

# Brain substrates underlying auditory speech priming in healthy listeners and listeners with schizophrenia

C. Wu<sup>1,2,3†</sup>, Y. Zheng<sup>4†</sup>, J. Li<sup>4†</sup>, H. Wu<sup>4</sup>, S. She<sup>4</sup>, S. Liu<sup>4</sup>, Y. Ning<sup>4</sup> and L. Li<sup>1,4,5\*</sup>

<sup>1</sup>School of Psychological and Cognitive Sciences, and Beijing Key Laboratory of Behavior and Mental Health, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing, People's Republic of China

<sup>2</sup>School of Life Sciences, Peking University, Beijing, People's Republic of China

<sup>3</sup>School of Psychology, Beijing Normal University, Beijing, People's Republic of China

<sup>4</sup>The Affiliated Brain Hospital of Guangzhou Medical University, Guangzhou, People's Republic of China

<sup>5</sup>Beijing Institute for Brain Disorders, Capital Medical University, Beijing, People's Republic of China

**Background.** Under 'cocktail party' listening conditions, healthy listeners and listeners with schizophrenia can use temporally pre-presented auditory speech-priming (ASP) stimuli to improve target-speech recognition, even though listeners with schizophrenia are more vulnerable to informational speech masking.

**Method.** Using functional magnetic resonance imaging, this study searched for both brain substrates underlying the unmasking effect of ASP in 16 healthy controls and 22 patients with schizophrenia, and brain substrates underlying schizophrenia-related speech-recognition deficits under speech-masking conditions.

**Results.** In both controls and patients, introducing the ASP condition (against the auditory non-speech-priming condition) not only activated the left superior temporal gyrus (STG) and left posterior middle temporal gyrus (pMTG), but also enhanced functional connectivity of the left STG/pMTG with the left caudate. It also enhanced functional connectivity of the left STG/pMTG with the left pars triangularis of the inferior frontal gyrus (TriIFG) in controls and that with the left Rolandic operculum in patients. The strength of functional connectivity between the left STG and left TriIFG was correlated with target-speech recognition under the speech-masking condition in both controls and patients, but reduced in patients.

**Conclusions.** The left STG/pMTG and their ASP-related functional connectivity with both the left caudate and some frontal regions (the left TriIFG in healthy listeners and the left Rolandic operculum in listeners with schizophrenia) are involved in the unmasking effect of ASP, possibly through facilitating the following processes: masker-signal inhibition, target-speech encoding, and speech production. The schizophrenia-related reduction of functional connectivity between the left STG and left TriIFG augments the vulnerability of speech recognition to speech masking.

Received 29 October 2015; Revised 6 October 2016; Accepted 11 October 2016; First published online 29 November 2016

**Key words:** Auditory speech priming, informational masking, pars triangularis of the inferior frontal gyrus, schizophrenia, speech recognition.

## Introduction

One of the difficult perceptual tasks with high perceptual load for human listeners is to recognize speech in a noisy 'cocktail party' speech-listening environment with multiple people talking (Cherry, 1953), because the attended speech is under not only energetic masking that occurs at the peripheral level but also informational masking that interferes with the processing of

the target talker's utterances at more central (i.e. cognitive) levels of processing (Schneider *et al.* 2007).

More specifically, under 'cocktail party' environments with multiple people talking (Cherry, 1953), two types of masking components contribute to the difficulty of speech recognition: energetic masking and informational masking (e.g. Freyman *et al.* 1999; Brungart *et al.* 2001; Li *et al.* 2004; Helfer & Freyman, 2005, 2009; Wu *et al.* 2005; Rakerd *et al.* 2006; Huang *et al.* 2008, 2009; Wu *et al.* 2012b). Energetic masking occurs when peripheral neural activity elicited by a signal is overwhelmed by that elicited by maskers, leading to a degraded or lost neural representation of the signal. Informational masking occurs when signals and maskers are similar along some informational dimensions, particularly when both target signals and maskers are speech sounds, causing confusion

\* Address for correspondence: L. Li, School of Psychological and Cognitive Sciences, Peking University, Beijing 100080, People's Republic of China.

(Email: liangli@pku.edu.cn)

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

between the target and masker and/or uncertainty regarding the target. Relative to energetic masking, informational masking involves more higher-order central processing (Schneider *et al.* 2007).

Under such adverse listening conditions, it is not unusual for a listener to ask the attended speaker repeat the sentence(s). The beneficial 'say-it-again' effect is caused by using some perceptually and/or cognitively unmasking primes, including prior knowledge about part of target-sentence content (Freyman *et al.* 2004; Yang *et al.* 2007; Helfer & Freyman, 2009; Wu *et al.* 2012b), familiarity with the target talker's voice (Yang *et al.* 2007; Huang *et al.* 2010), and visual working memory of the speaker's movements of speech articulators (Wu *et al.* 2013a, b).

In studies of auditory speech priming (ASP), when a meaningless (nonsense) target sentence with a number of keywords is co-presented with a two-talker speech masker (which causes informational masking of the target sentence at both perceptual and cognitive levels; see Schneider *et al.* 2007), recognition of the last keyword of the target sentence is improved if the early segment of the same sentence (including the first or the first two keywords) recited by the target talker's voice (i.e. the auditory speech prime) is temporally pre-presented in quiet before the target/masker co-presentation (Freyman *et al.* 2004; Yang *et al.* 2007; Wu *et al.* 2012b). Freyman *et al.* (2004) have suggested that the speech prime helps listeners focus attention more quickly on the target speech, thereby facilitating recognition of the last keyword in the target stream against informational masking. Up to date, the brain substrates specifically underlying the unmasking effect of ASP have not been reported in the literature. It has been reported that subjective clarity of degraded (noise-vocoding) speech can be enhanced by prior knowledge of speech content (Davis *et al.* 2005; Sohoglu *et al.* 2014) and the underlying mechanisms include top-down modulation of speech-signal processing at the processing stage of the auditory cortex (Sohoglu *et al.* 2012; Wild *et al.* 2012). It is of interest to know whether the brain substrates underlying auditory prime-induced unmasking of speech are similar to those underlying prior speech-knowledge-induced improvement of clarity of degraded speech.

Compared with that in healthy listeners, the correct percentage of target-keyword recognition in patients with schizophrenia is lower not only under noise-masking conditions but also under speech-masking conditions (Wu *et al.* 2012a, 2013b). More in detail, our previous studies have revealed that the difference in the threshold (in signal:masker ratio, SMR) for recognizing target keywords between people with first-episode schizophrenia and their healthy controls is 1.0 dB when the masker is steady-state speech-spectrum noise, but increases to 1.7 dB when the

masker is two-talker speech. Also, the threshold difference between people with chronic schizophrenia and their healthy controls is 1.9 dB when the masker is the noise, but increases to 3.0 dB when the masker is the speech. Interestingly, although people with schizophrenia perform worse than healthy listeners in speech recognition against informational masking (Wu *et al.* 2012a, 2013b), they retain the ability to use the temporally pre-presented auditory-speech primes to improve (unmask) target-speech recognition under informational masking conditions (Wu *et al.* 2012a). Knowing whether the brain substrates normally underlying the ASP in healthy listeners are also functional in listeners with schizophrenia is important for understanding the nature of the brain plasticity associated with schizophrenia.

It has been reported that people with schizophrenia experience more difficulties in filtering irrelevant/distracting sensory stimuli to prevent information overflow that leads to various cognitive dysfunctions (Gottesman & Gould, 2003; Braff & Light, 2005). They also perform worse than healthy listeners in speech perception particularly under informational speech-masking conditions (Wu *et al.* 2012a, 2013b; Zheng *et al.* 2016). Up to date, marked progress has been made in understanding the brain regions that are related to speech (informational) masking of target speech in healthy people (Scott *et al.* 2004, 2009; Ding & Simon, 2012; Scott & McGettigan, 2013; Evans *et al.* 2016). For example, the study of Evans *et al.* (2016) showed that both the mid-posterior superior temporal gyrus (STG) and the superior temporal sulcus exhibit higher activity as informational content of masking sounds is increased, suggesting that masking speech is processed within the same pathway for processing target speech. However, only a very small number of studies have been reported in the literature on brain substrates underlying perceptual cue (spatial or non-spatial)-induced unmasking of speech recognition in either healthy listeners or listeners with schizophrenia (Zheng *et al.* 2016). Using functional magnetic resonance imaging (fMRI), in both healthy listeners and listeners with schizophrenia, this study was to search for: (1) the brain substrates specifically underlying the unmasking effect of ASP; and (2) the brain mechanisms underlying schizophrenia-related speech-recognition deficits under speech-masking conditions.

## Method

### Participants

With the recruiting criteria used in our previous studies (Wu *et al.* 2012a, 2013b; Zheng *et al.* 2016), participants with schizophrenia, who were diagnosed with

the Structured Clinical Interview for DSM-IV (SCID) (First *et al.* 1996), were recruited in the Guangzhou Brain Hospital. Patients with diagnoses of schizoaffective or other psychotic disorders were not included. Demographics-matched healthy listeners (selected for controls) were recruited from the community around the hospital with the recruiting criteria used previously (Wu *et al.* 2012a, 2013b; Zheng *et al.* 2016). They were telephone interviewed first and then those who passed the interview were screened with the SCID 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 or a family history of psychiatric illness based on self-report.

