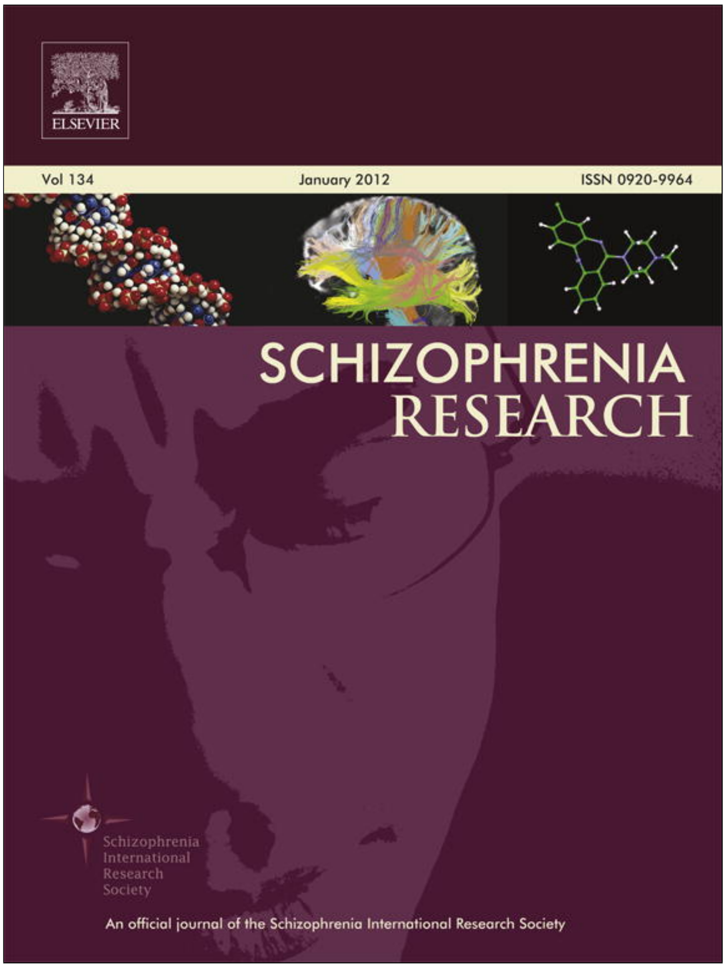
ELSEVIER
Vol 134
January 2012
ISSN 0920-9964
SCHIZOPHRENIA
RESEARCH
Schizophrenia
International
Research
Society
An official journal of the Schizophrenia International Research Society

Schizophrenia Research 134 (2012) 33–41

Contents lists available at SciVerse ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia

Chao Wu [a], Shuyang Cao [a], Fuchun Zhou [b], Chuanyue Wang [b], Xihong Wu [a], Liang Li [a,*]

[a] Department of Psychology, Department of Machine Intelligence, Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871, China

[b] Beijing Anding Hospital, Capital Medical University, Beijing 100088, China

ARTICLE INFO

Article history:
Received 24 April 2011
Received in revised form 17 September 2011
Accepted 18 September 2011
Available online 21 October 2011

Keywords:
Schizophrenia
Speech perception
Energetic masking
Informational masking
Content priming
Voice priming

ABSTRACT

In "cocktail-party" environments, although listeners feel it difficult to recognize attended speech due to both energetic masking and informational masking, they can use various perceptual/cognitive cues, such as content and voice primes, to facilitate their attention to target speech. In patients with schizophrenia, both speech-perception deficits and increased vulnerability to masking stimuli generally occur. This study investigated whether speech recognition in first-episode patients (FEPs) and chronic patients (CPs) of schizophrenia is more vulnerable to noise masking and/or speech masking than that in demographics-matched-healthy controls, and whether patients with schizophrenia can use primes to unmask speech. In a trial under the priming condition, before the target sentence containing three keywords was co-presented with a noise or speech masker, the prime (early part of the sentence including the first two keywords) was recited in quiet with the target-speaker's voice. The results show that in patients, target-speech recognition was more impaired under speech-masking conditions than noise-masking conditions, and the impairment in CPs (n = 22) was larger than that in FEPs (n = 12). Although working memory for holding prime-content information in patients, especially CPs, was more vulnerable to masking, especially speech masking, than that in healthy controls, patients were still able to use the prime to unmask the last keyword. Thus, in "cocktail-party" environments, speech recognition in people with schizophrenia is more vulnerable to masking, particularly informational masking, and the speech-recognition impairment augments as the illness progresses. However, people with schizophrenia can use the content/voice prime to reduce energetic masking and informational masking of target speech.

