Psychological Medicine (2016), **46**, 477–491. © Cambridge University Press 2015
doi:10.1017/S0033291715001828

ORIGINAL ARTICLE

Brain substrates of perceived spatial separation between speech sources under simulated reverberant listening conditions in schizophrenia

Y. Zheng[1]†, C. Wu[2]†, J. Li[1]†, H. Wu[1], S. She[1], S. Liu[1], H. Wu[1], L. Mao[2], Y. Ning[1] and L. Li[2,3]*

[1] *Guangzhou Brain Hospital, the Affiliated Hospital of Guangzhou Medical University, Guangzhou 510370, People's Republic of China*
[2] *Department of Psychology, School of Life Sciences, McGovern Institute for Brain Research at PKU, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871, People's Republic of China*
[3] *Beijing Institute for Brain Disorders, Capital Medical University, Beijing, People's Republic of China*

Background. People with schizophrenia recognize speech poorly under multiple-people-talking (informational masking) conditions. In reverberant environments, direct-wave signals from a speech source are perceptually integrated with the source reflections (the precedence effect), forming perceived spatial separation (PSS) between different sources and consequently improving target-speech recognition against informational masking. However, the brain substrates underlying the schizophrenia-related vulnerability to informational masking and whether schizophrenia affects the unmasking effect of PSS are largely unknown.

Method. Using psychoacoustic testing and functional magnetic resonance imaging, respectively, the speech recognition under either the PSS or perceived spatial co-location (PSC) condition and the underlying brain substrates were examined in 20 patients with schizophrenia and 16 healthy controls.

Results. Speech recognition was worse in patients than controls. Under the PSS (but not PSC) condition, speech recognition was correlated with activation of the superior parietal lobule (SPL), and target speech-induced activation of the SPL, precuneus, middle cingulate cortex and caudate significantly declined in patients. Moreover, the separation (PSS)-against-co-location (PSC) contrast revealed (1) activation of the SPL, precuneus and anterior cingulate cortex in controls, (2) suppression of the SPL and precuneus in patients, (3) activation of the pars triangularis of the inferior frontal gyrus and middle frontal gyrus in both controls and patients, (4) activation of the medial superior frontal gyrus in patients, and (5) impaired functional connectivity of the SPL in patients.

Conclusions. Introducing the PSS listening condition efficiently reveals both the brain substrates underlying schizophrenia-related speech-recognition deficits against informational masking and the schizophrenia-related neural compensatory strategy for impaired SPL functions.

Received 9 November 2014; Revised 9 August 2015; Accepted 24 August 2015; First published online 1 October 2015

Key words: Neural compensation, precedence effect, precuneus, schizophrenia, spatial attention, speech recognition, superior parietal lobule.

Introduction

People with schizophrenia perform poorly in various perceptual/cognitive tasks (Gjerde, 1983; Nuechterlein & Dawson, 1984; Fish & Granholm, 2008), particularly when the perceptual/cognitive workload is high (Nuechterlein *et al.* 1994; Seidman *et al.* 1998; Verleger *et al.* 2013). Consistently, in a 'cocktail party' environment with multiple people talking, people with schizophrenia perform worse in recognizing speech than healthy people because they are more vulnerable to informational speech masking (Wu *et al.* 2012, 2013). To date, the brain substrates underlying this schizophrenia-related vulnerability have not been reported in the literature.

When the 'cocktail party' environment is reverberant, it become even more difficult for listeners to attend to and recognize speech. To improve speech recognition under this adverse listening condition, listeners usually take advantage of various perceptual and/or cognitive cues available to facilitate perceptual segregation between speech sources. These cues include the precedence effect-induced perceived spatial separation (PSS) between target speech and masking speech (Freyman *et al.* 1999; Li *et al.* 2004,

---

* Address for correspondence: L. Li, Ph.D., Department of Psychology, Peking University, 5 Yiheyuan Road, Beijing 100871, People's Republic of China.
(Email: liangli@pku.edu.cn)

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

2013; Wu *et al.* 2005; Rakerd *et al.* 2006; Huang *et al.* 2008).

Spatial separation between a target sound and its masking sounds improves recognition of the target mainly by: (1) the acoustic effect of head shadowing that increases the signal-to-masker ratio (SMR) at the ear near the target; (2) the neurophysiological effect of the disparity in arriving-time difference between inputs to the two ears (i.e. the effect of the disparity in interaural time difference); and (3) the effect of selective attention that facilitates the target salience and reduces the masker salience (Li *et al.* 2013). In a reverberant environment, although numerous sound reflections bouncing from surfaces limit or even abolish both the head-shadowing and interaural-interaction effects, the perceptually unmasking effect still remains (Koehnke & Besing, 1996; Freyman *et al.* 1999; Darwin & Hukin, 2000; Culling *et al.* 2003; Zurek *et al.* 2004; Kidd *et al.* 2005). The persistence of the perceptually unmasking effect under reverberant conditions is based on the auditory precedence effect (see below).

Under a reverberant condition, the signal of the direct sound wave from a source can be perceptually integrated with signals of its time-delayed and linearly filtered reflections: attributes of the reflections are perceptually captured by the direct wave (Li *et al.* 2005), resulting in a single fused image of the source whose perceived location is around the location of the source (the precedence effect, Wallach *et al.* 1949; Zurek, 1980; Freyman *et al.* 1991; Huang *et al.* 2011). In a simulated reverberant environment, the precedence effect can cause PSS of a target signal from other disruptive stimuli (which will not be as highly correlated with the target signal). For example, when both a target speech and a masker are presented by each of the two spatially separated (e.g. left and right) loudspeakers and the right loudspeaker is the leading loudspeaker for both target and masker (simulating the two sources) with a leading time of 3 ms (meanwhile the left loudspeaker simulates the reflections), the listener will perceive both the target 'image' and the masker 'image' as coming from the right loudspeaker [i.e. the perceived spatial co-location (PSC) condition]. Under this PSC condition, if a listener attends to the target, it is difficult for the listener to ignore the spatially co-located masker. However, when the left loudspeaker becomes the leading one only for the masker, the listener will perceive the masker 'image' as coming from the left loudspeaker, with the target 'image' still as coming from the right loudspeaker (i.e. the PSS condition) (Li *et al.* 2004). Under the PSS condition, because the masker 'image' is outside of the attention focus to the target, the masker signal can be ignored or even suppressed. It has been confirmed that the precedence effect-induced PSS between target speech and masking speech markedly improves target-speech recognition by both enhancing spatial attention to the target and suppressing spatial attention to the masker (Freyman *et al.* 1999; Li *et al.* 2004, 2013; Wu *et al.* 2005; Rakerd *et al.* 2006; Huang *et al.* 2008). It should be noted that when the locations of the two spatially separated loudspeakers are symmetric to the listener, shifts between the two listening conditions do not change the SMR (in sound pressure level), compactness/diffusiveness, timbre, or loudness of the sounds. In other words, peripheral processes are not affected. To date, although progress has been made in understanding the brain regions involved in either speech perception or masking of speech (e.g. Scott & Wise, 2003; Scott *et al.* 2004; Ding & Simon, 2012; McGettigan *et al.* 2012; Scott & McGettigan, 2013), the brain substrates underlying the unmasking effect of PSS have not been reported in the literature.

People with schizophrenia both exhibit the intactness of the precedence effect (Mickey & Dalack, 2005) and can use temporally pre-presented auditory content/voice prime to improve their target-speech recognition (Wu *et al.* 2012). So far no studies have examined whether people with schizophrenia can also use the precedence-effect-induced PSS to improve target-speech recognition against informational speech masking. Moreover, it is not clear whether the neural mechanism underlying PSS in people with schizophrenia is the same as or different from that in healthy people.

Thus, the first goal of this study was to investigate the brain substrates underlying the schizophrenia-related deficits of speech recognition under the listening condition with informational speech masking under a simulated reverberant environment. The second goal of this study was to investigate whether the PSS between target speech and masking speech improves target-speech recognition in people with schizophrenia and whether the brain activation induced by the PSS in people with schizophrenia differs from that in healthy people.