Note that some potential patient participants were excluded from this study if they had co-morbid diagnoses, substance dependence and/or other conditions that affected experimental tests [including hearing loss, abuse and/or dependence of alcohol and/or drugs, a treatment of electroconvulsive therapy (ECT) within the past 3 months, a treatment of trihexyphenidyl hydrochloride with a dose of more than 6 mg/day, and/or an age younger than 18 years or older than 59 years]. For the purpose of improving sleeping, some of the patient participants received benzodiazepines based on doctors' advice.

Both patient participants and their guarantees gave their written informed consent for participation in this study. Demographics-matched healthy listeners (controls) were recruited from the community around the hospital with the recruiting criteria used previously (Wu *et al.* 2012a). The procedures of this study were approved by the Independent Ethics Committee of the Guangzhou Brain Hospital.

In all, 17 healthy participants and 25 patients participated in the study. One healthy listener and three patients were excluded from data analyses due to their excessive head movement (more than 2.5 mm in translation and/or 2.5° in rotation from the first volume in any axis) during fMRI scanning. The remaining 16 healthy listeners (seven males and nine females) and 22 patients (14 males and 8 females) were included in both behavioral testing and fMRI data analyses. 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. Their first language was Mandarin Chinese.

All patient participants were clinically stable during their participation and they received antipsychotic medications during this study with the average chlorpromazine equivalent of 521 mg/day based on the conversion factors described by Woods (2003). The locally validated version of the Positive and Negative Syndrome Scale Test (Si *et al.* 2004) was conducted on the day of fMRI scanning for all participants. The speech

recognition test was conducted after the fMRI scanning. The characteristics of the patients and healthy participants are shown in online Supplementary Table S1.

### Stimuli and procedures

There were three types of speech stimuli in total: target speech, masking speech, and priming speech. Spoken by a young female talker (Talker A), target speech stimuli were Chinese nonsense phrases with three words and each word contained two syllables. They were syntactically ordinary but not semantically meaningful (see Wu *et al.* 2013b). For example, the English translation of a phrase is '*retire his ocean*' (keywords are in italics). Obviously, the sentence frame provided no contextual support for recognizing keywords. The speech masker was a 47-s loop of digitally combined continuous recordings for Chinese nonsense sentences (whose keywords did not appear in target phrases) spoken by two other young female talkers (Talkers B and C; Yang *et al.* 2007). There were two types of priming conditions: (1) ASP condition – the prime (stimulus) was identical to the target phrase except that the last (second) keyword was replaced by a white noise with both the duration of the longest second keyword among the target phrases and the sound level 10 dB lower than that of target stimuli (following Freyman *et al.* 2004); (2) auditory non-speech priming (ANSP) condition (i.e. the speech-priming control condition) – the prime was a white noise with both the duration of the longest target phrase and the sound level 10 dB lower than that of target stimuli. In addition, during fMRI scanning a non-speech baseline stimulation condition (the controlling condition for both ASP and ANSP conditions) was used with the presentation of white noise whose duration was equal to that of the longest target phrase and whose sound level was 10 dB lower than that of target stimuli.

The whole-course scanning consisted of a 10-min auditory-priming functional run and an 8-min structure-scanning run. An event-related fMRI design was used for the functional run. In total, there were 60 scanning trials for the functional run (20 trials for each of the three conditions: ASP, ANSP, and baseline stimulation). For an individual participant, the 60 trials across the three conditions were presented with a random order. The sparse-imaging strategy was used to avoid the effect of machine scanning noise: Sound stimuli were presented only during the pause period between successive scanning periods (Hall *et al.* 1999). Also, to ensure that stimulus-evoked hemodynamic responses peaked within the scanning period (Wild *et al.* 2012), the stimulus presentation was so temporally positioned that the midpoint of stimulus

presentation occurred about 4200 ms before the onset of the next scanning.

In a scanning trial for examining the priming effect (Fig. 1), either the ASP stimulus or the ANSP stimulus was presented in quiet (without the masker presentation) 600 ms after the offset of the last scanning trial. Immediately after the prime presentation, the target and masker were co-presented and terminated simultaneously. To maintain participants' attention to stimulus presentation, a two-syllable word, either the last keyword in the target speech phrase (with the 50% possibility) or a different word was presented for 500 ms after the co-presentation of the target and masker speech. The participant was instructed to use a button-press with their right index finger to indicate whether the word on screen was the last keyword in the target-speech phrase.

The sound stimuli used in the fMRI scanning experiment were presented through a magnetic resonance-compatible pneumatic headphone system (SAMRTEC, China) driven by Presentation software (version 0.70) without introducing interaural time differences. The target-speech level was 60 dB sound pressure level (SPL) (after attenuation by earplugs) and the SMR was  $-4$  dB. Visual stimuli were presented through a liquid crystal display screen positioned on the head coil (SAMRTEC, China). 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.

A 3.0-Tesla Philips Achieva MRI scanner (Veenpluis 4-6,5680 DA Best, the Netherlands) was used to acquire blood oxygenation level-dependent (BOLD) gradient echo-planar images (spatial resolution:  $64 \times 64 \times 33$  matrix with the voxel size of  $3.44 \times 3.44 \times 4.6$  mm<sup>3</sup>; acquisition time: 2000 ms; time to repeat: 9000 ms; echo time: 30 ms; flip angle: 90°; field of view:  $211 \times 211$  mm<sup>2</sup>). It provided high-resolution T1-weighted structural images ( $256 \times 256 \times 188$  matrix with a spatial resolution of  $1 \times 1 \times 1$  mm<sup>3</sup>, repetition time: 8.2 ms; echo time: 3.8 ms; flip angle: 7°).

### fMRI data processing and analyses

#### Pre-processing

All fMRI data were processed and analysed using Statistical Parametric Mapping software (SPM8; the Wellcome Trust Centre for Neuroimaging, UK). The pre-processing of data includes the following four stages: (1) the functional images were corrected for head movements; (2) the anatomical images were co-registered with the mean realigned images and normalized to the standard template [UCLA Brain Mapping Center (ICBM) space] using the SPM8

unified segmentation routine; (3) all functional images were warped using deformation parameters generated from the normalization process, including re-sampling to a voxel size of  $3.0 \times 3.0 \times 4.0$  mm<sup>3</sup>; and (4) spatial smoothing was conducted using a Gaussian kernel with 8 mm full-width at half maximum (FWHM). Due to the long time to repeat this sparse-imaging paradigm, no slice timing was necessary.

#### Random-effect analyses

Random-effect analyses contained two processing levels. At the first level, the onsets and durations for the functional run were modeled using a general linear model according to the condition types. The three stimulation conditions (ASP, ANSP and baseline) were included in the model. Six realignment parameters of head movement were included to account for residual movement-related effects (Friston *et al.* 1996). At the second level, random-effect analyses were conducted based on the statistical parameter maps from each individual participant to allow population inference.

To assess the overall main effects of participant group and prime type, and the interaction, both the contrast image of 'ASP > baseline' and that of 'ANSP > baseline' from each participant were entered into a second-level  $2 \times 2$  (group  $\times$  prime type) full-factor analysis of variance (ANOVA). *F* contrasts were used.

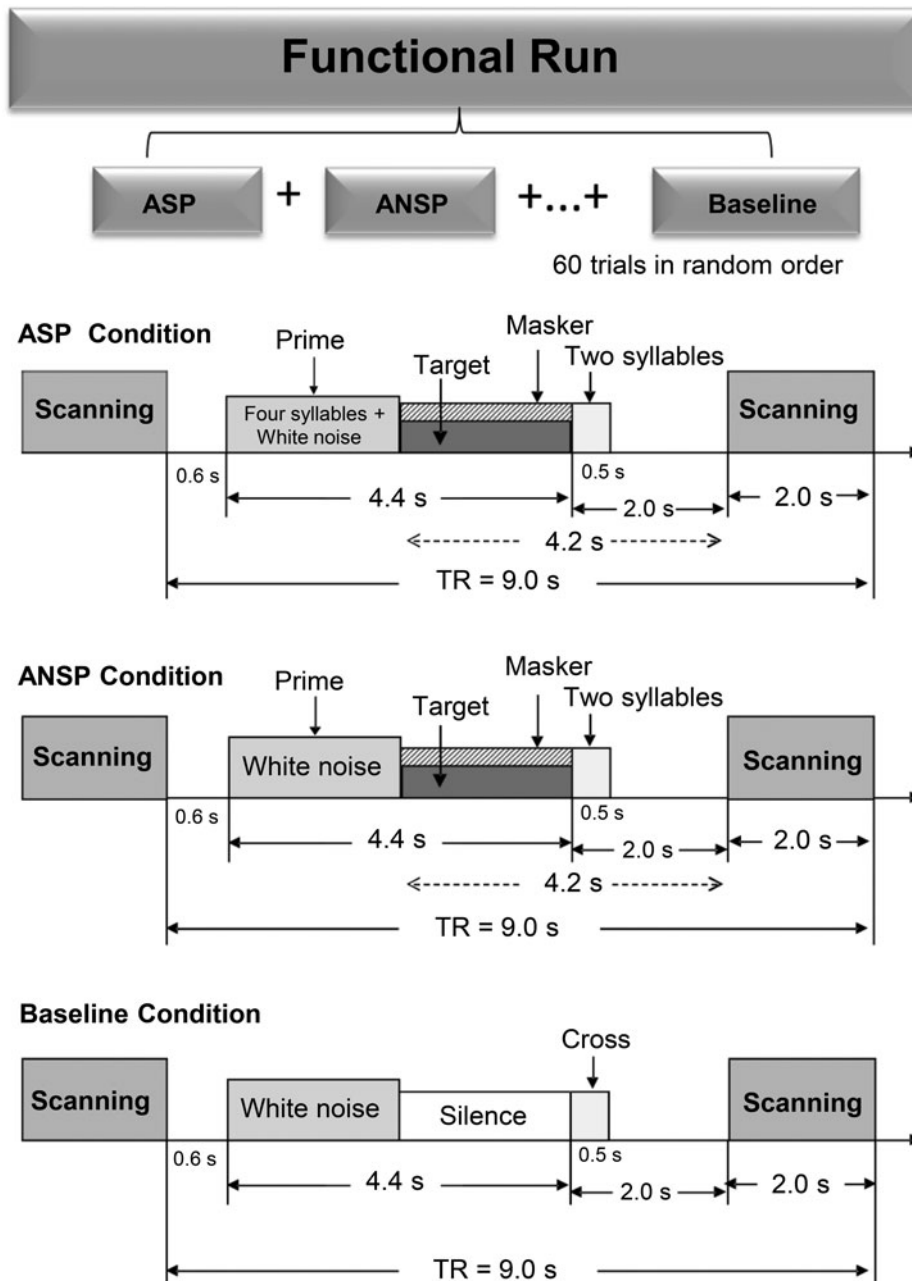
To localize the brain regions activated by speech stimulation conditions, the contrast images of 'ASP > baseline' and that of 'ANSP > baseline' from the first-level analyses in each participant were entered into second-level one-sample *t* tests for the patient group and the healthy-control group, respectively.