1. Introduction

Under "cocktail-party" environments with multiple-people talking (Cherry, 1953), there are two types of masking components contributing to speech-recognition difficulties: energetic masking and informational masking (e.g., Freyman et al., 1999; Brungart et al., 2001; Li et al., 2004; Helfer and Freyman, 2005, 2009; Wu et al., 2005; Rakerd et al., 2006; Huang et al., 2008, 2009). Energetic masking occurs when peripheral neural activity elicited by a signal is overwhelmed by that elicited by maskers, leading to a degraded or noisy neural representation of the signal. Steady-state speech-spectrum noise has been widely used for introducing energetic masking (noise-on-speech masking). Informational masking occurs when signals and maskers are similar along some informational dimensions, particularly when signals and maskers are speech sounds (speech-on-speech masking), causing confusion 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).

Compared to healthy listeners, patients with schizophrenia have impairments in both early-auditory processing (e.g., Javitt et al., 1997, 2000; Michie, 2001; Alain et al., 2002; Kasai et al., 2002a,b; Brenner et al., 2009) and speech perception (e.g., Bull and Venables, 1974; DeLisi, 2001; Condray et al., 2002; Kasai et al., 2002a,b; Lee et al., 2004; Titone and Levy, 2004; Condray, 2005). Moreover, people with schizophrenia are more vulnerable to auditory forward/backward masking and visual backward masking than their healthy controls (Green et al., 1999; Källstrand et al., 2002), perform poorly in deriving benefit from viewing a speaker's speech articulators against noise masking (Ross et al., 2007), and make more errors in the "babble"-masked speech tracking task (particularly for those with hallucination, Hoffman et al., 1999; Lee et al., 2004). However, it is not clear whether compared to that in healthy listeners, speech recognition in people with schizophrenia is more vulnerable to masking, particularly informational masking. The Ross et al. study (2007) shows that compared to healthy participants, schizophrenic participants did not have any deficits in

* Corresponding author at: Department of Psychology, Peking University, Beijing 100871, China. Fax: +86 10 6276 1081.
E-mail address: liangli@pku.edu.cn (L. Li).

0920-9964/$ – see front matter
doi:10.1016/j.schres.2011.09.019

recognizing monosyllabic words against the background of a pink noise (also called 1/*f* noise whose power spectral density is inversely proportional to the frequency *f*), suggesting that low-load speech recognition is not impaired under energetic-masking conditions.

Although listeners usually feel it difficult to recognize target speech under "cocktail-party" conditions, they can use various perceptual/cognitive cues to reduce masking of target speech. These cues include perceived spatial separation between target and masker (Freyman et al., 1999; Arbogast et al., 2002; Li et al., 2004; Wu et al., 2005; Rakerd et al., 2006; Huang et al., 2008, 2009), prior knowledge about where and/or when target speech will occur (Kidd et al., 2005; Best et al., 2007), prior knowledge about part of the target-sentence content (Freyman et al., 2004; Yang et al., 2007; Ezzatian et al., 2011), familiarity of the target-talker's voice (Brungart et al., 2001; Freyman et al., 2004; Newman and Evers, 2007; Yang et al., 2007; Helfer and Freyman, 2009; Huang et al., 2010; Ezzatian et al., 2011), and viewing a speaker's movements of the speech articulators (lipreading) (Sumby and Pollack, 1954; Summerfield, 1979; Rosenblum et al., 1996; Grant and Seitz, 2000; Rudmann et al., 2003; Helfer and Freyman, 2005). It appears that any cue that facilitates listeners' selective attention to target speech can improve recognition of target speech against masking stimuli. However, it is not clear whether people with schizophrenia are able to use perceptual/cognitive cues to unmask target speech.

In the line of studies of the effects of content priming on speech recognition against a masker (Freyman et al., 2004; Yang et al., 2007; Ezzatian et al., 2011), recognition of the last (the third) keyword in a three-keyword nonsense target sentence, which is co-presented with a masker (noise or speech), is improved if the content prime (early part of the target sentence including the first and second keywords) is pre-presented in quiet. Since target sentences used for studying the priming effect are not semantically meaningful (so called nonsense sentences), listeners receive no contextual support from the content prime for guessing the last keyword. Freyman et al. (2004) suggest that the prime helps participants attend more quickly to the target sentence, thereby facilitating recognition of the last keyword against masking. Moreover, working memory, a brain system that temporarily stores and processes information to direct ongoing cognitive processes (Baddeley, 1981, 1992), is critical for the priming effect, because working memory temporally holds the prime information during the target/masker co-presentation. Thus, any masker-induced disruptions of the working memory of prime information may result in impaired recall of the content prime. It is of interest to know whether people with schizophrenia are still able to use a prime to unmask target speech, and whether recalling the prime is affected by noise masking and/or speech masking in both people with schizophrenia and healthy controls.

The present study was to investigate (1) whether speech recognition in people with schizophrenia is more vulnerable to noise masking and/or speech masking than their healthy controls, (2) whether the accuracy of recalling the prime is affected by noise masking and/or speech masking in people with schizophrenia and healthy controls, and (3) whether people with schizophrenia are still able to use a prime to unmask target speech.

