the PSC condition against the masker-only condition revealed the involvement of FC of the STG in speech listening against informational masking.

In healthy participants, the contrast of the PSS condition against the masker-only condition revealed enhanced FC of the left STG with the left postcentral cortex, the left SPL and right precuneus, and enhanced FC of the right STG with the left IPL. Also, the contrast of the PSC condition against the masker-only condition revealed enhanced FC of the left STG with the left SPL and right precuneus.

Our previous studies (using the same participants) have shown that the SPL is the critical region that is involved in the unmasking effect of introducing the PSS condition on speech recognition against informational masking (Zheng et al., 2016). Also, other previous studies have shown that under "cocktail-party" listening conditions, the SPL is involved in directing attention to one particular talker (Hill and Miller, 2010), processing of spatial attributes (Renier et al., 2009), and suppressing irrelevant distracters to ensure accurate target selection in the competition between target and distracters (Pollmann et al., 2003; Krueger et al., 2007; Zheng et al., 2016). Moreover, the precuneus is involved in computing the exact spatial location of the target sound source (Zundorf et al., 2013), and the IPL is involved in sensor–motor integration

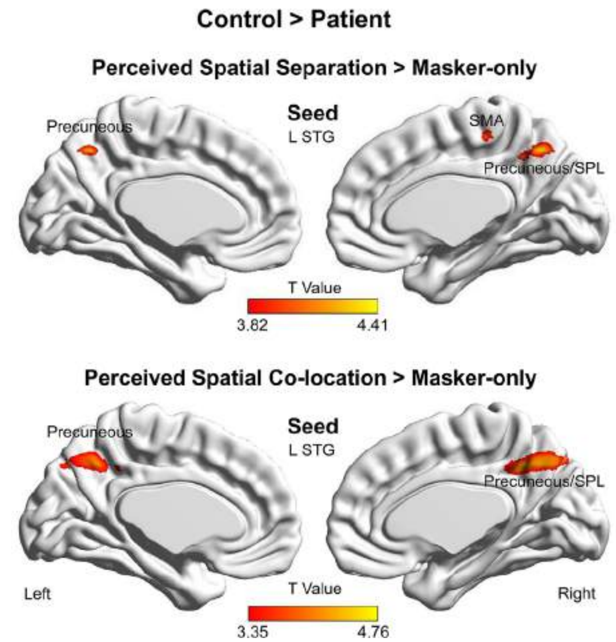

Fig. 6. Differences in PPI analyses of functional connectivity of left STG between healthy controls and people with schizophrenia. The maps are displayed on the group-average structural image. A cluster-defining threshold (CDT) of $p = 0.001$ ($T = 3.82$ for the "PSS > masker-only" condition and $T = 3.35$ for the "PSC > masker-only" condition) and a FWE-corrected threshold of $p = 0.05$ was used. SMA = supplementary motor area; SPL = superior parietal lobule.

during speech perception (Callan et al., 2004; Wilson and Iacoboni, 2006; Du et al., 2014). Thus, FC of the STG with these parietal areas (SPL, precuneus, and IPL) plays a role in facilitating speech listening against informational speech masking by integrating various processes including directing attention to one particular talker, processing of spatial attributes, suppressing irrelevant distracters, sound localization, and sensor–motor interaction during speech perception.

In this study, participants with schizophrenia, however, exhibited their specific FC patterns of the STG. Enhanced FC of the STG was observed with the left MCC, right OperIFG, and right paracentral lobule. Normally, these brain regions are related to speech processing, error detection, response selection, attention control, and working memory (Vouloumanos et al., 2001; Carreiras et al., 2007; Shackman et al., 2011; Zhang et al., 2011; Apps et al., 2013). More in detail, for example, the MCC is activated during a variety of cognitive tasks including conflict monitoring, error detection, response selection and attention control (Shackman et al., 2011; Apps et al., 2013). The right IFG is involved in not only detection of speech stimuli (Vouloumanos et al., 2001) but also speech production including lexical decision (Carreiras et al., 2007). Thus, we propose that the schizophrenia-related patterns of FC of the STG reflect a schizophrenia-related compensatory strategy for the functional impairments of the parietal regions (e.g., SPL and precuneus, Zheng et al., 2016). Obviously, the schizophrenia-induced compensatory strategy is not sufficiently effective in maintaining speech recognition against informational masking. The compensatory mechanism specifically underlying schizophrenia (Tan et al., 2007) is an important issue and needs further investigation.