Method

Participants

Patients with schizophrenia were recruited in the Guangzhou Brain Hospital, with age ranging from 18 to 59 years. Their diagnoses were based on the Structured Clinical Interview for DSM-IV (SCID) (First *et al.* 2012). Some potential patient participants were excluded due to hearing loss, alcohol/drug abuse, a treatment with electroconvulsive therapy within the past 6 months, and/or a treatment of trihexyphenidyl hydrochloride with a dose of more than 6 mg/day. Demographics-matched healthy participants (controls) were recruited from communities around the hospital. They showed no significant differences in age, gender and education level compared with participants with schizophrenia. They were telephone interviewed first and then examined with the SCID.

All participants were right-handed and showed normal pure-tone hearing thresholds (<30 dB hearing level) between 125 and 8000 Hz. Their first language was Mandarin Chinese. Both participants and patient participants' guarantees gave written informed consent for the participation in this study. The locally validated version of the Positive and Negative Syndrome Scale (PANSS) tests (Si *et al.* 2004) was conducted on the day of behavioural testing or functional magnetic resonance imaging (fMRI) scanning.

A total of 18 healthy participants and 24 patients with schizophrenia participated in both the behavioural testing and fMRI scanning. However, due to excessive head movement and/or failure in following the instructions to button press, two healthy participants and four patient participants were excluded. Hence, a total of 16 healthy participants and 20 patients (see online Supplementary Table S1) were included in both behavioural data and fMRI data analyses.

The procedures of this study were approved by the Independent Ethics Committee of the Guangzhou Brain Hospital.

Stimuli and equipment

Speech stimuli were Chinese nonsense sentences, which are syntactically correct but not semantically meaningful (see Helfer, 1997; Yang *et al.* 2007). For example, the English translation of a Chinese nonsense sentence is 'One appreciation could retire his ocean' (the three keywords are underlined). Obviously, the sentence frame provides no contextual support for recognizing keywords. Target sentences were spoken by a young female talker (talker A). 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) spoken by two different young female talkers (talkers B and C) (Yang *et al.* 2007).

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 (Brungart *et al.* 2005; Qu *et al.* 2008, 2009). More specifically, HRTF data were first derived from the PKU-IOA HRTF database (Qu *et al.* 2008), which was based on the acoustic filtering by a Knowles Electronic Manikin for Acoustic Research (Knowles Electronics, Inc., USA). The speech signals were filtered with the HRTFs to simulate source locations at 90° left and 90° right to the listener in the azimuth, respectively.

Moreover, based on the precedence effect paradigm, both the target and masker were simulated as being presented by each of the two spatially separated 'loudspeakers' with the inter-source interval of 3 ms. For example, under the PSS condition, 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 fused masker image was perceived as coming from the right location. Also, under the PSC condition, 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.

In the behavioural test, the acoustic signals were calibrated by a sound-level meter (AUDit and System 824; Larson Davis, USA), transformed from a notebook-computer sound card (ATI SB450 AC97, Beijing, China) to headphones (model HDA 600; Wedemark, Germany) and presented to the participant at the sound pressure level of 60 dB. The sound pressure level of the speech masker was adjusted to produce four SMRs: −8, −4, 0 and 4 dB.

In the fMRI experiment, the acoustic stimuli were presented through a magnetic resonance-compatible pneumatic headphone system (SAMRTEC; China) driven by Presentation software (version 0.70). The target level was about 60 dB sound pressure level (after attenuation by earplugs) and the SMR was −4 dB. A 3.0-Tesla Philips Achieva MRI scanner (Veenpluis 4-6, the Netherlands) was used to acquire blood oxygenation level-dependent (BOLD) gradient echo-planar images (spatial resolution: 64 × 64 × 33 matrix with 3.44 × 3.44 × 4.6 mm^3; acquisition time: 2000 ms; time to repeat: 9000 ms; echo time: 30 ms; flip angle: 90°; field of view: 211 × 211 mm^2). It provided high-resolution T1-weighted structural images [256 × 256 × 188 matrix with a spatial resolution of 1 × 1 × 1 mm^3, repetition time (TR): 8.2 ms; echo time: 3.8 ms; flip angle: 7°].

Design and procedures

Behavioural testing

There were two within-subject variables: (1) listening condition (PSS, PSC); and (2) SMR (−8, −4, 0 and 4 dB). For each participant, there were eight testing conditions and 12 trials (also 12 target-sentence presentations) for each condition. The presentation order for the eight combinations of spatial condition and SMR were partially counterbalanced across participants in each participant group using a Latin square order.

In a trial, the participant, who was seated at the centre of a quiet room in the hospital, pressed the 'Enter' key on a computer keyboard to start the masker presentation. About 1 s later, the target sentence was presented. Then the target sentence terminated with the masker. After the masker/target co-presentation was finished, the participant was instructed to loudly repeat the whole target sentence 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 three keywords had been identified correctly.

A logistic psychometric function,

$$y = 1/[1 + e^{-\sigma(x-\mu)}],$$

was fit to individual participants' speech recognition performances, using the Levenberg–Marquardt method, where y is the probability of correct recognition of the keywords, x is the SMR corresponding to y, μ (threshold in dB) is the SMR corresponding to 50% correct on the psychometric function, and σ determines the slope of the psychometric function (Li *et al.* 2004; Yang *et al.* 2007). Thus, μ could be used to evaluate the behavioural performance (a lower μ represented a better performance). The $\Delta\mu$ (difference in μ between the PSS and PSC conditions) presented the unmasking effect induced by PSS. Analysis of variance (ANOVA) followed by *post-hoc* tests (when necessary) was performed using SPSS 16.0 software (USA). The null hypothesis was rejected at the level of 0.05.

fMRI experiments

The whole-course scanning consisted of an 8-min run for localizing the auditory cortex, two 10-min functional runs for associating with auditory speech perception, and an 8-min run for scanning brain structures. There were 60 trials for each of the two functional runs, 20 for each of the three listening conditions (PSS, PSC and baseline stimulation) with a random presenting order for individual participants. Thus, 40 images were collected for each condition and 120 images were obtained in total. The masker-only condition was used as the baseline to highlight the effect of target-speech presentation.

The sparse-imaging strategy was used to avoid the effect of machine noise on the auditory task: stimuli were presented only during the resting period between successive scanning periods (Hall *et al.* 1999). In each trial, to ensure that the stimulus-evoked haemodynamic responses peaked within the scanning period (Wild *et al.* 2012), the stimulus presentation was so temporally positioned that the stimulus-presentation midpoint occurred 4100 ms before the following scanning onset.

In a scanning trial (online Supplementary Fig. S1), the two-talker masker was presented in quiet 800 ms after the last scanning trial. About 1 s later, the target sentence was presented. Then the target sentence terminated with the masker. To maintain participants' attention to target speech, participants were instructed to press the left button on a response box using their right index finger if they heard a target sentence, and press the right button if they did not.

All participants who participated in fMRI scanning were screened for MR safety prior to scanning. Participants were also provided a brief training to ensure that they understood both the instructions and their press-button responses under each of the stimulus conditions.

fMRI data processing and analyses

Preprocessing

All fMRI data were processed and analysed using Statistical Parametric Mapping (SPM8; the Wellcome Trust Centre for Neuroimaging, UK). The procedures included: (1) correcting functional images for head movements; (2) co-registering anatomical images with the mean realigned images and then normalizing to a standard template (ICBM space) using the SPM8 unified segmentation routine; (3) warping all functional images using deformation parameters generated from the normalization process, including re-sampling to a voxel size of $3.0 \times 3.0 \times 4.0\,\text{mm}^3$; (4) spatial smoothing with a Gaussian kernel with 8-mm full-width at half maximum (FWHM). Due to the long TR of this sparse-imaging paradigm, no slice timing was used.

Random-effect analyses

A hierarchical random-effect model with two levels was used in statistical analyses in SPM8. At the first level, the onset and duration of each run were modelled using a general linear model according to the condition types. The second and third functional runs were modelled as one session within the design matrix, and three conditions (separation, co-location and baseline) were included in the model. Six realignment parameters were included to account for residual movement-related effects (Friston *et al.* 1996).