It is known that from the illness onset to more chronic stages, people with schizophrenia experience progressive increases in auditory sensory dysfunction (Rabinowicz et al., 2000), deficits of auditory selective attention (Oie et al., 1998), and impairments of working memory (Ohrmann et al., 2007; Pantelis et al., 2009; Zanello et al., 2009; for a review see Forbes et al., 2009). In addition, people with schizophrenia have progressively augmented reduction of regional cerebral blood flow in frontal and posterior lobes (Kanahara et al., 2009) and damage to gray and/or white matter (Okugawa et al., 2002; Federspiel et al., 2006; DeLisi, 2008; Friedman et al., 2008; Gasparotti et al., 2009). Thus, this study also investigated the illness-duration effect by using two patient groups: patients with first-episode schizophrenia and patients with chronic schizophrenia.

2. Materials and methods

2.1. Participants

All the patients used in this study were recruited in the Beijing Anding Hospital. Their diagnoses were obtained based on the Structured Clinical Interview for DSM-IV (SCID-DSM-IV; First et al., 1996). These patients received antipsychotic medications during the test of this study (Table 1). Averagely, the chlorpromazine equivalent was 254 ± 69 mg/day for first episode patients (FEPs) and 611 ± 111 mg/day for chronic patients (CPs). The equivalent was calculated using conversion factors described by Woods (2003). Some potential patient participants were excluded if they had healthy condition(s) affecting experimental tests (including hearing loss, alcohol and/or drug abuse and dependence, and a treatment of the electroconvulsive therapy (ECT) within the past six months, a treatment of trihexyphenidyl hydrochloride with a dose of more than 6 mg/day, and/or an age younger than 18 or older than 59). For the purpose of improving sleeping, some of the patient participants received benzodiazepines based on doctors' advice. All patients were clinically stable at the time of their participation.

Potential healthy controls (FEP-Cs and CP-Cs) were recruited from the communities near the hospital. They were telephone interviewed first and then were screened with the SCID-DSM-IV as used for patient participants. None of the healthy volunteers had a history of Axis I psychiatric disorder as defined by the DSM-IV.

Twelve FEPs with their 12 demographics-matched healthy controls (FEP-Cs), and 22 CPs with their 22 demographics-matched

Table 1
Characteristics of patients and healthy controls.

	First-episode patients (FEPs)	Healthy control for FEPs	Chronic patients (CPs)	Healthy controls for CPs
Individual number	12	12	22	22
Age (years ± SEM)	24.1 (5.8)	24.0 (4.5)	43.3 (9.3)	39.1 (8.8)
Male%	67	67	82	77
Education (years)	12.1 (2.2)	12.6 (2.8)	10.9 (1.7)	11.2 (2.0)
MID (years)	0.75 (0.56)	NA	16.59 (8.30)	NA
PANSS	93.42 (22.98)	NA	92.24 (23.02)	NA
P-scale	23.92 (6.57)	NA	19.94 (6.28)	NA
N-scale	23.58 (8.72)	NA	26.461 (7.12)	NA
G-scale	46.75 (12.26)	NA	46.00 (13.08)	NA
PASAT				
Total-response	41.75 (6.27)a	47.25 (3.36)	31.44 (11.19)b	44.89 (6.22)
Total-correct	37.92 (11.01)	43.92 (4.25)	25.17 (10.39)b	42.61 (6.44)
CRR	89.56 (18.58)	92.87 (4.69)	82.50 (19.92)a	95.08 (8.44)
Digit Span	8.50 (2.71)	9.18 (1.33)	7.89 (1.32)a	9.16 (1.07)
Diagnostic subtype	N		N	
paranoid	7		13	
Non-paranoid	5		9	
Hallucination	N		N	
P3 ≥ 4	6		7	
Antipsychotic medications	N		N	
Typical	0		7	
Atypical*	12		19	
Chlorpromazine equivalent	Mean:254 SD:69 Range: 150–400		Mean:611 SD:111 Range: 350–850	

SEM = sample error of mean. SD = standard deviation. PANSS = positive and negative syndrome scale. P3 = acoustic hallucinition. MID = mean illness duration. NA = not applicable. PASAT = paced auditory serial addition test. CRR = (percent correct/percent response) × 100.
a. Schizophrenia patients different from control group mean, $p < 0.05$.
b. Schizophrenia patients different from control group mean, $p < 0.01$.
* Note that 7 patients received 2 different antipsychotic medications.

healthy controls (CP-Cs) participated in the study. Table 1 shows the main characteristics of each of the four participant groups. All the participants were right-handed and did not show any pure-tone hearing impairments for each ear at the frequencies of 125, 256, 512, 1024, and 2048 Hz. They gave their written informed consent to participate in this study. The procedures of this study were approved by the Independent Ethics Committee (IEC) of the Beijing Anding Hospital, Capital Medical University.