To reveal the brain regions related to the 'ASP effect (ASP > ANSP)' in each participant group, contrast images of 'ASP > ANSP' from the first-level analyses in each participant were entered into the second-level group one-sample *t* tests in the healthy control group and the patient group, separately. For whole-brain analyses, peak signals that were statistically significant at the *p* value less than 0.05 [family-wise error (FWE) corrected] were reported.

#### Psychophysiological interaction (PPI) analyses

PPI analyses (Friston *et al.* 1997) were performed to identify the brain regions showing significantly increased functional connectivity with the most critical brain structures (seeds) related to the ASP condition compared with the ANSP condition. The coordinates of the peak voxel from the contrast of 'ASP > ANSP' in random-effect analyses were used as the landmarks for the individual seed voxels. A seed region in each participant was defined as a sphere with a 5 mm radius centered at the peak voxel. The time series of





**Fig. 1.** Illustration of a functional scanning run that comprised 60 trials [20 trials for each of the three conditions: auditory speech priming (ASP), auditory non-speech priming (ANSP), and non-speech stimulation baseline] presented in random order. Sparse temporal sampling scanning was used. Trial structures of each of the three conditions for the functional run are illustrated separately. The temporal midpoint of the sound stimulus was presented 4200 ms prior to the onset of the next scanning. TR, Time to repeat.

seed regions were then extracted, and the PPI regressors, which reflected the interaction between psychological variable (ASP *v.* ANSP) and the activation time course of the seed regions, were calculated.

The individual contrast images, which reflected the effects of PPI between the seed regions and other brain areas, were subsequently subjected to the second-level one-sample *t* tests in each of the

participant groups to identify the brain regions showing increased co-variation with the activity of the seed regions in analyses of the ASP condition against the ANSP condition. Then individual participants' contrast images were entered into the second-level two-sample *t* tests for group comparisons. In PPI analyses, peak signals that were statistically significant at a *p* value less than 0.05 (FDR corrected) were reported.

### Functional connectivity analyses (partial correlation)

It has been established that partial correlation can be used as an effective measure of functional connectivity between a given pair of brain regions by attenuating the contribution of other sources of covariance (Hampson *et al.* 2002; Liu *et al.* 2008, 2014). In this study, to search for schizophrenia-related changes in functional connectivity of speech processing-related brain regions, both the activation clusters from 'ASP > baseline' and those from 'ANSP > baseline' SPM files in healthy controls and patient participants were extracted and used as the seed regions for examining the functional connectivity ( $p < 0.05$ , FWE corrected) (MarsBaR: region of interest toolbox for SPM; <http://marsbar.sourceforge.net/>). The clusters representing the same brain region activated by 'ASP > baseline' and 'ANSP > baseline' in patients and controls were combined as the seed for functional connectivity analyses of this brain region. The correlation coefficient obtained by partial correlation between a pair of variables after filtering out contributions from all other variables was included in the dataset (Salvador *et al.* 2005; Liu *et al.* 2008, 2014) and converted to Z-scores with the Fischer transform as the strengths of functional connectivity. Thus, significant differences in functional connectivity between healthy controls and patients were identified by significant differences in Z-score (see Liu *et al.* 2014).

### Correlation between speech-recognition performance and functional connectivity

Spearman correlation analyses were performed using SPSS 16.0 software to investigate the correlation between the behavioral performance (averaged percentage correct of target speech recognition across the ASP and ANSP listening conditions) and the strength of functional connectivity (Z-scores of functional connectivity) of a pair of brain regions. The null hypothesis was rejected at the level of 0.05.

### Behavioral testing

The behavioral testing was conducted after the fMRI scanning experiment. Acoustic signals, calibrated by a sound-level meter (AUDit and System 824; Larson Davis, Inc., USA), were delivered from a notebook-computer sound card (ATI SB450 AC97) and presented to participants via earphones (model HDA 600). The target speech level was 60 dB SPL. The SPL of the speech masker was adjusted to produce two SMRs:  $-4$  and  $-8$  dB. There were two within-subject variables: (1) priming type (ASP, ANSP), and (2) SMR ( $-8$ ,  $-4$  dB). For each participant, there were four testing conditions and 20 trials (also 20 target-sentence

presentations) for each condition. The presentation order for the four combinations of priming type and SMR were partially counterbalanced across participants using a Latin square order.

In a trial, the participant, who was seated at the center of a quiet room in the hospital, pressed the 'Enter' key on a computer keyboard to start the presentation of the auditory priming stimulus. Either an ASP stimulus or an ANSP stimulus was presented in quiet (without the masker presentation) after the key press. Immediately after the prime-presentation phrase, the target and masker were presented and terminated simultaneously. After the masker/target co-presentation was finished, the participant was instructed to loudly repeat the whole target phrases as best as he/she could. The experimenters, who sat quietly behind the participant, scored whether each of the two syllables for each of the two keywords in the target phrases had been identified correctly.

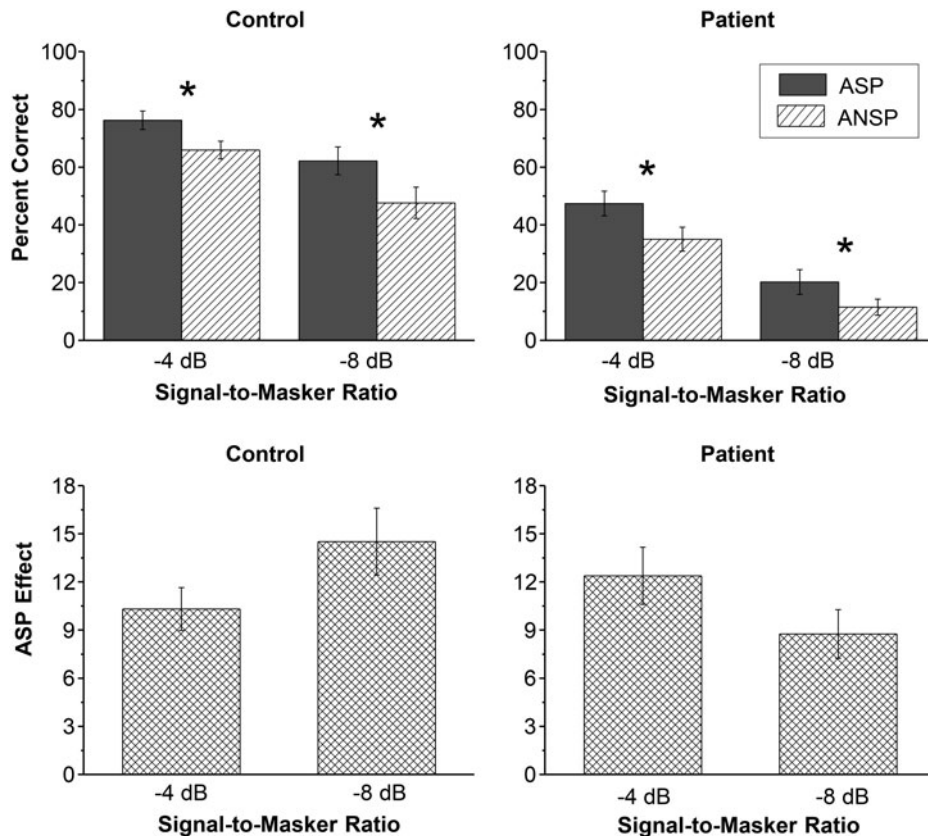
Target speech recognition performance was valued by the percentage of correct response of the second keywords in target phrases. The percentage correct scores were entered into ANOVAs or Spearman correlation with the strength of functional connectivity (Z-scores of functional connectivity) of brain region pairs using SPSS 16.0 software (USA). The null hypothesis was rejected at the level of 0.05.