Random-effect analyses were conducted based on the statistical parameter maps from each participant to allow population inference. Contrast images of the 'separation > baseline', 'co-location > baseline' and 'separation > co-location' 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. Contrasts between the two participant groups were performed by second-level two-sample t tests (in SPM8). For the whole-brain analysis, only the peak signals that were statistically significant at the p value less than 0.05 [false discovery rate (FDR) corrected, with the activation size larger than 10 contiguous voxels] were reported.

Three main contrasts between conditions were computed: (1) 'separation > baseline'; (2) 'co-location > baseline'; and (3) 'separation > co-location'. The difference in the contrast of 'separation > baseline' and that of 'co-location > baseline' between controls and patients were used to reveal the schizophrenia-related abnormal brain regions involved in target speech recognition when target sound and masking sound come from different or the same perceptual direction. The contrast of 'separation > co-location' was computed to reveal the brain areas involved in the unmasking effect induced by PSS.

Region-of-interest (ROI) analyses

ROI analyses were conducted to examine the role that the brain regions played in either the target speech-recognition performance against informational masking or the speech-unmasking effect induced by PSS. The ROIs were defined by the activation clusters extracted from the SPM files of the 'separation > baseline', 'co-location > baseline' or 'separation > co-location' contrasts (MarsBaR: ROI toolbox for SPM; http://marsbar.sourceforge.net/) in healthy control and patient participants (FDR corrected). The parameter estimates of signal intensity for each ROI were extracted from each individual participant. Then the mean contrast value (CV) for each condition (i.e. the parameter estimate for PSS condition minus the parameter estimate for baseline, averaged across participants) for each ROI were calculated (Wild *et al.* 2012). Spearman correlation analyses were performed using SPSS 16.0 software to investigate the correlation between: (1) the intensity of brain activities (CV); (2) behavioural performance (μ or $\Delta\mu$); and (3) psychiatric symptoms scores (i.e. the positive symptom scores of the PANSS, negative symptom scores of PANSS, and total scores of PANSS). The null hypothesis was rejected at the level of 0.05.

Psychophysiological interaction (PPI) analyses

PPI analyses (Friston *et al.* 1997) were performed to identify which brain regions showed significantly increased functional connectivity with the activity of the most critical brain structures (seeds) related to PSS compared with PSC. The coordinates of the peak voxel from the contrast of 'separation > co-location' in random-effects analyses were used as the landmarks for the individual seed voxels. A seed region in each participant was defined as a sphere with 5-mm radius centred at the peak voxel. The time series of seed regions were then extracted, and the PPI regressors which reflected the interaction between psychological variable (PSS *v*. PSC) 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 PSS condition against the PSC 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 p values less than 0.05 (FDR corrected) were reported.

Results

Behavioural testing

Fig. 1*a* shows comparisons in group mean percentage correct recognition of keywords in target speech between the healthy control group (left panel) and the patient group (right panel) as a function of SMR (along with the group mean best-fitting psychometric functions) under either the PSS condition or the PSC condition. Fig. 1*b* shows the threshold μ of the speech-recognition performance for each of the two participant groups under either the PSS or PSC condition. Obviously, healthy controls had better speech-recognition performance than patients. Interestingly, similar to healthy controls, patient participants were able to use the PSS cue to improve their target-speech recognition (Fig. 1*c*).

To statistically examine the differences in percentage correct performance between the participant groups, a 2 (group: control, patient) by 2 (spatial condition: separation, co-location) by 4 (SMR: −8, −4, 0, and 4 dB) three-way ANOVA showed that the main effects of group ($F_{1,272} = 216.179$, $p < 0.001$), spatial condition ($F_{1,272} = 16.266$, $p < 0.001$) and SMR ($F_{3,272} = 74.318$, $p < 0.001$) were all significant. However, all the two-way interactions and the three-way interaction were not significant.

For the threshold μ, a 2 (group: control, patient) by 2 (spatial type: separation, co-location) ANOVA showed that the main effects of group ($F_{1,68} = 47.397$, $p < 0.001$) and spatial type ($F_{1,68} = 15.080$, $p = 0.004$) were significant, but the two-way interaction was not significant ($F_{1,68} = 0.670$, $p = 0.416$).

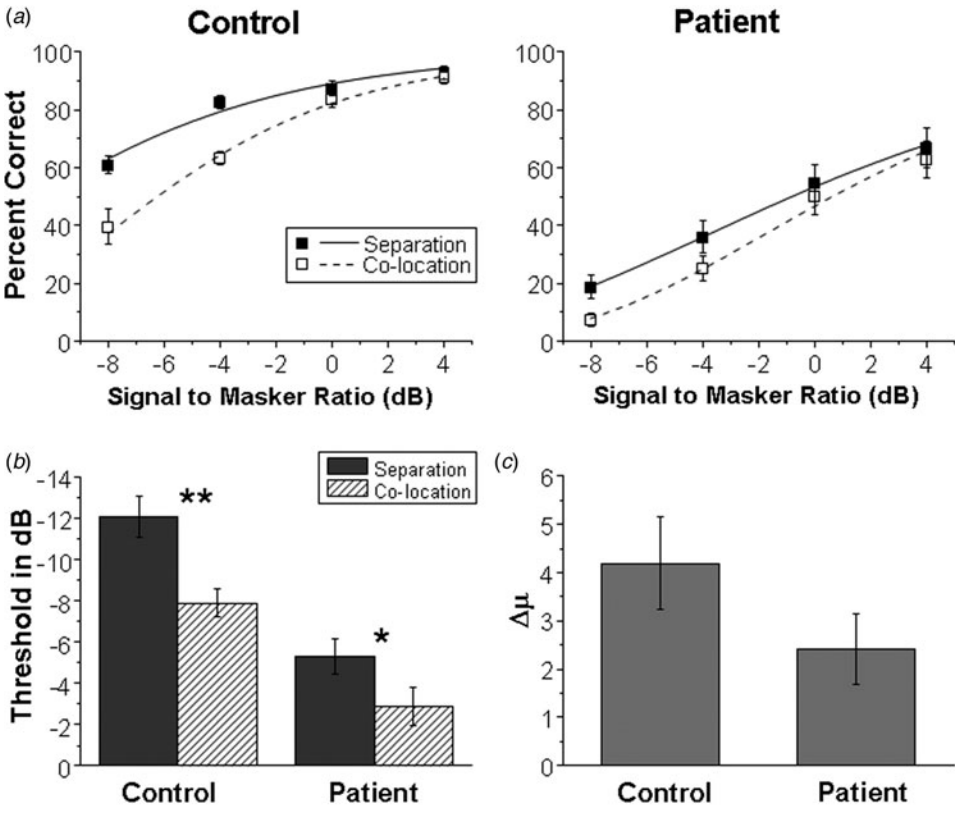


Fig. 1. (*a*) Comparisons in group mean percentage correct recognition of the target sentence between the healthy group (left panel) and the patient group (right panel) as a function of the signal-to-masker ratio along with the group mean best-fitting psychometric functions, under either the perceived spatial co-location (PSC) condition or the perceived spatial separation (PSS) condition. (*b*) Group mean thresholds (μ in dB) for 50% correct recognition of the keywords in target sentences in the two participant groups under either the PSS or PSC condition. * $p < 0.05$, ** $p < 0.01$. (*c*) No group difference in the change in threshold μ ($\Delta\mu$ in dB) was induced by the PSS condition (compared with the PSC condition). Values are means, with standard errors represented by vertical bars.

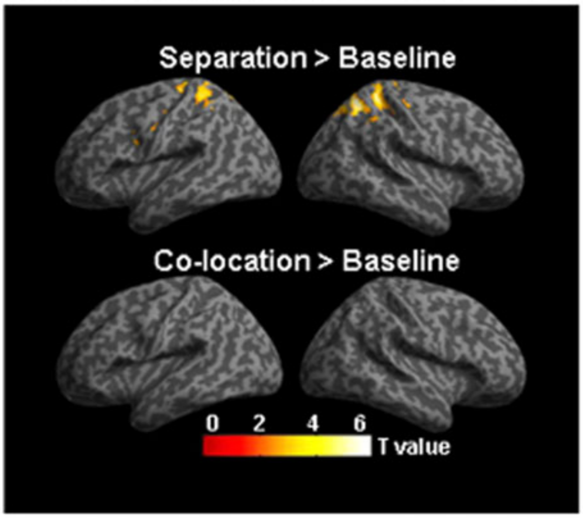
The threshold reduction ($\Delta\mu$) induced by PSS was the difference in μ between the PSS condition and the PSC condition. A one-way ANOVA showed that there was no significant difference in $\Delta\mu$ between the groups ($F_{1,34} = 2.255$, $p = 0.142$).