2.2. Equipment and materials

The participant was seated at the center of a quite room in the hospital. Acoustic signals, calibrated by a sound-level meter (AUDit and System 824, Larson Davis, USA), were delivered from a notebook-computer sound card (ATI SB450 AC97) to earphones (Model HDA 200) and bilaterally presented to the participant at the sound pressure level (SPL) of 60 dBA without any binaural differences. The SPLs of the masker were adjusted to produce four signal-to-noise ratios (SNRs): −8, −4, 0 and 4 dB.

Speech stimuli were Chinese nonsense sentences, which are syntactically correct but not semantically meaningful. Direct English translations of the sentences are similar but not identical to the English nonsense sentences developed by Helfer (1997). For example, the English translation of a Chinese nonsense sentence is "One <u>appreciation</u> could <u>retire</u> his <u>ocean</u>" (keywords are underlined). Each of the Chinese sentences has 12 syllables (also 12 characters) including three keywords with two syllables for each. The sentence frame cannot provide any contextual support for recognizing the keywords. The development of the Chinese nonsense sentences is described by Yang et al. (2007).

Both target speech and priming stimuli were spoken by three young-female talkers (Talkers A, B and C). In a trial, the priming stimulus was identical to the target sentence except that the last keyword was replaced by a white noise burst, whose duration was equal to that of the longest of the last (third) keywords in all the target sentences, and whose level was 10 dB lower than that of preceding part of the sentence (Freyman et al., 2004). The voice reciting the priming stimulus was identical to that reciting the target sentence for the trial.

The noise masker was a stream of steady-state speech-spectrum noise (Yang et al., 2007). 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 D and E) (Yang et al., 2007).

2.3. Procedures

The locally validated version of the Positive and Negative Syndrome Scale (PANSS) tests (Si et al., 2004), Paced Auditory Serial Addition Test (PASAT) (Shi et al., 2005; Tombaugh, 2006), and Digital Span were conducted one day before the test for patient participants. The order of the neuropsychological tests was randomized. Group-mean scores for the positive, negative, and general psychopathological scales did not significantly differ between the first-episode-patient group and the chronic-patient group.

For the speech-recognition test, there were three within-subject variables: (1) masker type (noise, speech), (2) priming condition (no-priming, priming), and (3) SNR (−8, −4, 0 and 4 dB). For each participant, there were 16 test conditions in total and 12 trials (also 12 target sentences) were used in each condition. Under a condition, each of the three target (and priming) voices had the equal chance to be used, and the presentation order for the voices was arranged in a random manner. The presentation order for the four masker-type/priming-condition combinations was partially counterbalanced across participants in each group using a Latin square order, and the presentation order of the four SNRs was arranged randomly for each masker-type/priming-condition combination.

In a trial with the presentation of the prime (under the priming condition), the prime was presented in quiet following the participant pressed the "Enter" key on a computer keyboard. Immediately after the priming phase, the masker was presented and about 1 s later the target sentence was started. Then the target sentence terminated with the masker. In a trial without the presentation of the prime (under the no-priming condition), either the noise masker or the speech masker was presented after the key press, and then the target sentence was started about 1 s after the masker onset. 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.

2.4. Statistical analyses

Analysis of variance (ANOVA) followed by Bonferroni post hoc tests (when necessary) was performed using SPSS 13.0 software. The null hypothesis was rejected at the level of 0.05.

To investigate whether speech recognition was different between patients and their demographics-matched healthy controls, recognition of all the three keywords in target speech under the no-priming condition was first examined by comparing the group-mean thresholds of the psychometrical function. One of the advantages to test the threshold was to remove the SNR factor from the statistics. However, to examine whether recognition of the prime keywords was affected by the noise or speech masker, the SNR was the critical factor. Finally, to study the priming effect on recognition of the last (third) keyword, thresholds of the psychometric function were compared between the no-priming condition and the priming condition, leading to that the SNR factor was removed from the statistics.

3. Results

3.1. Recognition of the three keywords under the no-priming condition

To investigate whether speech recognition was different between patients and their demographics-matched healthy controls, recognition of all the three keywords in target speech under the no-priming condition was first examined.

A logistic psychometric function,

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

was fit to each individual participant's data, using the Levenberg–Marquardt method (Wolfram, 1991), where y is the probability of correct recognition of the keywords, x is the SNR corresponding to y, μ (the threshold) is the SNR corresponding to 50% correct on the psychometric function, and σ determines the slope of the psychometric function.

Fig. 1 shows group-mean percent-correct recognition of the three keywords in target sentences as a function of SNR along with the group-mean best-fitting psychometric functions in both the two patient (FEP and CP) groups and the two demographics-matched healthy control (FEP-C and CP-C) groups under the no-priming condition, when the masker was either noise (top panels) or speech (bottom panels). Fig. 2 shows comparisons in the group-mean threshold μ for recognizing the three keywords across participant groups and masking conditions. Obviously, both Fig. 1 and Fig. 2 show that under the no-priming condition, recognition of the keywords was worse (higher in the threshold μ) in patients with schizophrenia than in healthy controls, and it was worse under the speech-masking condition than under the noise-masking condition for all the four participant groups.