## Results

### The speech-priming effect and the participant-group effect on speech recognition

Fig. 2 shows comparisons in group-mean percentage correct recognition of the last target keyword (top panels) and those in group-mean ASP-induced speech-recognition improvement (bottom panels, the ASP effect: the difference in percentage correct between the ASP condition and the ANSP condition) between healthy listeners and patients when the SMR was either  $-4$  or  $-8$  dB.

To statistically examine the differences in percentage of correct identification of target speech across experimental conditions, a 2 (group: control, patient) by 2 (priming type: ASP, ANSP) by 2 (SMR:  $-4$ ,  $-8$  dB) three-way ANOVA showed that the main effects of group ( $F_{1,144} = 138.76$ ,  $p < 0.001$ ), priming type ( $F_{1,144} = 13.27$ ,  $p < 0.001$ ) and SMR ( $F_{1,144} = 52.27$ ,  $p < 0.001$ ) were all significant. However, all the two-way interactions and the three-way interaction were not significant. *Post-hoc* tests further confirmed that consistent with previous studies (Wu *et al.* 2012a, 2013b), patients performed worse in speech recognition than healthy controls (Fig. 2 top panels) when SMR was either  $-4$  dB (for each of the two priming conditions,  $p < 0.001$ )



**Fig. 2.** Top panels: Comparisons in group-mean percentage-correct recognition of target keywords between the healthy-listener group and the patient group when the signal:masker ratio (SMR) was either  $-4$  or  $-8$  dB. Bottom panels: Comparisons in group-mean auditory speech priming (ASP)-induced improvement in recognition of target keywords between the healthy listener group and the patient group when the SMR was either  $-4$  or  $-8$  dB. ANSP, Auditory non-speech priming. \*  $p < 0.05$ .

or  $-8$  dB (for each of the two priming conditions,  $p < 0.001$ ).

Moreover, similar to healthy listeners (Fig. 2 bottom panels), patients were able to use the ASP stimulus to improve their target-speech recognition when the SMR was either  $-4$  dB ( $F_{1,72} = 6.880$ ,  $p = 0.011$ ) or  $-8$  dB ( $F_{1,72} = 7.192$ ,  $p = 0.009$ ).

#### Brain regions activated by speech priming

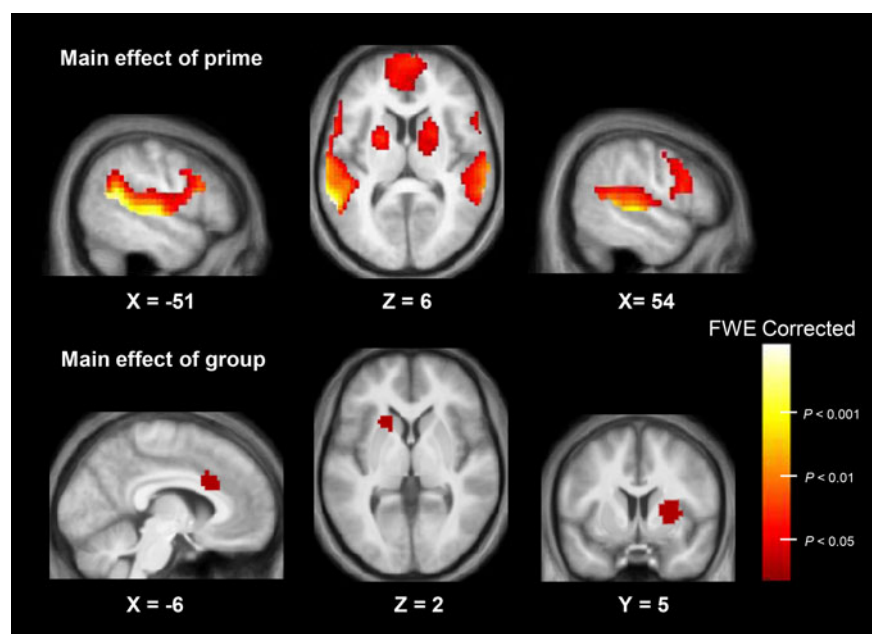
A 2 (group: control, patient) by 2 (prime type: ASP, ANSP) ANOVA of brain-image data showed that the main effect of priming type (ASP *v.* ANSP) revealed significant activation in the bilateral STG, bilateral supplementary motor area (SMA), bilateral medial superior frontal area, left posterior middle temporal gyrus (pMTG), right inferior frontal gyrus (IFG), right insular, right precentral, right putamen and right caudate (Fig. 3 top panel and online Supplementary Table S2; *F* contrasts were significant at  $p < 0.05$  with FWE correction). The main effect of group (control *v.* patient) elicited significant activation in the left anterior

cingulate cortex, left caudate and right putamen at  $p < 0.001$  uncorrected but not at  $p < 0.05$  correction with FDR (Fig. 3 bottom panel). The interaction was not significant at either  $p < 0.001$  uncorrected or  $p < 0.05$  correction with FDR. To further explore the brain substrates underlying the ASP effects, we conducted one-sample *t* tests and PPI analyses in the healthy control group and the patient group separately.

In healthy listeners, compared with the ANSP condition, introducing the ASP condition significantly enhanced BOLD signals in the bilateral STG, bilateral MTG, left pMTG and left putamen ( $p < 0.05$ , FWE corrected) (Fig. 4a, online Supplementary Table S3).

In patients, introducing the ASP condition (against the ANSP condition) significantly enhanced BOLD signals only in the left STG and left pMTG ( $p < 0.05$ , FWE corrected) (Fig. 4a, online Supplementary Table S3). Thus, both the left STG and left pMTG were activated by the ASP > ANSP listening-condition contrast in both healthy listeners and patients.

We also analysed the data in which the incorrect button-press trials were removed in each participant.



**Fig. 3.** Top panels: Voxels that exhibited a main effect of priming type. The activation map is thresholded at  $p < 0.001$  uncorrected and overlaid on the group-average structural image. The color scale indicates the  $p$  value corrected family-wise for type I error (FWE). Bottom panels: Voxels that exhibited a main effect of group type. The activation map is thresholded at  $p < 0.001$  uncorrected and overlaid on the structural image averaged across control group and patient group; no voxels survive the  $p$  value false discovery rate corrected for type I error.

For healthy controls, due to the high correct button-press performance (online Supplementary Fig. S1), the brain-image results without incorrect button press trials were similar to those with both correct trials and incorrect trials (online Supplementary Fig. S2 bottom panels and online Supplementary Table S4). For patient participants, however, activation in the bilateral STG, left MTG and left pMTG became enhanced by removal of incorrect trials (online Supplementary Fig. S2 bottom panels and online Supplementary Table S4).

#### *Functional connectivity of the left STG/pMTG in speech priming*

Since both the left STG and left MTG were activated by the ASP > ANSP condition contrast in both healthy listeners and patients, PPI analyses were then conducted to identify the brain regions showing enhanced co-variation specifically with the left STG/MTG induced by the ASP > ANSP contrast. In healthy listeners, enhanced functional connectivity was observed (1) between the left STG and the following brain regions: the left middle frontal gyrus, left pars triangularis of the IFG (TriIFG), left insular, left caudate, left MTG and bilateral putamen, and (2) between the left pMTG and the following brain regions: the left TriIFG, left insular, left caudate and left supra marginal area (Fig. 4b, online Supplementary Table S5).

In patients, enhanced functional connectivity included the ones (1) between the left STG and the following two brain regions: the left Rolandic operculum and the left caudate, and (2) between the left pMTG and the following two brain regions: the left Rolandic operculum and the left caudate (Fig. 4b, online Supplementary Table S5).