Notably, compared to healthy listeners, people with schizophrenia exhibit significantly reduced FC of left STG with parietal regions (SPL/precuneus) and SMA during target-speech listening against informational masking. As mentioned above, FC of the STG with parietal areas including SPL and precuneus facilitates speech listening against informational speech masking in healthy people by integrating directing attention to one particular talker, processing spatial attributes, and suppressing irrelevant distracters. The SMA plays a role in planning, preparing, controlling and executing complex movements (Nachev et al., 2008; Price, 2012). Thus, the reduced FC of STG with precuneus, SPL, and SMA may be related to the poorer performance in target-speech identification against speech masking in people with schizophrenia due to dysfunction in integrating the processes of sound localization, target-attention, masker suppression, and motor control.

In summary, this study suggests that FC of the STG with the parietal areas (including SPL and precuneus) normally underlies the listening of target speech against informational speech masking by suppressing masking signals, enhancing attention to sound location, and regulating speech motor processing. The schizophrenia-associated pattern of FC of the STG with the MCC, OperIFG, and the paracentral lobule may reflect the schizophrenia-related neural compensatory strategy. The reduced FC of the STG with the parietal areas in people with schizophrenia may be associated with the increased vulnerability to informational speech masking.

Acknowledgments—This work was supported by the National Natural Science Foundation of China (81671334, 81601168, 31170985), Planed Science and Technology Projects of Guangzhou (2014Y2-00105), Guangzhou Municipal Key Discipline in Medicine for Guangzhou Brain Hospital (GBH2014-ZD06, GBH2014-QN04), the Chinese National Key Clinical Program in Psychiatry to Guangzhou Brain Hospital (201201004), Beijing Municipal Science & Tech Commission (Z161100002616017), and the China Postdoctoral Science Foundation Program (2016T90050).

REFERENCES

Apps MA, Lockwood PL, Balsters JH (2013) The role of the midcingulate cortex in monitoring others' decisions. Front Neurosci 7:251.

Bidelman GM, Dexter L (2015) Bilinguals at the "cocktail party": dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. Brain Lang 143:32–41.

Braff DL, Light GA (2005) The use of neurophysiological endophenotypes to understand the genetic basis of schizophrenia. Dialogues Clin Neurosci 7:125–135.

Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. Neuroimage 22:1182–1194.

Carreiras M, Mechelli A, Estevez A, Price CJ (2007) Brain activation for lexical decision and reading aloud: two sides of the same coin? J Cogn Neurosci 19:433–444.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci USA 109:11854–11859.

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735.

Dole M, Meunier F, Hoen M (2014) Functional correlates of the speech-in-noise perception impairment in dyslexia: an MRI study. Neuropsychologia 60:103–114.

Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci USA 111:7126–7131.

Evans S, Mcgettigan C, Agnew ZK, Rosen S, Scott SK (2016) Getting the cocktail party started: masking effects in speech perception. J Cogn Neurosci 28:483–500.

First MB, Spitzer RL, Miriam G, Williams JBW (1997) Structured clinical interview for DSM-IV axis I disorders : SCID-I: clinical version: administration booklet. American Psychiatric Press.

Freyman RL, Helfer KS, McCall DD, Clifton RK (1999) The role of perceived spatial separation in the unmasking of speech. J Acoust Soc Am 106:3578–3588.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 160:636–645.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) ''Sparse'' temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223.

Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cereb Cortex 20:583–590.

Huang Y, Huang Q, Chen X, Qu T, Wu X, Li L (2008) Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults. Hear Res 244:51–65.

Krueger F, Fischer R, Heinecke A, Hagendorf H (2007) An fMRI investigation into the neural mechanisms of spatial attentional selection in a location-based negative priming task. Brain Res 1174:110–119.

Lee SH, Chung YC, Yang JC, Kim YK, Suh KY (2004) Abnormal speech perception in schizophrenia with auditory hallucinations. Acta Neuropsychiatr 16:154–159.

Li L, Daneman M, Qi JG, Schneider BA (2004) Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? J Exp Psychol Human 30:1077–1091.

McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. Neuroimage 61:1277–1286.

Nachev P, Kennard C, Husain M (2008) Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci 9:856–869.