The difference in threshold μ ($\Delta\mu$) between the patient group and the matched healthy-control group at each of the masking conditions

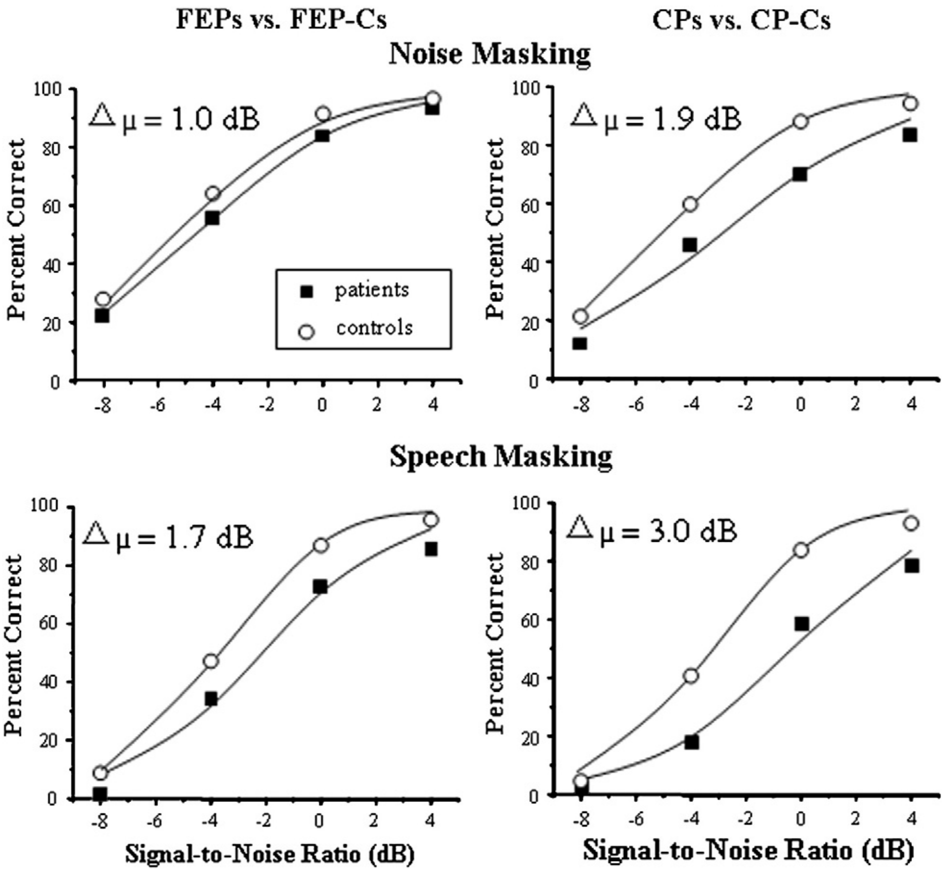

Fig. 1. Group-mean percent-correct recognition of the three keywords in target sentences as a function of SNR along with the group-mean best-fitting psychometric functions (curves) for the patient groups (solid squares) and the demographics-matched-control groups (open circles) under the no-priming conditions, when the masker was noise (top panels) or speech (bottom panels). The difference in threshold μ (Δμ) between the patient group and the matched healthy-control group at each of the masking conditions was also indicated in each panel. FEPs, first-episode patients; FEP-Cs, healthy controls for first-episode patients; CPs, chronic patients; CP-Cs, healthy controls for chronic patients.

was also indicated in Fig. 1. Generally, the difference between CPs and CP-Cs was larger than that between FEPs and FEP-Cs; the difference between patients and matched healthy controls was larger under the speech-masking condition than under the noise-masking condition.

A 4 (group: FEP, FEP-C, CP, CP-C) by 2 (masker type: noise, speech) two-way ANOVA shows that the main effect of group on the threshold μ was significant ($F_{3,128}=26.366$, $p<0.01$), the main effect of masker type was significant ($F_{1,128}=46.060$, $p<0.01$), but the interaction between two factors was not significant ($F_{3,128}=0.549$, $p=0.650$). Post Hoc tests for the group effect show that the threshold difference between the FEP group and the FEP-C group was significant ($p<0.01$), the difference between the CP group and the CP-C group was significant ($p<0.01$), the difference between the two patient groups (FEP and CP) was significant ($p<0.01$), but the difference between the two healthy control groups (FEP-C and CP-C) was not significant ($p=0.084$).