Online Supplementary Fig. S2 shows that the mean percentage correct button-press response for detecting target keywords during fMRI scanning was poorer in patients than healthy controls.

fMRI experiment

Difference in brain regions activated by simulated 'cocktail party' listening conditions between healthy controls and patients

Under the PSS listening condition, significantly reduced BOLD signals in the bilateral superior parietal lobule (SPL), right precuneus, left middle cingulate cortex (MCC) and left caudate ($p < 0.05$, FDR corrected) were found in patients compared with healthy controls (Fig. 2*a*, online Supplementary Table S2). Under the PSC condition, no significant difference in brain activation was found between controls and patients ($p < 0.05$, FDR corrected). Moreover, Spearman correlation analyses were conducted to identify the brain regions whose activation was related to the target-speech recognition. The ROIs included the brain structures listed in online Supplementary Table S2. The parameter estimates of signal intensity for each ROI were extracted from each individual participant, and then both the mean CV for the PSS condition (the parameter estimate for the PSS condition minus the parameter estimate for the baseline, and averaged across participants) and the mean CV for the PSC condition (the parameter estimate for the PSC condition minus the parameter estimate for the baseline, and averaged across participants) for each ROI were calculated (Wild *et al.* 2012). Under the PSS condition (Fig. 2), significant negative correlation was found between the threshold μ of speech-recognition performance and the CV for 'PSS > baseline' in the left SPL ($r = -0.550$, $p = 0.027$) and right SPL ($r = -0.538$, $p = 0.021$) in healthy controls, (*a*) Group Differences Related to either PSS or PSC Condition



(*b*) Correlation between Brain Signal and Recognition Threshold μ

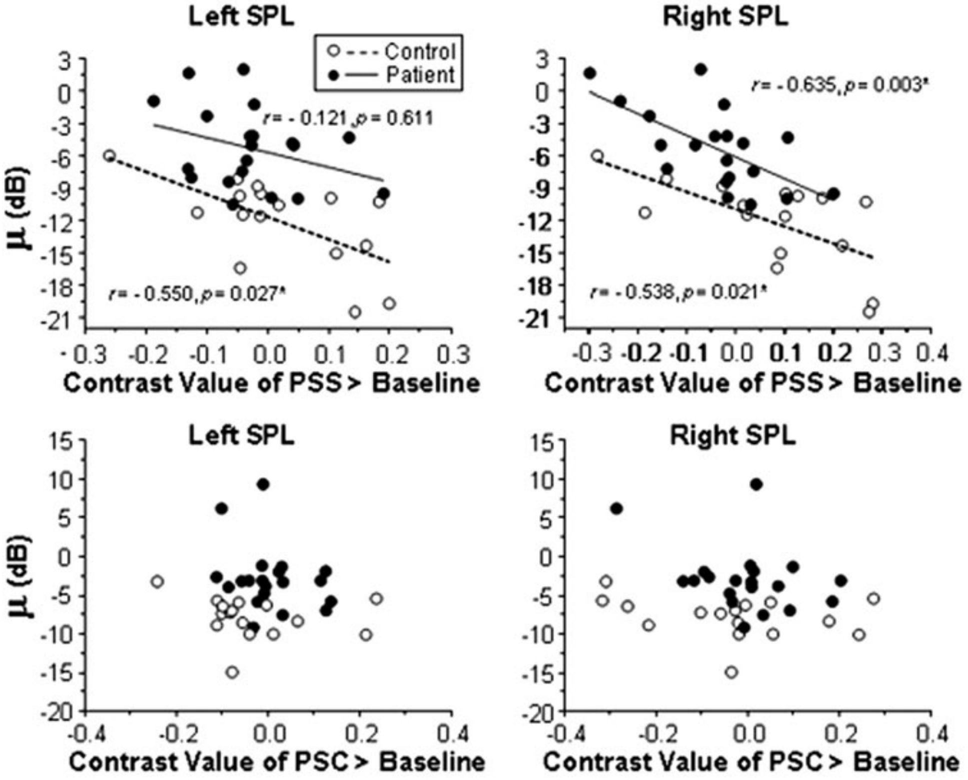


Fig. 2. (*a*) Brain regions associated with significant activation differences between controls and patients under either the 'perceived spatial separation (PSS) > baseline' condition or the 'perceived spatial co-location (PSC) > baseline' condition. The 'PSS > baseline' contrast represents the brain activation related to target-speech recognition against informational masking when the target and masker were perceived from different locations. The 'PSC > baseline' contrast represents the brain activation related to target-speech recognition against informational masking when the target and masker were perceived from the same location. The activation map was thresholded at $p < 0.05$ (false discovery rate corrected with activation of more than 10 contiguous voxels) and overlaid on the canonical template of SPM8 software. (*b*) The contrast value (CV) of 'PSS > baseline' (difference in percentage of signal change between the separation condition and baseline condition) in the left superior parietal lobule (SPL) was negatively correlated with the μ value in healthy controls but not patients. The CV of 'PSS > baseline' in the right SPL was negatively correlated with the μ value for both controls and patients. The CV of 'PSC > baseline' in the bilateral SPL showed no significant correlation with the μ value for both controls and patients.

and in the right SPL ($r = -0.635$, $p = 0.003$) in patients. Under the PSC condition (Fig. 2*b*), no significant correlation was found between μ and CV for the contrast of 'PSC > baseline'.

Brain regions activated by the PSS condition against the PSC condition

Fig. 3*a* and online Supplementary Table S3 show that in healthy controls, compared with the PSC listening condition, introducing the PSS listening condition significantly enhanced BOLD signals in the bilateral SPL, right precuneus, bilateral anterior cingulate cortex (ACC), left caudate, pars triangularis of the left inferior frontal gyri (TriIFG) and left middle frontal gyrus (MFG) ($p < 0.05$, FDR corrected). Also, Spearman correlation analyses showed that the PSS condition-induced BOLD signal enhancement (CV for 'PSS > PSC') in the left ACC was significantly correlated with the performance threshold reduction ($\Delta\mu$) induced by the PSS condition ($r = 0.568$, $p = 0.009$) (Fig. 3*b*). However, compared with the PSS condition, introducing the PSC condition did not induce significant changes in BOLD signal ($p < 0.05$, FDR corrected).

Fig. 3*a* and online Supplementary Table S3 also show that in patients, compared with the PSC condition, introducing the PSS condition did not activate either the SPL or the precuneus, but activated the left MFG, left TriIFG and left medial superior frontal gyrus (mSFG). Also, Spearman correlation analyses showed that the PSS condition-induced BOLD signal enhancement (CV for 'PSS > PSC') in the left mSFG was significantly correlated with the $\Delta\mu$ ($r = 0.552$, $p = 0.013$) (Fig. 3*b*). Moreover, introducing the PSC condition (compared with the PSS condition) enhanced BOLD signals in the bilateral SPL and bilateral precuneus, implying a deactivation of the SPL by the shift from the PSC condition to the PSS condition.

Significantly reduced BOLD signals in the contrast of 'PSS > PSC' in the bilateral SPL were revealed in patients compared with healthy controls (Fig. 3*c* and online Supplementary Table S4).

Correlation between brain activation induced by the PSS condition and psychotic symptoms

Significantly positive correlations were found between the CV of the left mSFG for the 'PSS > PSC' contrast and the positive symptom score of the PANSS, and between the CV and the P6-item score (suspiciousness/persecution) of the PANSS (Fig. 4).