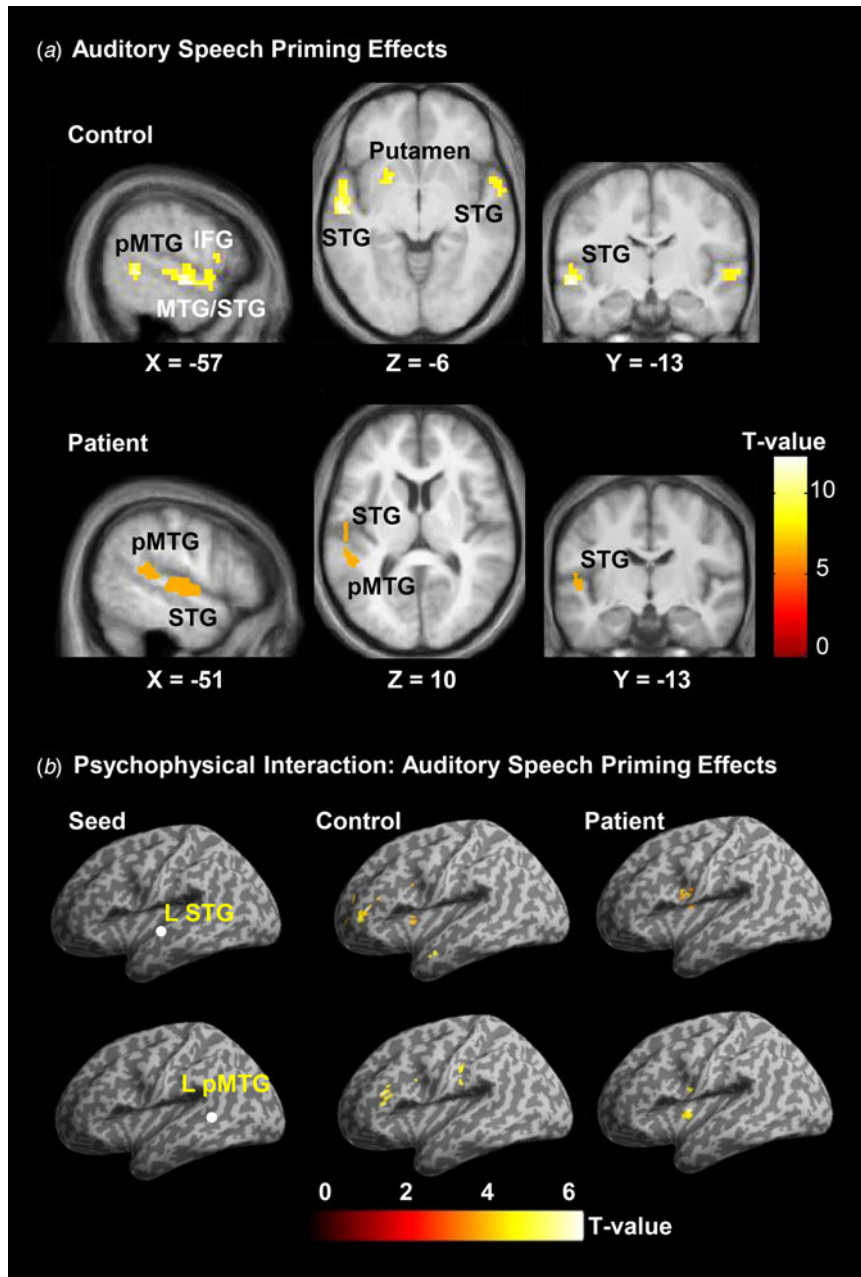
Thus, the left caudate was the only brain structure showing significantly enhanced functional connectivity with both the left STG and left pMTG in both healthy listeners and patients. Moreover, particularly in healthy listeners, the left STG/pMTG had functional connectivity with the left TriIFG and left insular; particularly in patients, the left STG/pMTG had functional connectivity with the left Rolandic operculum. No significant differences in PPI were found between healthy controls and patients in two-sample  $t$  tests.

#### *Functional connectivity related to speech recognition under speech-masking conditions*

*Brain regions activated by either the ASP listening condition or the ANSP listening condition*

As mentioned in the Introduction, listeners with schizophrenia perform worse than healthy listeners in speech recognition against informational masking (Wu et al. 2012a, 2013b). In this study, the (non-speech) noise baseline-stimulation condition was used as the reference for examining the brain regions and





**Fig. 4.** (a) Activated brain regions associated with the contrast of the auditory speech priming (ASP) listening condition against the auditory non-speech priming (ANSP) listening condition in healthy listeners and patients with schizophrenia. The activation maps were thresholded at  $p < 0.05$  (family-wise error corrected) and overlaid on the group-average structural image. (b) Psychophysiological interaction analyses in healthy listeners (middle column) and listeners with schizophrenia (right column) for revealing the ASP effect (ASP > ANSP)-related functional connectivity of the left superior temporal gyrus/middle temporal gyrus (STG/MTG). Locations of seed regions (left column) are overlaid on the template of SPM8, and the activation maps are overlaid on a template brain with inflated cortex of SPM8. All peaks are significant at  $p < 0.05$  (false discovery rate corrected). pMTG, Posterior middle temporal gyrus; IFG, inferior frontal gyrus; L, left.

functional connectivity induced by the speech-listening conditions.

In both healthy listeners and patient listeners, compared with the (non-speech) noise baseline-stimulation condition, the ASP listening condition significantly

enhanced BOLD signals in the bilateral STG, left putamen, left TriIFG, right precentral and right SMA ( $p < 0.05$ , FWE corrected) (online Supplementary Fig. S3 and online Supplementary Table S6). In addition, specifically in healthy controls, the activated brain regions

also included the right putamen and bilateral thalamus. Specifically in patients, the activated brain regions also included the left precentral, left inferior parietal lobule (IPL), right insular, right lingual and vermis of cerebellum ( $p < 0.05$ , FWE corrected).

Also as shown by online Supplementary Fig. S3 and online Supplementary Table S6, compared with the non-speech baseline-stimulation condition, in healthy controls the ANSP condition significantly enhanced BOLD signals in the bilateral STG, right putamen, and right SMA, left precentral and left TriIFG ( $p < 0.05$ , FWE corrected). In patients, the ANSP condition significantly enhanced BOLD signals in the bilateral STG, bilateral insular, left pars opercularis of the IFG (OperIFG), left IPL, left precentral, left cerebellum, right mid-cingulate cortex (MCC), right precentral and right lingual cortex ( $p < 0.05$ , FWE corrected).

#### *Schizophrenia-related changes in functional connectivity for speech listening*

Based on the results of brain regions activated by either the ASP listening condition (ASP > baseline) or the ANSP listening conditions (ANSP > baseline), functional connectivity of these activated brain areas were examined using partial correlation analyses. The brain region clusters activated by either the ASP- or ANSP-listening condition (against the non-speech baseline listening condition) were used as the seed regions (online Supplementary Table S6) ( $p < 0.05$ , FWE corrected) for conducting the comparison in functional connectivity between healthy listeners and patients. More specifically, in a brain region, clusters activated by the ASP-listening condition and those by the ANSP-listening condition were combined across healthy listeners and patient listeners to form a seed region for functional connectivity analyses (for example, all the four activated STG clusters shown in online Supplementary Table S6 across the two priming conditions and the two participant groups were combined to form the STG seed region for the functional connectivity analyses). In total, 20 clusters of seed regions were obtained and used for examining speech recognition-related functional connectivity in both healthy controls and patients: the left STG, right STG, left putamen, right putamen, left TriIFG, right TriIFG, left OperIFG, left precentral, right precentral, right SMA, left insular, right insular, left IPL, left postcentral, right MCC, left thalamus, right thalamus, right lingual, left cerebellum and vermis.

Then Z-scores (which represented the strength of functional connectivity) were obtained based on partial correlation analyses between time series of the total of 190 seed-region pairs for each participant, and compared between healthy controls and patients to reveal

the significantly changed functional connectivity (which was related to the schizophrenia-related poor performance in target-speech recognition). Fig. 5a shows Z-scores of all the seed-region pairs whose strengths of functional connectivity were significantly lower in patients than in healthy listeners.

#### *Correlation between strength of functional connectivity and speech-recognition performance*

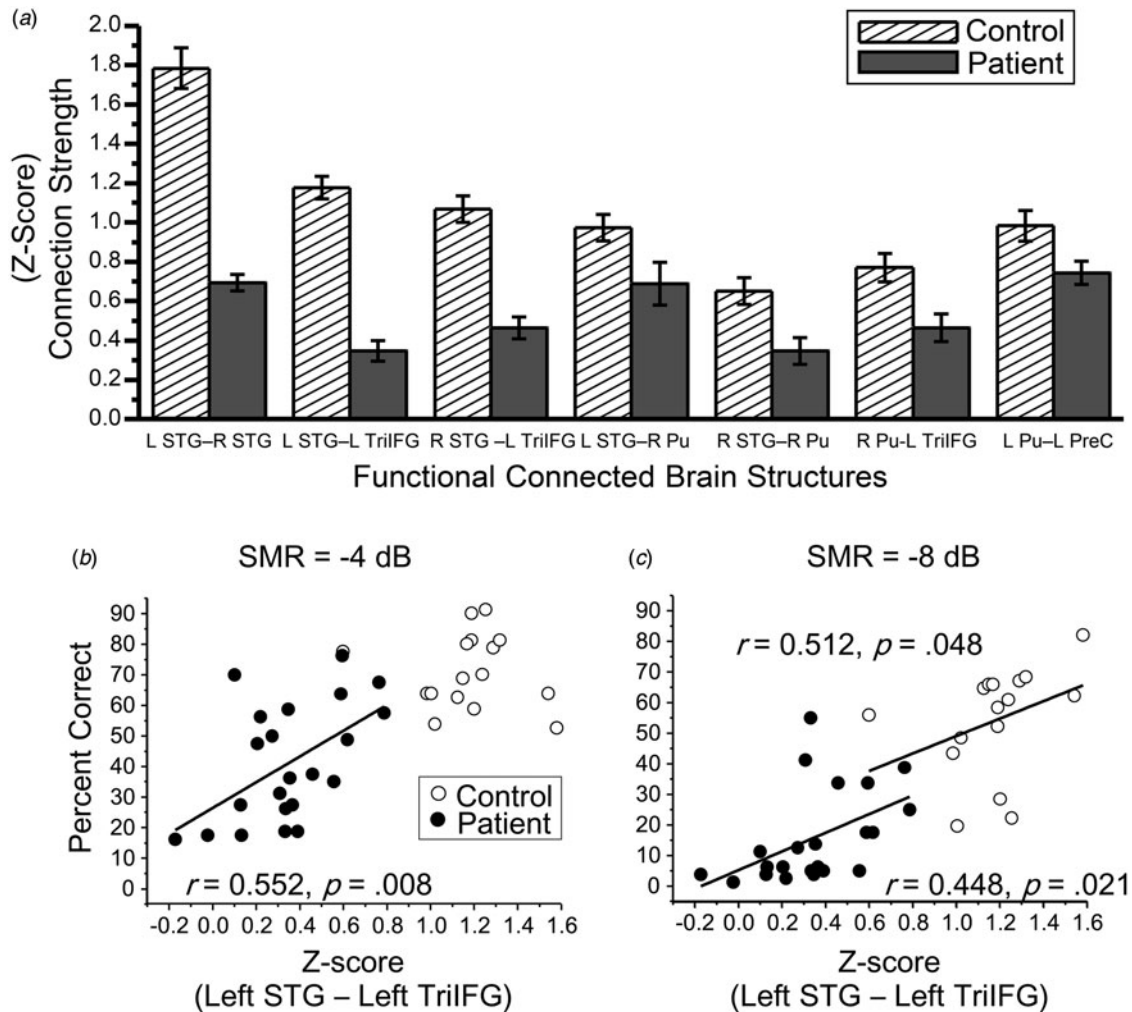
Spearman correlation analyses were conducted between the percentage correct of target-speech recognition and each of the Z-scores for the seed-region pairs shown in Fig. 5a in both healthy listeners and patients. The results showed that the percentage correct of target-speech recognition was significantly correlated only with the Z-score for functional connectivity between the left STG and the left TriIFG in healthy controls when the SMR was  $-8$  dB ( $r = 0.512$ ,  $p = 0.048$ ) and in patients when the SMR was either  $-8$  dB ( $r = 0.488$ ,  $p = 0.021$ ) or  $-4$  dB ( $r = 0.552$ ,  $p = 0.008$ ) (Fig. 5b and c). Also, not only was speech-recognition performance in individual patients generally poorer than that in individual healthy listeners, but also the Z-scores (representing the strength of functional connectivity) in individual patients were lower than those in individual healthy listeners.