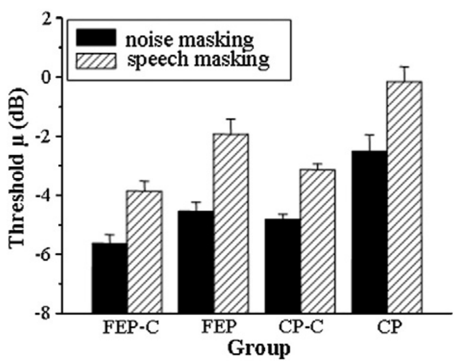

Fig. 2. Group-mean thresholds (μ) for correct recognition of the three keywords in target sentences when the masker was noise or speech for each of the four participant groups under the no-priming condition. Error bars indicate the standard errors of the mean. FEP, first-episode patient; FEP-C, first-episode-patient healthy control; CP, chronic patient; CP-C, chronic control.

3.2. Recognition of the prime keywords under the priming condition

Fig. 3 shows group-mean percent-correct recognition of the prime keywords (first and second keywords) as a function of SNR along with the group-mean best-fitting psychometric function in each of the four participant groups under the priming condition, when the masker was either noise (top panels) or speech (bottom panels). As predicted, the correct percent of recalling the prime keywords was high even at low SNRs, because the prime keywords were presented in quiet to participants before the co-presentation of the (whole-course) target speech and the masker, suggesting that the working memory of prime-content information was sufficiently maintained. However, in both healthy and patient participants, the prime-recalling performance was not perfect but a function of the SNR, especially under the speech-masking condition.

To examine whether FEPs and FEP-Cs were different in recalling the prime keywords and whether the SNR effect was significant under the priming condition when the masker was either noise or speech, a 2 (group) by 2 (masker) by 4 (SNR) three-way ANOVA was conducted, which showed that the interaction between group and masker type was significant ($F_{1,\ 176}=7.875$, $p=0.006$). However, the interactions between SNR and masker type, between SNR and group, and between the three factors were not significant ($p>0.800$). Separate 2 (group) by 4 (SNR) two-way ANOVAs show that when the masker was noise,

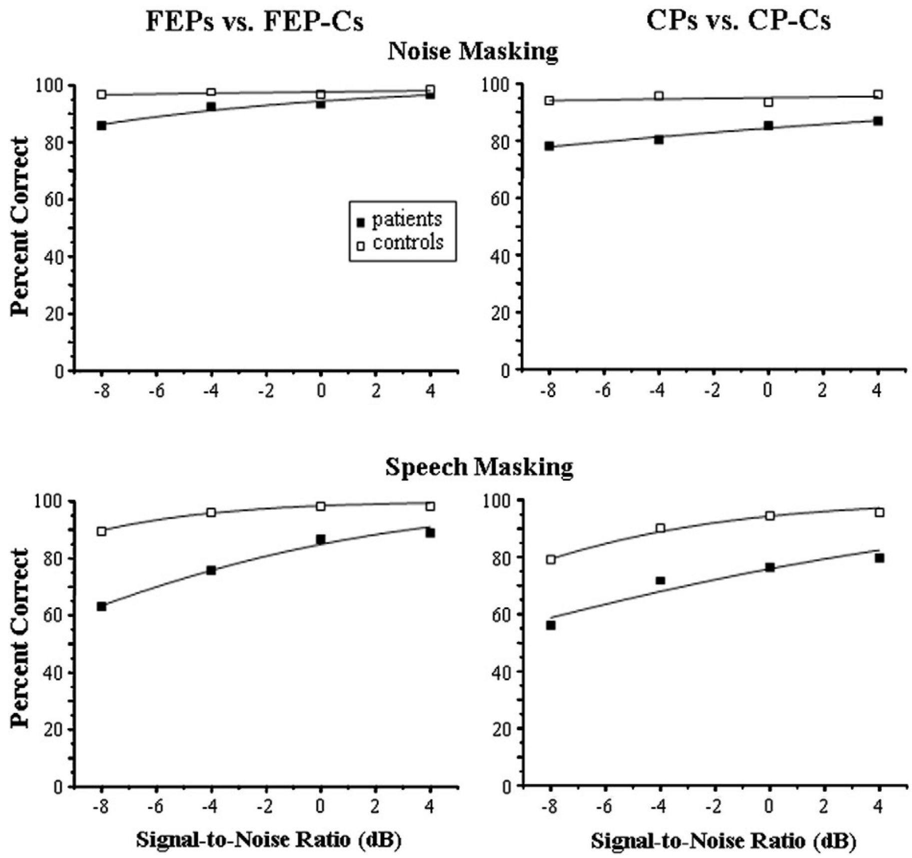

Fig. 3. Group-mean percent-correct recognition of the prime keywords (the first and second keywords) in target sentences as a function of SNR along with the group-mean best-fitting psychometric function (curve) in the patient group (filled squares) and demographics-matched healthy control group (open squares) under the priming condition, when the masker was either noise (upper panels) or speech (lower panels). For abbreviations see Fig. 2.