PPI analyses

PPI analyses were conducted to identify the brain regions that exhibited functional connectivity with the bilateral SPL for the PSS > PSC contrast in controls and patients. The results showed that in healthy controls the enhanced functional connectivity of the left SPL was observed with the right TriIFG, bilateral caudate and left thalamus, and the enhanced functional connectivity of the right SPL was observed with the right TriIFG, right caudate and left thalamus. In patients, however, the significantly enhanced functional connectivity of the right SPL was observed with the left parahippocampus gyrus (PHG) and left angular (Fig. 5*a* and online Supplementary Table S5).

Compared with those in healthy controls, in patients the functional connectivities of the left SPL with the pars opercularis of the right inferior frontal gyrus (OperIFG), right rolandic operculum (RO) and right caudate were significantly reduced (Fig. 5*b* and online Supplementary Table S6). Also, the functional connectivities of the right SPL in patients with the right OperIFG and right RO were also significantly reduced (Fig. 5*b* and online Supplementary Table S6).

Discussion

Schizophrenia-related deficits of speech recognition against informational masking

The behavioural test results of this study indicate that in people with schizophrenia, speech recognition against a two-talker speech masker is worse than that in healthy listeners when both spatial attention and masker suppression are involved in the performance. Thus, the results support the view that the augmented vulnerability of speech recognition to informational masking is an essential feature of schizophrenia (Wu *et al.* 2012, 2013).

This study for first time discovers that speech recognition against informational masking is correlated with the target-speech-induced activation of the bilateral SPL in healthy listeners and that of the right SPL in listeners with schizophrenia under the PSS condition. The loss of correlation between the speech-recognition performance (threshold μ) and the target-speech-induced activation of the left SPL in listeners with schizophrenia suggests a schizophrenia-related functional impairment of the left SPL under 'cocktail party' listening conditions. Also, under the PSS condition, compared with healthy listeners, the target speech-induced brain activation in listeners with schizophrenia declined not only in the SPL but also in the right precuneus, left caudate and left MCC.

It has been reported that in 'cocktail party' listening situations, the SPL is involved in directing attention to one particular talker (Hill & Miller, 2010) and becomes transiently activated during voluntary shifts of

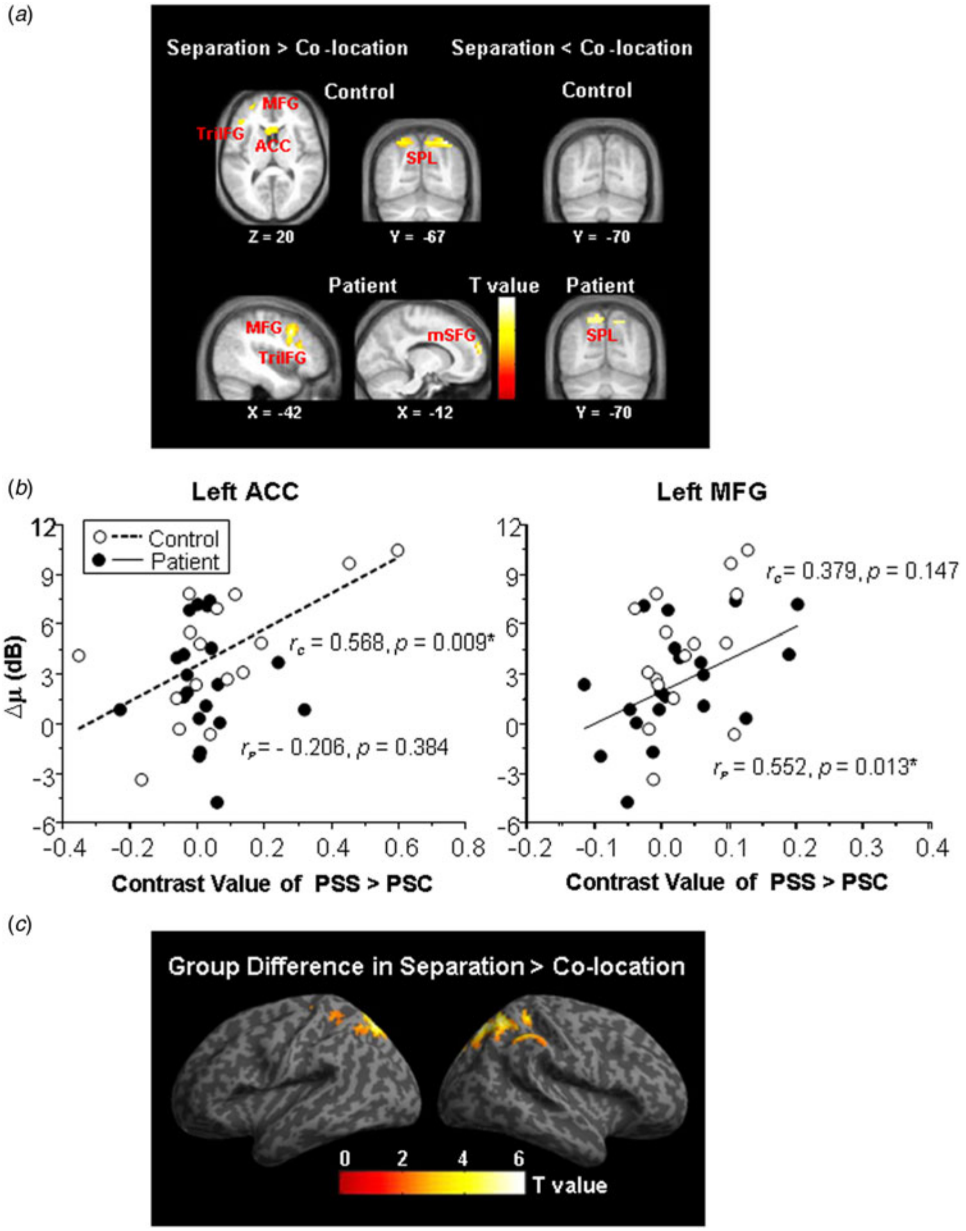


Fig. 3. (*a*) Brain regions activated by the contrast of 'perceived spatial separation (PSS) > perceived spatial co-location (PSC)' in the healthy control group (upper panel) and those in the patient group (lower panel). The activation map was thresholded at $p < 0.05$ [false discovery rate (FDR) corrected with activation of more than 10 contiguous voxels] and overlaid on the group average structural image. ACC, Anterior cingulate cortex; TriIFG, pars triangularis of inferior frontal gyrus; MFG, middle frontal gyrus; mSFG, medial superior frontal gyrus; SPL, superior parietal lobule. (*b*) The contrast value (CV) of 'PSS > PSC' in the left ACC was positively correlated with the $\Delta\mu$ value only in controls; the CV of 'PSS > PSC' in the left MFG was positively correlated with the $\Delta\mu$ value only in patients. (*c*) Difference in brain regions activated by the CV of 'PSS > PSC' between controls and patients. The activation map was thresholded at $p < 0.05$ (FDR corrected with activation of more than 10 contiguous voxels) and overlaid on the canonical template of SPM8 software.

attention (Yantis *et al.* 2002; Shomstein & Yantis, 2006; Serences & Yantis, 2007). Moreover, the precuneus is involved in computing the exact spatial location of the target sound source (Zündorf *et al.* 2013). Also, 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). Clearly, the SPL, precuneus and MCC are the critical components in the network mediating attention/response selection (Schumacher *et al.* 2003; Cavanna & Trimble, 2006; Shackman *et al.* 2011). Particularly the SPL and precuneus are

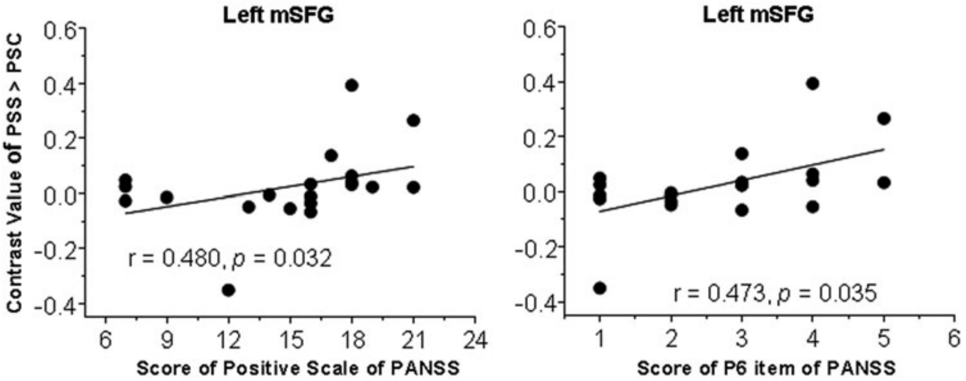


Fig. 4. Spearman correlation analyses showed that the contrast value of 'perceived spatial separation (PSS) > perceived spatial co-location (PSC)' in the left medial superior frontal gyrus (mSFG) was positively correlated with the positive symptom score of Positive and Negative Syndrome Scale (PANSS) (left panel), and the P6-item score (suspiciousness/persecution) of the PANSS (right panel) in patient participants.

specialized for the processing of spatial attributes (Renier *et al.* 2009).