## Discussion

### *Brain substrates activated by ASP*

This study confirms previous reports that in both healthy listeners (Freyman et al. 2004; Yang et al. 2007; Helfer & Freyman, 2009; Huang et al. 2010; Wu et al. 2012b, 2013a) and listeners with schizophrenia (Wu et al. 2012b), target-speech recognition against informational masking can be improved by temporally pre-presenting an early part of the target speech (i.e. the speech prime). The behavioral results of this study did not show a significant difference in the priming (unmasking) effect on speech recognition between the two participant groups. More importantly, this study for the first time reveals the brain substrates specifically underlying the speech prime-induced facilitating effect on speech recognition against speech masking in both healthy listeners and listeners with schizophrenia.

In healthy listeners, the BOLD-signal contrast between the ASP condition and the ANSP condition reveals significant enhancement of activation in the following brain regions: the bilateral STG, bilateral pMTG, left MTG, left superior temporal pole and left putamen. In listeners with schizophrenia, the significantly activated brain regions include the left STG and left pMTG. Since both the left STG and left



**Fig. 5.** (a) Strengths of functional connectivity (Z-score) of brain seed-region pairs, which were significantly different between healthy listeners and patients. (b) and (c) Individual participants' Z-scores for functional connectivity (the abscissa) between the left superior temporal gyrus (L STG) and the left pars triangularis of the inferior frontal gyrus (L TriIFG) were significantly correlated with individual participants' percentage correct recognition of target speech (the ordinate) in the patient group when the signal:masker ratio (SMR) was  $-4$  or  $-8$  dB, and in the healthy listener group when the SMR was  $-8$  dB. R, Right; Pu, putamen; PreC, precentral.

pMTG are activated by the ASP>ANSP contrast in both healthy listeners and listeners with schizophrenia, these two temporal lobe areas must be the most critical to the unmasking effect of ASP.

It has been suggested that the STG, particularly the anterior part of the STG, is an early stage in the cortical network for speech perception, mediating speech-sound identity information (Hickok & Poeppel, 2004; Ahveninen *et al.* 2006; Rauschecker & Scott, 2009, 2015). The involvement of the left STG in speech perception at the phonetic, lexical-semantic and syntactic levels has been confirmed by both brain imaging (Friederici *et al.* 2003; Scott & Wise, 2003) and functional lesion studies (Boatman, 2004). The left pMTG is involved in retrieval of lexical-syntactic information

from the mental lexicon in long-term memory (Snijders *et al.* 2009) and has a functional relationship with the left IFG in both processing syntactic structures and constructing grammatical representations of spoken language (Tyler *et al.* 2010, 2013; Walz *et al.* 2010; Papoutsis *et al.* 2011). Also, the left pMTG is involved in sentence-level prosody processing in Chinese-speech perception (Tong *et al.* 2005). Moreover, previous studies have confirmed both the function of the left rostral STG in working memory of sound identities (Arnott *et al.* 2005) and the function of the left pMTG in the storage of lexical representation (Lau *et al.* 2008). These memory functions of the left STG/pMTG are particularly critical to the memory-based speech priming effect. Thus, the speech priming-induced activation

of the left STG/pMTG indicates a speech priming-related facilitation of cortical representation of target-speech signals.

It should be noted that in this study, only younger-adult females' voices were used. Since previous studies have shown that male and female voices are represented as distinct auditory objects in the human non-primary auditory cortex (Weston *et al.* 2015), it is necessary in the future to examine whether similar results can be obtained when male-voice stimuli are used.

#### *ASP-related functional connectivity of the left STG/pMTG*

Relative to the ANSP condition, the ASP condition enhances functional connectivity of the left STG/pMTG with the left caudate in both healthy listeners and listeners with schizophrenia. In addition, introducing the ASP condition (against the ANSP condition) enhances functional connectivity of the left STG/pMTG with the left TriIFG, MFG, supramarginal gyrus, insular cortex and putamen in healthy listeners, and with the left Rolandic operculum in listeners with schizophrenia.

The ASP-related functional connectivity of the left STG/pMTG in healthy listeners suggests some critical mechanisms normally underlying the unmasking effect of ASP: (1) since the caudate contributes to speech inhibition and even more general response inhibition (Menon *et al.* 2001; Ketteler *et al.* 2008; Li *et al.* 2008; Ali *et al.* 2010) and the left IFG plays a role in signal selections among competing sources (Thompson-Schill *et al.* 2005), the unmasking effect of ASP should be based on both selective attention to target speech and suppression of disruptive speech signals; (2) the left IFG also contributes to multiple speech processes, including syntactic unifications on the basis of interplay with the left pMTG, speech meaning selection, sentence re-interpretation, and speech production (Herholz *et al.* 1996; Paulesu *et al.* 1997; Papathanassiou *et al.* 2000; Schuhmann *et al.* 2009; Snijders *et al.* 2009; Rodd *et al.* 2012). Since the involvement of the speech-motor system in speech perception is important for facilitating speech perception particularly under 'cocktail party' listening conditions (Wu *et al.* 2014), functional connectivity of the left STG/pMTG with the left TriIFG normally contributes to the unmasking effect of ASP through not only facilitating masking-speech suppression and target-speech representation (e.g. syntactic unification), but also inducing speech production.

The results of this study reveal that introducing the ASP condition (against the ANSP condition) both activates the STG/pMTG and enhances functional

connectivity between the STG/pMTG and the frontal cortex. Thus, it is of importance to know whether the enhanced functional connectivity between the STG/pMTG and the frontal cortex indicates a top-down modulation mechanism underlying the ASP-induced unmasking of speech.

It is known that the left Rolandic operculum is involved in sentence-level speech prosody processing (Ischebeck *et al.* 2008), speech production (Tonkonogy & Goodglass, 1981), and syntactic encoding during speech production (Indefrey *et al.* 2001). It is not clear whether this schizophrenia-enhanced functional connectivity between the left STG/pMTG and the left Rolandic operculum reflects schizophrenia-related neural plasticity underlying the unmasking effect of the speech-priming stimulus. The compensatory mechanism specifically underlying schizophrenia is an important issue (Tan *et al.* 2007).

#### *Functional connectivity specifically related to schizophrenia-induced deficits in speech recognition against informational masking*

The behavioral-testing results of this study support previous reports that under 'cocktail party' listening conditions (Schneider *et al.* 2007), listeners with schizophrenia perform poorly in recognizing target speech (Wu *et al.* 2012a, 2013b; Zheng *et al.* 2016). This study also reveals the brain substrates underlying the schizophrenia-related deficits in speech recognition under the priming conditions: The strength of functional connectivity between a number of speech-processing-related brain regions (including the STG, MTG, putamen, TriIFG, IPL, precentral cortex and SMA) declines in listeners with schizophrenia. Particularly, in listeners with schizophrenia, the strength of functional connectivity between the left STG/pMTG and the left TriIFG is not only correlated with speech-recognition performance but is also lower than that in healthy listeners. Iwashiro *et al.* (2012) have shown that the gray matter volume reduces in the TriIFG in both individuals at clinical high-risk for psychosis and those with first episode of schizophrenia. The declined functional connectivity between the left STG/pMTG and left TriIFG as revealed by this study further confirms the schizophrenia-related functional deficits in the left TriIFG.