the main effect of SNR was not significant ($F_{3,88}=1.961$, $p=0.126$), the main effect of group was significant ($F_{1,88}=7.602$, $p=0.001$), and the interaction between SNR and group was not significant ($F_{3,88}=1.027$, $p=0.385$). When the masker was speech, the main effect of SNR was significant ($F_{3,88}=4.590$, $p=0.005$), and main effect of group was significant ($F_{3,88}=20.854$, $p<0.001$), but the interaction between SNR and group was not significant ($F_{3,88}=1.143$, $p=0.336$). The results suggest that (1) recall of the prime keywords in FEPs was worse than that in FEP-Cs, and (2) the speech masker, but not the noise masker, significantly affected recall of prime keywords in both FEPs and FEP-Cs. The reduced recall of the prime keywords in FEPs under the speech-masking condition was consistent with the reduced score of "total-response" in the PASAT in FEPs (Table 1).

To examine whether CPs and CP-Cs were different in recognizing the prime keywords and whether the SNR effect was significant, a 2 (group) by 2 (masker) by 4 (SNR) three-way ANOVA shows that the interaction between SNR and masker was significant ($F_{3,\ 335}=4.982$, $p=0.002$) and the interaction between group and masker was significant ($F_{1,\ 335}=4.878$, $p=0.023$). However, the interaction between SNR and group and three-way interaction were not significant ($p>0.3$). Separate 2 (group) by 4 (SNR) two-way ANOVAs show that when the masker was noise, the main effect of SNR was not significant ($F_{3,\ 168}=1.743$, $p=0.160$), the main effect of group was significant ($F_{1,\ 168}=47.891$, $p=0.001$), and the interaction between the two factors was not significant ($F_{3,\ 168}=1.219$, $p=0.304$). When the masker was speech, the main effect of SNR was significant ($F_{3,\ 168}=12.617$, $p<0.001$), the main effect of group was significant ($F_{3,\ 168}=56.463$, $p<0.001$), but the interaction between SNR and group was not significant ($F_{3,\ 168}=0.409$, $p=0.746$). The results suggest that (1) recall of prime keywords in CPs was worse than that in CP-Cs under each of the two masking conditions, (2) the speech masker, but not the noise masker, affected recall of prime keywords in both CPs and CP-Cs. Fig. 3 also indicates that under speech-masking conditions, the performance difference between CPs and CP-Cs was larger than that between FEPs and FEP-Cs. In CPs, the reduced recall of the prime keywords under the speech-masking condition was consistent with the reduction in both the score of "total-response" and the scope of "total-correct" in the PASAT, and consistent with the reduction in the scores of CRR and digital span (Table 1).

3.3. Priming effect on recognition of the last (third) keyword

Figs. 4 and 5 show group-mean percent-correct recognition of the last (third) keyword as a function of SNR along with the group-mean best-fitting psychometric function in the two younger-participant (FEP and FEP-C) groups and the two older-participant (CP and CP-C) groups, respectively, under either the no-priming condition or the priming condition, when the masker was either noise (top panels) or speech (bottom panels). Fig. 6 shows comparisons between group-mean thresholds for recognizing the last keywords across the four participant groups and the two masking conditions. Obviously, presenting the prime markedly improved recognition of the last keyword in each of the participant groups when the masker was either noise or speech.

A 2 (priming condition) by 4 (group) by 2 (masker type) three-way ANOVA shows that the main effect of priming condition on the threshold μ for recognizing the last keyword was significant($F_{3,\ 256}=39.564$, $p<0.001$), the main effect of group was significant ($F_{1,\ 256}=59.588$, $p<0.001$), and the main effect of masker was significant ($F_{1,\ 256}=20.337$, $p<0.001$). However, all the two-way interactions and the three-way interaction were not significant (for all, $p>0.100$). Post Hoc tests for the group effect show that the difference between