On the other hand, the SPL plays a role in suppressing irrelevant distracters to ensure accurate target selection in the competition between target and distracters (Wojciulik & Kanwisher, 1999; Pollmann *et al.* 2003; Krueger *et al.* 2007). Also, 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).

Thus, the brain regions beyond the traditional auditory system, including the posteromedial parietal cortex (containing the SPL and precuneus) and caudate, are normally involved in speech listening under reverberant conditions with multiple people talking, based on the integrated functions of spatial processing, attention direction and irrelevant-stimulus suppression.

It has been reported that there is a progressive reduction in working memory-induced modulation of functional connectivity between the right SPL and the frontal cortex across healthy controls, at-risk mental state participants and first-episode schizophrenic patients (Schmidt *et al.* 2013). A recent study has even shown that abnormal connectivity between the right SPL and right MFG during working memory processing is already evident in individuals with an at-risk mental state and is related to psychiatric symptoms (Schmidt *et al.* 2014). Thus, the right SPL is functionally impaired in people with schizophrenia. By introducing the listening condition with PSS between the target-speech 'image' and the masker-speech 'image', this study reveals the most critical brain substrates (particularly the SPL, precuneus, caudate and MCC) that underlie schizophrenia-related deficits of speech recognition under 'cocktail party' listening conditions when both spatial attention and masker suppression are required.

Brain regions involved in PSS of speech sources

In both the current study and previous studies, shifts between the PSS and PSC listening conditions did not affect peripheral processes (Li *et al.* 2004). Thus, the contrast in BOLD signal for the separation condition against the co-location condition represents cortical activities that are associated with higher-order perceptual processes. Previous studies have shown that the auditory cortex is involved in the masking of speech (Scott & McGettigan, 2013). This study further shows that the perceptual separation between target speech and masking speech involves an enhancement of BOLD signals in some brain regions beyond the traditional central auditory system: the bilateral SPL, right precuneus, bilateral ACC, left caudate, left TriIFG and left MFG.

As mentioned above, both the SPL and precuneus are critical in mediating spatial attention/response selection (Yantis *et al.* 2002; Schumacher *et al.* 2003; Cavanna & Trimble, 2006; Shomstein & Yantis, 2006; Serences & Yantis, 2007; Renier *et al.* 2009; Hill & Miller, 2010; Zündorf *et al.* 2013). In fact, both the ACC and MFG are also involved in suppressing irrelevant distracters to ensure accurate target selection in the competition between target and distracters (ACC: Fornito *et al.* 2011; Schulz *et al.* 2011; Shenhav *et al.* 2013; MFG: Lesh *et al.* 2011; Sokol-Hessner *et al.* 2012; Jeurissen *et al.* 2014). When target speech and masking speech are perceived as spatially separated, although the listener's spatial attention to the perceived target location is facilitated, the masker also draws the listener's attention (especially when the SMR is low, e.g.

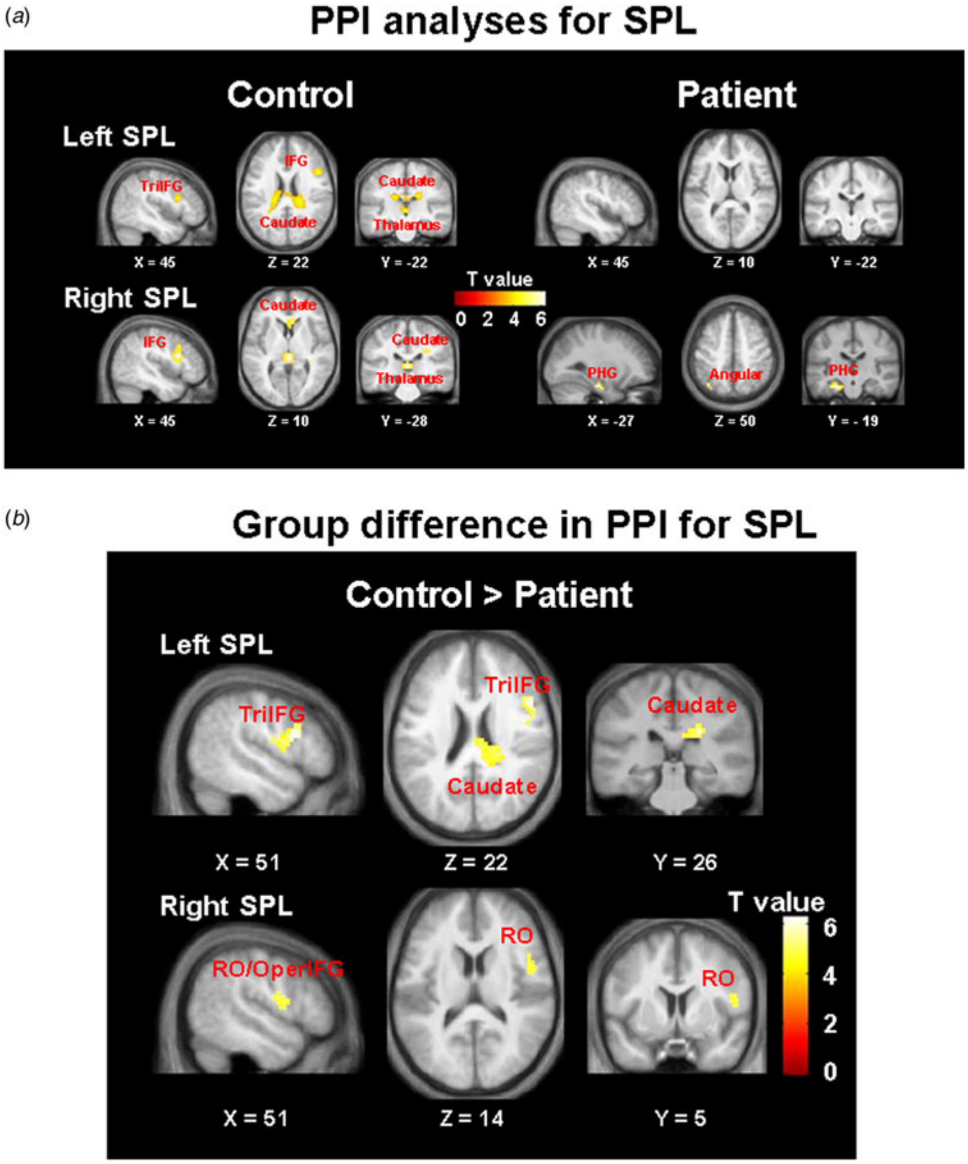


Fig. 5. (*a*) Psychophysiological interaction (PPI) analyses of functional connectivities of the superior parietal lobule (SPL) associated with the 'perceived spatial separation > perceived spatial co-location' contrast in controls and patients. (*b*) Difference in PPI analyses of functional connectivities of the SPL between controls and patients. The activation maps are displayed on the group average structural image. All peaks were significant at $p < 0.05$ (false discovery rate corrected with activation more than 10 contiguous voxels). TriIFG, Pars triangularis of the left inferior frontal gyri; IFG, inferior frontal gyrus; PHG, parahippocampal gyrus; RO, rolandic operculum; OperIFG, pars opercularis of the inferior frontal gyrus.

−4 dB). The spatial attention/response competition occurs between the two perceived locations. Thus, the separation-induced activation of these brain regions in healthy listeners represents a normal engagement of neural processing for facilitating attention to the target speech.