In addition to the involvement of the STG, MTG and TriIFG in speech processing as discussed above, the putamen engages in both syntactic processing (Friederici *et al.* 2003) and speech initiation through its functional connectivity with both the left IFG and the left lateral temporal cortex (Booth *et al.* 2007). Moreover, the IPL is involved in interpreting sensory information (Zhang & Li, 2014) and works as a sensorimotor interface between the representation of acoustic signals in the auditory



cortex and the prediction of sensory consequences of articulatory gestures in the premotor cortex (Olson & Berryhill, 2009; Price, 2012; Du *et al.* 2014). Also, the precentral gyrus is part of the premotor cortex involved in speech processing (Callan *et al.* 2010) including phoneme segmentation (Sato *et al.* 2009), categorical speech perception in noise (Du *et al.* 2014) and speech production (Price, 2012). Finally, the SMA plays a role in planning, preparing, controlling and executing complex movements (Nachev *et al.* 2008; Price, 2012; Laviolette *et al.* 2013; Gao *et al.* 2014). The reduced strength of functional connectivity between the STG, MTG, putamen, TrIFG, IPL, precentral cortex and SMA may also account for the declined speech recognition under speech-masking conditions in listeners with schizophrenia. Indeed, previous studies have shown that functional connectivity between these brain regions is not only involved in auditory and/or speech processing (Jeong *et al.* 2009; Mhuircheartaigh *et al.* 2010; Price, 2010; Turken & Dronkers, 2011; Sundermann & Pfeleiderer, 2012; Zhang & Li, 2014; Kireev *et al.* 2015; Munoz-Lopez *et al.* 2015) but also impaired in people with schizophrenia (Ford *et al.* 2002; Liu *et al.* 2012; Leroux *et al.* 2014; Pu *et al.* 2014; Zhang & Li, 2014). Moreover, this study reveals that the strength of functional connectivity between the left STG and left TrIFG is significantly correlated with the speech-recognition performance under the informational masking condition in both healthy listeners and listeners with schizophrenia, and reduces in listeners with schizophrenia. More studies are needed to further examine how schizophrenia-related deficits in functional connectivity between the left STG and left TrIFG affect recognition of speech under the informational masking condition.

## Conclusions

- (1) Under 'cocktail party' listening conditions, both healthy listeners and listeners with schizophrenia are able to use temporally pre-presented ASP stimuli to improve their target-speech recognition against informational speech masking. The most critical brain regions underlying this unmasking effect are the left STG and the left pMTG in both healthy listeners and listeners with schizophrenia.
- (2) In both healthy listeners and listeners with schizophrenia, the left STG/pMTG has ASP-related functional connectivity with the left caudate, suggesting that the inhibitory function of the left caudate is involved in the unmasking effect of ASP.
- (3) Normally, the left STG and left pMTG have ASP-related functional connectivity with the left TrIFG, suggesting an enhanced functional integration among the following three processes: speech-

signal selection, speech-signal representation and speech production.

- (4) In listeners with schizophrenia, enhanced ASP-related functional connectivity of the left STG/pMTG and the left Rolandic operculum may indicate an ASP-related facilitation in sentence prosody processing and speech production.
- (5) The poor performance of speech recognition against informational masking in people with schizophrenia is associated with reduced strength of functional connectivity between a number of speech-processing-related brain regions, particularly associated with the impaired functional connectivity between the left STG and the left TrIFG.

## Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291716002816>

## Acknowledgements

Huahui Li assisted in many aspects of this work. This work was supported by the '973' National Basic Research Program of China (2015|CB351800, 2011CB707805); the National High Technology Research and Development Program of China (863 Program: 2015AA016306); the National Natural Science Foundation of China (31170985); the Planned Science and Technology Projects of Guangzhou (2014Y2-00105); the Guangzhou Municipal Key Disciplines in Medicine for Guangzhou Brain Hospital (GBH2014-QN04, GBH2014-ZD06); the Chinese National Key Clinical Program in Psychiatry (201201004); and the China Postdoctoral Science Foundation General Program (2013M530453).

## Declaration of Interest

None.

## References

- Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006). Task-modulated 'what' and 'where' pathways in human auditory cortex. *Proceedings of the National Academy of Sciences USA* **103**, 14608–14613.
- Ali N, Green DW, Kherif F, Devlin JT, Price CJ (2010). The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience* **22**, 2369–2386.
- Arnott SR, Grady CL, Hevenor SJ, Graham S, Alain C (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience* **17**, 819–831.

- Boatman D (2004). Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* **92**, 47–65.
- Booth JR, Wood L, Lu D, Houk JC, Bitan T (2007). The role of the basal ganglia and cerebellum in language processing. *Brain Research* **1133**, 136–144.
- Braff DL, Light GA (2005). The use of neurophysiological endophenotypes to understand the genetic basis of schizophrenia. *Dialogues in Clinical Neuroscience* **7**, 125–135.
- Brunart DS, Simpson BD, Ericson MA, Scott KR (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America* **110**, 2527–2538.
- Callan D, Callan A, Gamez M, Sato MA, Kawato M (2010). Premotor cortex mediates perceptual performance. *NeuroImage* **51**, 844–858.
- Cherry EC (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America* **25**, 975–979.
- Davis MH, Johnsruide IS, Hervais-Adelman A, Taylor K, McGettigan C (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General* **134**, 222–241.
- Ding N, Simon JZ (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences USA* **109**, 11854–11859.
- Du Y, Buchsbaum BR, Grady CL, Alain C (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences 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. *Journal of Cognitive Neuroscience* **28**, 483–500.
- First MB, Gibbon M, Spitzer RL, Williams JBW (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID), Clinician Version*. American Psychiatric Press: Washington, DC.
- Ford JM, Mathalon DH, Whitfield S, Faustman WO, Roth WT (2002). Reduced communication between frontal and temporal lobes during talking in schizophrenia. *Biological Psychiatry* **51**, 485–492.
- Freyman RL, Balakrishnan U, Helfer KS (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America* **115**, 2246–2256.
- Freyman RL, Helfer KS, McCall DD, Clifton RK (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America* **106**, 3578–3588.
- Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex* **13**, 170–177.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* **6**, 218–229.
- Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine* **35**, 346–355.
- Gao Q, Tao Z, Zhang M, Chen H (2014). Differential contribution of bilateral supplementary motor area to the effective connectivity networks induced by task conditions using dynamic causal modeling. *Brain Connectivity* **4**, 256–264.
- Gottesman II, Gould TD (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of 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. *Human Brain Mapping* **7**, 213–223.
- Hampson M, Peterson BS, Skudlarski P, Gatenby JC, Gore JC (2002). Detection of functional connectivity using temporal correlations in MR images. *Human Brain Mapping* **15**, 247–262.
- Helfer KS, Freyman RL (2005). The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America* **117**, 842–849.
- Helfer KS, Freyman RL (2009). Lexical and indexical cues in masking by competing speech. *Journal of the Acoustical Society of America* **125**, 447–456.
- Herholz K, Thiel A, Wienhard K, Pietrzyk U, von Stockhausen HM, Karbe H, Kessler J, Bruckbauer T, Halber M, Heiss WD (1996). Individual functional anatomy of verb generation. *NeuroImage* **3**, 185–194.
- Hickok G, Poeppel D (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99.
- 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. *Hearing Research* **244**, 51–65.
- Huang Y, Huang Q, Chen X, Wu X, Li L (2009). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *Journal of Experimental Psychology: Human Perception and Performance* **35**, 1618–1628.
- Huang Y, Xu L, Wu X, Li L (2010). The effect of voice cuing on releasing speech from informational masking disappears in older adults. *Ear and Hearing* **31**, 579–583.
- Indefrey P, Brown CM, Hellwig F, Amunts K, Herzog H, Seitz RJ, Hagoort P (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Sciences USA* **98**, 5933–5936.
- Ischebeck AK, Friederici AD, Alter K (2008). Processing prosodic boundaries in natural and hummed speech: an fMRI study. *Cerebral Cortex* **18**, 541–552.
- Iwashiro N, Suga M, Takano Y, Inoue H, Natsubori T, Satomura Y, Koike S, Yahata N, Murakami M, Katsura M, Gonoji W, Sasaki H, Takao H, Abe O, Kasai K, Yamasue H (2012). Localized gray matter volume reductions in the pars triangularis of the inferior frontal gyrus in individuals at clinical high-risk for psychosis and first episode for schizophrenia. *Schizophrenia Research* **137**, 124–131.