Nakai T, Kato C, Matsuo K (2005) An FMRI study to investigate auditory attention: a model of the cocktail party phenomenon. Magnet Reson Med 4:75–82.

Pollmann S, Weidner R, Humphreys GW, Olivers CN, Muller K, Lohmann G, Wiggins CJ, Watson DG (2003) Separating distractor rejection and target detection in posterior parietal cortex–an event-related fMRI study of visual marking. Neuroimage 18:310–323.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.

Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62:816–847.

Rakerd B, Aaronson NL, Hartmann WM (2006) Release from speech-on-speech masking by adding a delayed masker at a different location. J Acoust Soc Am 119:1597–1605.

Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP (2009) Multisensory integration of sounds and vibrotactile stimuli in processing streams for ''what'' and ''where''. J Neurosci 29:10950–10960.

Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ (2007) Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophr Res 97:173–183.

Scott SK, McGettigan C (2013) The neural processing of masked speech. Hear Res 303:58–66.

Scott SK, Rosen S, Wickham L, Wise RJ (2004) A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. J Acoust Soc Am 115:813–821.

Scott SK, Rosen S, Beaman CP, Davis JP, Wise RJ (2009) The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. J Acoust Soc Am 125:1737–1743.

Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ (2011) The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat Rev Neurosci 12:154–167.

Si T, Yang J, Shu L, Wang X, Kong Q, Zhou M, Li X (2004) The reliability, validity of PANSS (Chinese version), and its implication. Chin Mental Health J 18:45–47.

Tan HY, Callicott JH, Weinberger DR (2007) Dysfunctional and compensatory prefrontal cortical systems, genes and the pathogenesis of schizophrenia. Cereb Cortex 17(S1):i171–i181.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF (2001) Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci 13:994–1005.

Wild CJ, Davis MH, Johnsrude IS (2012) Human auditory cortex is sensitive to the perceived clarity of speech. Neuroimage 60:1490–1502.

Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage 33:316–325.

Woods SW (2003) Chlorpromazine equivalent doses for the newer atypical antipsychotics. J Clin Psychiat 64:663–667.

Wu X, Wang C, Jing C, Qu H, Li W, Wu Y, Schneider BA, Liang L (2005) The effect of perceived spatial separation on informational masking of Chinese speech. Hear Res 199:1–10.

Wu C, Cao S, Zhou F, Wang C, Wu X, Li L (2012) Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia. Schizophr Res 134:33–41.

Wu C, Li H, Tian Q, Wu X, Wang C, Li L (2013) Disappearance of the unmasking effect of temporally pre-presented lipreading cues on speech recognition in people with chronic schizophrenia. Schizophr Res 150:594–595.

Wu C, Zheng Y, Li J, Zhang B, Li R, Wu H, She S, Liu S, Peng H, Ning Y, Li L (2017b) Activation and functional connectivity of the left inferior temporal gyrus during visual speech priming in healthy listeners and listeners with schizophrenia. Front Neuroscience 11:1–13.

Wu C, Zheng Y, Li J, Wu H, She S, Liu S, Ning Y, Li L (2017a) Brain substrates underlying auditory speech priming in healthy listeners and listeners with schizophrenia. Psychol Med 47:837–852.

Yang Z, Chen J, Huang Q, Wu X, Wu Y, Schneider BA, Li L (2007) The effect of voice cuing on releasing Chinese speech from informational masking. Speech Commun 49:892–904.

Zhang Y, Meyers EM, Bichot NP, Serre T, Poggio TA, Desimone R (2011) Object decoding with attention in inferior temporal cortex. Proc Natl Acad Sci USA 108:8850–8855.

Zhang L, Vander Meer L, Opmeer EM, Marsman JC, Ruhe HG, Aleman A (2016) Altered functional connectivity during self- and close other-reflection in patients with bipolar disorder with past

psychosis and patients with schizophrenia. Neuropsychologia 93:97–105.
Zheng Y, Wu C, Li J, Wu H, She S, Liu S, Mao L, Ning Y, Li L (2016) Brain substrates of perceived spatial separation between speech sources under simulated reverberant listening conditions in schizophrenia. Psychol Med 46:477–491.
Zundorf IC, Lewald J, Karnath HO (2013) Neural correlates of sound localization in complex acoustic environments. PLoS ONE 8: e64259.

(Received 14 March 2017, Accepted 22 June 2017)
(Available online 1 July 2017)