Also as mention above, the SPL and caudate also play a role in suppressing irrelevant distracters (Wojciulik & Kanwisher, 1999; Menon *et al.* 2001; Pollmann *et al.* 2003; Krueger *et al.* 2007; Ketteler *et al.* 2008; Li *et al.* 2008; Ali *et al.* 2010). The separation-induced enhancement of activation in the SPL and caudate represents an enhanced suppression of the masker (whose auditory image becomes outside the attention focus under the perceptual separation condition).

As mentioned in the Introduction, people with schizophrenia can use temporally pre-presented content cues (Wu *et al.* 2012) to unmask target speech and exhibit the intactness of the precedence effect (Mickey & Dalack, 2005). For the first time, this study provides evidence showing that similar to healthy listeners, listeners with schizophrenia are able to benefit from the PSS to improve their recognition of target speech. However, as discovered by this study, the underlying brain substrates are different between listeners with schizophrenia and healthy listeners. One of the most striking findings of this study is that introducing the PSS listening condition (relative to the PSC condition) activates the bilateral SPL and right precuneus in healthy listeners but suppresses the bilateral SPL and bilateral precuneus in patients with schizophrenia.

Moreover, the separation-induced left ACC activation is significantly correlated with the separation-induced speech-recognition improvement (measured by $\Delta\mu$) in healthy listeners but not in patients; the separation-induced left MFG activation is significantly correlated with the separation-induced speech-recognition improvement ($\Delta\mu$) in patients but not in healthy listeners. We propose that normally the left ACC plays a role in specifically mediating the unmasking effect of PSS, and the perceptual separation-induced activation of the left MFG reflects a neural compensatory strategy for the impairment of both spatial attention to target signals and suppression of irrelevant signals that involve the ACC and SPL.

Interestingly, the PSS listening condition activates the mSFG only in patients, in whom this perceptual separation-induced activation of the left mSFG is correlated with the patients' positive syndrome. Thus, there is a link between the positive syndrome and the PSS-induced signal gating that involves the activation of the left mSFG. Previous studies have shown that the mSFG is involved in controlling goal-directed behaviour through the stable maintenance of task sets (Dosenbach *et al.* 2007), selecting action sets (Rushworth *et al.* 2004) and representing an increase of attentional load (Mazoyer *et al.* 2002; Vickery & Jiang, 2009). Thus, people with schizophrenia may utilize a different neural strategy to achieve PSS between speech sources when the SPL activation becomes abnormal in certain task situations. Since the recruitment of compensatory networks in people with schizophrenia is an important issue (Tan *et al.* 2007), the separation-induced activation in the mSFG suggests that under reverberant environments there can be certain plasticity of neural mechanisms underlying PSS between speech sources.

PSS-related functional connectivities

Previous studies have suggested that 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) and lexical-tone production (Liu *et al.* 2006). Since the motor-system involvement in speech perception is task-load dependent and the function of the auditory–motor link is important for speech perception under adverse listening conditions (Wu *et al.* 2014), the normally enhanced functional connectivity of the SPL with the IFG suggests an enhanced involvement of the speech-production system to deal with 'cocktail party' speech listening situations. In healthy listeners the PSS listening condition enhances the functional connectivities of the bilateral SPL with the right TriIFG, bilateral caudate and left thalamus (online Supplementary Table S5). In listeners with schizophrenia the functional connectivities of the SPL significantly decline (online Supplementary Table S6). As a substitution, the functional connectivities of the right SPL with the left PHG and left angular are enhanced by the PSS listening condition (online Supplementary Table S5).

The results further support the view that introducing the PSS listening condition is an efficient way to reveal the schizophrenia-related functional changes of the brain structures involved in speech recognition under 'cocktail party' listening conditions.

The default mode network (DMN)

The DMN exhibits both correlated activity among the network component brain structures at resting status and decreased activation during performance of cognitive tasks (Raichle *et al.* 2001; Whitfield-Gabrieli *et al.* 2009; Andrews-Hanna, 2012; Zhang & Li, 2012). One of the findings of this study is that a shift from the PSC condition to the PSS condition (i.e. the 'separation > co-location' analysis), which normally reduces the perceptual load for processing target speech against informational masking, reveals not only an increased activation of the precuneus and the ventral ACC (two of the component brain structures in the DMN) in healthy listeners, but also a decreased activation of the precuneus in patients with schizophrenia. In addition, the 'separation > co-location' contrast reveals an increased activation of the mSFG (another component structure in the DMN) in patients with schizophrenia but not in healthy controls. The results of this study both confirm previous findings that schizophrenia markedly affects the DMN (Whitfield-Gabrieli *et al.* 2009; Repovs *et al.* 2011; Yu *et al.* 2012; Andrews-Hanna *et al.* 2014; Gao *et al.* 2015) and reveal that different component structures of the DMN are affected differently by schizophrenia. Particularly, the activation of the mSFG, which is correlated with the positive syndrome of patients (this study; Gao *et al.* 2015), may reflect a specific schizophrenia-related compensatory strategy for the functional impairments of the precuneus and ACC.

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291715001828

Acknowledgements

This work was supported by the '973' National Basic Research Program of China (2011CB707805), the Chinese National Key Clinical Program in Psychiatry to Guangzhou Brain Hospital (201201004), the Planned Science and Technology Projects of Guangzhou (2014Y2-00105), the China Postdoctoral Science Foundation General Program (2013M530453) and the National Natural Science Foundation of China (31170985). Huahui Li assisted in many aspects of this work.

Declaration of Interest

None.

References

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.

Andrews-Hanna JR (2012). The brain's default network and its adaptive role in internal mentation. *Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry* **18**, 251–270.

Andrews-Hanna JR, Smallwood J, Spreng RN (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences* **1316**, 29–52.

Apps MAJ, Lockwood PL, Balsters JH (2013). The role of the midcingulate cortex in monitoring others' decisions. *Frontiers in Neuroscience* **7**, 251.

Brungart DS, Simpson BD, Freyman RL (2005). Precedence-based speech segregation in a virtual auditory environment. *Journal of the Acoustical Society of America* **118**, 3241–3251.

Carreiras M, Mechelli A, Estévez A, Price CJ (2007). Brain activation for lexical decision and reading aloud: two sides of the same coin? *Journal of Cognitive Neuroscience* **19**, 433–444.

Cavanna AE, Trimble MR (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* **129**, 564–583.

Culling JF, Hodder KI, Toh CY (2003). Effects of reverberation on perceptual segregation of competing voices. *Journal of the Acoustical Society of America* **114**, 2871–2876.

Darwin CJ, Hukin RW (2000). Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *Journal of the Acoustical Society of America* **108**, 335–342.

Ding N, Simon JZ (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11854–11859.

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11073–11078.

First MB, Spitzer RL, Gibbon M, Williams JB (2012). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), Clinician Version. Administration Booklet*. American Psychiatric Press, Inc.: Washington, DC.

Fish SC, Granholm E (2008). Easier tasks can have higher processing loads: task difficulty and cognitive resource limitations in schizophrenia. *Journal of Abnormal Psychology* **117**, 355–363.

Fornito A, Yoon J, Zalesky A, Bullmore ET, Carter CS (2011). General and specific functional connectivity disturbances in first-episode schizophrenia during cognitive control performance. *Biological Psychiatry* **70**, 64–72.

Freyman RL, Clifton RK, Litovsky RY (1991). Dynamic processes in the precedence effect. *Journal of the Acoustical Society of America* **90**, 874–884.

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.

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 B, Wang Y, Liu W, Chen Z, Zhou H, Yang J, Cohen Z, Zhu Y, Zang Y (2015). Spontaneous activity associated with delusions of schizophrenia in the left medial superior frontal gyrus: a resting-state fMRI study. *PLOS ONE* **10**, e0133766.

Gjerde PF (1983). Attentional capacity dysfunction and arousal in schizophrenia. *Psychological Bulletin* **93**, 57–72.

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.

Helfer KS (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech, Language, and Hearing Research* **40**, 432–443.

Hill KT, Miller LM (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral 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. *Hearing Research* **244**, 51–65.