- Jeong B, Wible CG, Hashimoto R, Kubicki M** (2009). Functional and anatomical connectivity abnormalities in left inferior frontal gyrus in schizophrenia. *Human Brain Mapping* **30**, 4138–4151.
- Ketteler D, Kastrau F, Vohn R, Huber W** (2008). The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. *NeuroImage* **39**, 2002–2009.
- Kireev M, Slioussar N, Korotkov AD, Chernigovskaya TV, Medvedev SV** (2015). Changes in functional connectivity within the fronto-temporal brain network induced by regular and irregular Russian verb production. *Frontiers in Human Neuroscience* **9**, 36.
- Lau EF, Phillips C, Poeppel D** (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience* **9**, 920–933.
- Laviolette L, Niérat M-C, Hudson AL, Raux M, Allard É, Similowski T** (2013). The supplementary motor area exerts a tonic excitatory influence on corticospinal projections to phrenic motoneurons in awake humans. *PLOS ONE* **8**, e62258.
- Leroux E, Delcroix N, Dollfus S** (2014). Left fronto-temporal dysconnectivity within the language network in schizophrenia: an fMRI and DTI study. *Psychiatry Research* **223**, 261–267.
- Li CS, Yan P, Sinha R, Lee TW** (2008). Subcortical processes of motor response inhibition during a stop signal task. *NeuroImage* **41**, 1352–1363.
- 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? *Journal of Experimental Psychology. Human Perception and Performance* **30**, 1077–1091.
- Liu H, Kaneko Y, Ouyang X, Li L, Hao Y, Chen EY, Jiang T, Zhou Y, Liu Z** (2012). Schizophrenic patients and their unaffected siblings share increased resting-state connectivity in the task-negative network but not its anticorrelated task-positive network. *Schizophrenia Bulletin* **38**, 285–294.
- Liu Y, Liang M, Zhou Y, He Y, Hao Y, Song M, Yu C, Liu H, Liu Z, Jiang T** (2008). Disrupted small-world networks in schizophrenia. *Brain* **131**, 945–961.
- Liu Y, Yu C, Zhang X, Liu J, Duan Y, Alexander-Bloch AF, Liu B, Jiang T, Bullmore E** (2014). Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. *Cerebral Cortex* **24**, 1422–1435.
- Menon V, Adleman NE, White CD, Glover GH, Reiss AL** (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping* **12**, 131–143.
- Mhuircheartaigh RN, Rosenorn-Lanng D, Wise R, Jbabdi S, Rogers R, Tracey I** (2010). Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: a functional magnetic resonance imaging study using propofol. *Journal of Neuroscience* **30**, 9095–9102.
- Munoz-Lopez M, Insausti R, Mohedano-Moriano A, Mishkin M, Saunders RC** (2015). Anatomical pathways for auditory memory II: information from rostral superior temporal gyrus to dorsolateral temporal pole and medial temporal cortex. *Frontiers in Neuroscience* **9**, 158.
- Nachev P, Kennard C, Husain M** (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews. Neuroscience* **9**, 856–869.
- Olson IR, Berryhill M** (2009). Some surprising findings on the involvement of the parietal lobe in human memory. *Neurobiology of Learning and Memory* **91**, 155–165.
- Papathanassiou D, Etard O, Mellet E, Zago L, Mazoyer B, Tzourio-Mazoyer N** (2000). A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. *NeuroImage* **11**, 347–357.
- Papoutsi M, Stamatakis EA, Griffiths J, Marslen-Wilson WD, Tyler LK** (2011). Is left fronto-temporal connectivity essential for syntax? Effective connectivity, tractography and performance in left-hemisphere damaged patients. *NeuroImage* **58**, 656–664.
- Paulesu E, Goldacre B, Scifo P, Cappa SF, Gilardi MC, Castiglioni I, Perani D, Fazio F** (1997). Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. *Neuroreport* **8**, 2011–2017.
- Price CJ** (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences* **1191**, 62–88.
- 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.
- Pu W, Rolls ET, Guo S, Liu H, Yu Y, Xue Z, Feng J, Liu Z** (2014). Altered functional connectivity links in neuroleptic-naïve and neuroleptic-treated patients with schizophrenia, and their relation to symptoms including volition. *NeuroImage Clinical* **6**, 463–474.
- Rakerd B, Aaronson NL, Hartmann WM** (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *Journal of the Acoustical Society of America* **119**, 1597–1605.
- Rauschecker JP, Scott SK** (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience* **12**, 718–724.
- Rauschecker JP, Scott SK** (2015). Pathways and streams in the auditory cortex: an update on how work in nonhuman primates has contributed to our understanding of human speech processing. In *Neurobiology of Language* (ed. G Hickok and SL Small), pp. 287–298. Elsevier Inc.: Waltham, MA.
- Rodd JM, Johnsrude IS, Davis MH** (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex* **22**, 1761–1773.
- Salvador R, Suckling J, Schwarzbauer C, Bullmore E** (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **360**, 937–946.
- Sato M, Tremblay P, Gracco VL** (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language* **111**, 1–7.
- Schneider BA, Li L, Daneman M** (2007). How competing speech interferes with speech comprehension in everyday

- listening situations. *Journal of the American Academy of Audiology* **18**, 559–572.
- Schuhmann T, Schiller NO, Goebel R, Sack AT** (2009). The temporal characteristics of functional activation in Broca's area during overt picture naming. *Cortex* **45**, 1111–1116.
- Scott SK, McGettigan C** (2013). The neural processing of masked speech. *Hearing Research* **303**, 58–66.
- 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. *Journal of the Acoustical Society of America* **125**, 1737–1743.
- 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. *Journal of the Acoustical Society of America* **115**, 813–821.
- Scott SK, Wise RJ** (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication* **41**, 23–34.
- 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. *Chinese Mental Health Journal* **18**, 45–47.
- Snijders TM, Vosse T, Kempen G, Van Berkum JJ, Petersson KM, Hagoort P** (2009). Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cerebral Cortex* **19**, 1493–1503.
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH** (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience* **32**, 8443–8453.
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH** (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology. Human Perception and Performance* **40**, 186–199.
- Sundermann B, Pfeleiderer B** (2012). Functional connectivity profile of the human inferior frontal junction: involvement in a cognitive control network. *BMC Neuroscience* **13**, 119.
- Tan HY, Callicott JH, Weinberger DR** (2007). Dysfunctional and compensatory prefrontal cortical systems, genes and the pathogenesis of schizophrenia. *Cerebral Cortex* **17**, i171–i181.
- Thompson-Schill SL, Bedny M, Goldberg RF** (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology* **15**, 219–224.
- Tong Y, Gandour J, Talavage T, Wong D, Dziedzic M, Xu Y, Li X, Lowe M** (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage* **28**, 417–428.
- Tonkonogy J, Goodglass H** (1981). Language function, foot of the third frontal gyrus, and rolandic operculum. *Archives of Neurology* **38**, 486–490.
- Turken AU, Dronkers NF** (2011). The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Frontiers in Systems Neuroscience* **5**, 1.
- Tyler LK, Cheung TP, Devereux BJ, Clarke A** (2013). Syntactic computations in the language network: characterizing dynamic network properties using representational similarity analysis. *Frontiers in Psychology* **4**, 271.
- Tyler LK, Wright P, Randall B, Marslen-Wilson WD, Stamatakis EA** (2010). Reorganization of syntactic processing following left-hemisphere brain damage: does right-hemisphere activity preserve function? *Brain* **133**, 3396–3408.
- Walz JM, Goldman RI, Carapezza M, Muraskin J, Brown TR, Sajda P** (2013). Simultaneous EEG-fMRI reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. *Journal of Neuroscience* **33**, 19212–19222.
- Weston PSJ, Hunter MD, Sokhi DS, Wilkinson ID, Woodruff PWR** (2015). Discrimination of voice gender in the human auditory cortex. *NeuroImage* **105**, 208–214.
- Wild CJ, Davis MH, Johnsrude IS** (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage* **60**, 1490–1502.
- Woods SW** (2003). Chlorpromazine equivalent doses for the newer atypical antipsychotics. *Journal of Clinical Psychiatry* **64**, 663–667.
- Wu C, Cao S, Wu X, Li L** (2013a). Temporally pre-presented lipreading cues release speech from informational masking. *Journal of the Acoustical Society of America* **133**, E1281–E1285.
- Wu C, Cao S, Zhou F, Wang C, Wu X, Li L** (2012a). Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia. *Schizophrenia Research* **134**, 33–41.
- Wu C, Li H, Tian Q, Wu X, Wang C, Li L** (2013b). Disappearance of the unmasking effect of temporally pre-presented lipreading cues on speech recognition in people with chronic schizophrenia. *Schizophrenia Research* **150**, 594–595.
- Wu M, Li H, Gao Y, Lei M, Teng X, Wu X, Li L** (2012b). Adding irrelevant information to the content prime reduces the prime-induced unmasking effect on speech recognition. *Hearing Research* **283**, 136–143.
- Wu X, Wang C, Chen J, Qu H, Li W, Wu Y, Schneider BA, Li L** (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hearing Research* **199**, 1–10.
- Wu Z, Chen M, Wu X, Li L** (2014). Interaction between auditory and motor systems in speech perception. *Neuroscience Bulletin* **30**, 490–496.
- Yang Z, Chen J, Huang Q, Wu X, Wu Y, Schneider BA, Liang L** (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication* **49**, 892–904.
- Zhang S, Li C-SR** (2014). Functional clustering of the human inferior parietal lobule by whole-brain connectivity mapping of resting-state functional magnetic resonance imaging signals. *Brain Connectivity* **4**, 53–69.
- Zheng Y, Wu C, Li J, Wu H, She S, Liu S, Wu H, Mao L, Ning Y, Li L** (2016). Brain substrates of perceived spatial separation between speech sources under simulated reverberant listening conditions in schizophrenia. *Psychological Medicine* **46**, 477–491.