Huang Y, Li J, Zou X, Qu T, Wu X, Mao L, Wu Y, Li L (2011). Perceptual fusion tendency of speech sounds. *Journal of Cognitive Neuroscience* **23**, 1003–1014.

Jeurissen D, Sack AT, Roebroeck A, Russ BE, Pascual-Leone A (2014). TMS affects moral judgment, showing the role of

DLPFC and TPJ in cognitive and emotional processing. *Frontiers in Neuroscience* **8**, 18.

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.

Kidd G, Mason CR, Brughera A, Hartmann WM (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica United with Acustica* **91**, 526–536.

Koehnke J, Besing JM (1996). A procedure for testing speech intelligibility in a virtual listening environment. *Ear and Hearing* **17**, 211–217.

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 Research* **1174**, 110–119.

Lesh TA, Niendam TA, Minzenberg MJ, Carter CS (2011). Cognitive control deficits in schizophrenia: mechanisms and meaning. *Neuropsychopharmacology* **36**, 316–338.

Li CSR, Yan P, Sinha R, Lee TW (2008). Subcortical processes of motor response inhibition during a stop signal task. *NeuroImage* **41**, 1352–1363.

Li H, Kong L, Wu X, Li L (2013). Primitive auditory memory is correlated with spatial unmasking that is based on direct-reflection integration. *PLOS ONE* **8**, e63106.

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.

Li L, Qi JG, He Y, Alain C, Schneider BA (2005). Attribute capture in the precedence effect for long-duration noise sounds. *Hearing Research* **202**, 235–247.

Liu L, Peng D, Ding G, Jin Z, Zhang L, Li K, Chen C (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *NeuroImage* **29**, 515–523.

Mazoyer P, Wicker B, Fonlupt P (2002). A neural network elicited by parametric manipulation of the attention load. *Neuroreport* **13**, 2331–2334.

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK (2012). Speech comprehension aided by multiple modalities: behavioural and neural interactions. *Neuropsychologia* **50**, 762–776.

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.

Mickey BJ, Dalack GW (2005). Auditory gating in schizophrenia: a pilot study of the precedence effect. *Schizophrenia Research* **73**, 327–331.

Nuechterlein KH, Dawson ME (1984). Information processing and attentional functioning in the developmental course of schizophrenic disorders. *Schizophrenia Bulletin* **10**, 160–203.

Nuechterlein KH, Dawson ME, Green MF (1994). Information-processing abnormalities as neuropsychological vulnerability indicators for schizophrenia. *Acta Psychiatrica Scandinavica* **90**, 71–79.

Pollmann S, Weidner R, Humphreys GW, Olivers CN, Müller 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.

Qu T, Xiao Z, Gong M, Huang Y, Li X, Wu X (2008). Distance dependent head-related transfer function database of KEMAR. In *IEEE International Conference on Audio, Language and Image Processing*, 2008, pp. 466–470. IEEE. http://www.ieee.org/conferences_events/conferences/conferencedetails/index.html?Conf_ID=13439.

Qu T, Xiao Z, Gong M, Huang Y, Li X, Wu X (2009). Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. *IEEE Transactions on Audio, Speech, and Language Processing* **17**, 1124–1132.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 676–682.

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.

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". *Journal of Neuroscience* **29**, 10950–10960.

Repovs G, Csernansky JG, Barch DM (2011). Brain network connectivity in individuals with schizophrenia and their siblings. *Biological Psychiatry* **69**, 967–973.

Rushworth M, Walton ME, Kennerley SW, Bannerman DM (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences* **8**, 410–417.

Schmidt A, Smieskova R, Aston J, Simon A, Allen P, Fusar-Poli P, McGuire PK, Riecher-Rössler A, Stephan KE, Borgwardt S (2013). Brain connectivity abnormalities predating the onset of psychosis: correlation with the effect of medication. *JAMA Psychiatry* **70**, 903–912.

Schmidt A, Smieskova R, Simon A, Allen P, Fusar-Poli P, McGuire PK, Bendfeldt K, Aston J, Lang UE, Walter M, Radue EW, Rössler AR, Borgwardt SJ (2014). Abnormal effective connectivity and psychopathological symptoms in the psychosis high-risk state. *Journal of Psychiatry and Neuroscience* **39**, 239–248.

Schulz KP, Bédard A-CV, Czarnecki R, Fan J (2011). Preparatory activity and connectivity in dorsal anterior cingulate cortex for cognitive control. *NeuroImage* **57**, 242–250.

Schumacher EH, Elston PA, D'Esposito M (2003). Neural evidence for representation-specific response selection. *Journal of Cognitive Neuroscience* **15**, 1111–1121.

Scott SK, McGettigan C (2013). The neural processing of masked speech. *Hearing Research* **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. *Journal of the Acoustical Society of America* **115**, 813–821.

Scott SK, Wise RJS (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication* **41**, 23–34.

Seidman LJ, Van Manen K, Turner WM, Gamser DM, Faraone SV, Goldstein JM, Tsuang MT (1998). The effects of increasing resource demand on vigilance performance in adults with schizophrenia or developmental attentional/learning disorders: a preliminary study. *Schizophrenia Research* **34**, 101–112.

Serences JT, Yantis S (2007). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cerebral Cortex* **17**, 284–293.

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. *Nature Review Neuroscience* **12**, 154–167.

Shenhav A, Botvinick MM, Cohen JD (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240.

Shomstein S, Yantis S (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *Journal of Neuroscience* **26**, 435–439.

Si TM, Yang JZ, Shu L, Wang XL, Kong QM, Zhou M, Li XN (2004). The reliability, validity of PANSS (Chinese version), and its implication. *Chinese Mental Health Journal* **18**, 45–47.

Sokol-Hessner P, Hutcherson C, Hare T, Rangel A (2012). Decision value computation in DLPFC and VMPFC adjusts to the available decision time. *European Journal of Neuroscience* **35**, 1065–1074.

Tan H, Callicott JH, Weinberger DR (2007). Dysfunctional and compensatory prefrontal cortical systems, genes and the pathogenesis of schizophrenia. *Cerebral Cortex* **17**, i171–i181.

Verleger R, Talamo S, Simmer J, Śmigasiewicz K, Lencer R (2013). Neurophysiological sensitivity to attentional overload in patients with psychotic disorders. *Clinical Neurophysiology* **124**, 881–892.

Vickery TJ, Jiang YV (2009). Inferior parietal lobule supports decision making under uncertainty in humans. *Cerebral Cortex* **19**, 916–925.

Vouloumanos A, Kiehl K, Werker J, Liddle P (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience* **13**, 994–1005.

Wallach H, Newman EB, Rosenzweig MR (1949). The precedence effect in sound localization. *American Journal of Psychology* **62**, 315–336.

Whitfield-Gabrieli S, Thermenos HW, Milanovic S, Tsuang MT, Faraone SV, McCarley RW, Shenton ME, Green AI, Nieto-Castanon A, LaViolette P, Wojcik J, Gabrieli JD, Seidman LJ (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1279–1284.

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

Wojciulik E, Kanwisher N (1999). The generality of parietal involvement in visual attention. *Neuron* **23**, 747–764.

Wu C, Cao S, Wu X, Li L (2013). Temporally pre-presented lipreading cues release speech from informational masking. *Journal of the Acoustical Society of America* **133**, L281–L285.

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. *Schizophrenia Research* **134**, 33–41.

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 ZM, Chen ML, Wu XH, 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, Li L (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication* **49**, 892–904.

Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience* **5**, 995–1002.

Yu Q, Allen EA, Sui J, Arbabshirani MR, Pearlson G, Calhoun VD (2012). Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. *Current Topics in Medicinal Chemistry* **12**, 2415–2425.

Zhang S, Li C-SR (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *NeuroImage* **59**, 3548–3562.

Zündorf IC, Lewald J, Karnath H (2013). Neural correlates of sound localization in complex acoustic environments. *PLOS ONE* **8**, e64259.

Zurek PM (1980). The precedence effect and its possible role in the avoidance of interaural ambiguities. *Journal of the Acoustical Society of America* **67**, 952–964.

Zurek PM, Freyman RL, Balakrishnan U (2004). Auditory target detection in reverberation. *Journal of the Acoustical Society of America* **115**, 1609–1620.