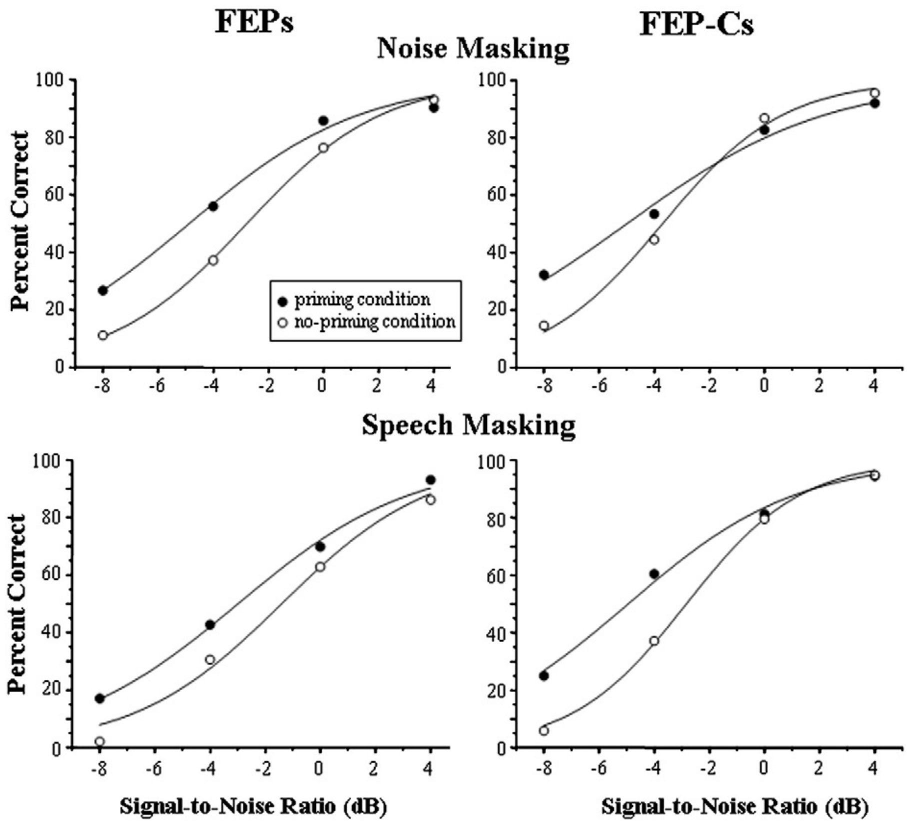

Fig. 4. Group-mean percent-correct recognition of the last keyword in target sentences as a function of SNR along with the group-mean best-fitting psychometric function (curve) in the first episode patient group and demographics-matched healthy control group under the no-priming (open circles) and priming conditions (solid circles), when the masker was either noise (upper panels) or speech (lower panels). For abbreviations see Fig. 2.

FEPs and FEP-Cs was significant ($p = 0.003$), the difference between CPs and CP-Cs was significant ($p < 0.001$), the difference between the two patient groups was significant ($p < 0.001$), but the difference between the two healthy-control groups was not significant ($p = 0.235$).

4. Discussion

4.1. Recognition of target-speech keywords under the no-priming condition

First of all, this study examined whether recognition of the three target-speech keywords in patients with schizophrenia is more vulnerable to noise masking and/or speech masking, compared to that in their demographics-matched healthy controls. The results show that under the conditions without presenting the prime, recognition of the keywords was significantly worse in patients than that in healthy controls. The results are in agreement with previous studies reporting that people with schizophrenia are more susceptible to auditory masking (Green et al., 1999; Källstrand et al., 2002; Lee et al., 2004) and experience more difficulties in speech perception (Bull and Venables, 1974; DeLisi, 2001; Condray et al., 2002; Kasai et al., 2002a,b; Lee et al., 2004; Titone and Levy, 2004; Condray, 2005).

Moreover, although the number of participants in each of the two younger-participant groups (FEP, FEP-C) was smaller than that in each of the two older-participant groups (CP, CP-C), the difference in recognizing the keywords between patients and their healthy controls was apparently both masker-type dependent and illness-duration dependent. Specifically, when the masker was noise, the difference in the threshold μ for recognizing the three keywords between FEPs and FEP-Cs was 1.0 dB, and that between CPs and CP-Cs was 1.9 dB. When the masker was speech, the difference in the μ between FEPs and FEP-Cs was 1.7 dB, and that between CPs and CP-Cs was 3.0 dB. The small difference in μ between patients and matched-healthy controls under the noise-masking condition in this study may be partially in agreement with the Ross et al. report (2007) showing that compared to healthy participants (mean age = 35 years), schizophrenic participants (mean age = 39 years) did not have any deficits in recognizing monosyllabic words under noise-masking conditions.

As mentioned in the Introduction, in addition to energetic masking, speech maskers induce informational masking of target speech (Freyman et al., 1999; Brungart et al., 2001; Li et al., 2004; Helfer and Freyman, 2005, 2009; Wu et al., 2005; Rakerd et al., 2006; Huang et al., 2008, 2009). The increased difference in recognizing target keywords between patients and their healthy controls under the speech-masking condition suggests that speech recognition in patients with schizophrenia is more vulnerable to informational masking than energetic masking, because speech-on-speech masking involves additional interfering influence at higher perceptual/linguistic processing levels than masking of speech by noise (for a review see Schneider et al., 2007).

Also, the results suggest that compared to FEPs, CPs performed more poorly in recognizing target keywords under each of the two masking conditions. Since there were no significant differences between the two healthy control groups, the illness-duration difference for patient participants, but not the age difference, might play a role in affecting target-speech recognition against noise masking and speech masking. A possible explanation of the difference between the two patient groups is that due to progressively augmented impairment in auditory selective attention (Oie et al., 1998) and damage to gray and/or white matter (Okugawa et al., 2002; Federspiel et al., 2006; DeLisi, 2008; Friedman et al., 2008; Gasparotti et al., 2009), recognition of speech is more vulnerable to masking, especially informational masking, in chronic patients than in first-episode patients. Note that since the participant

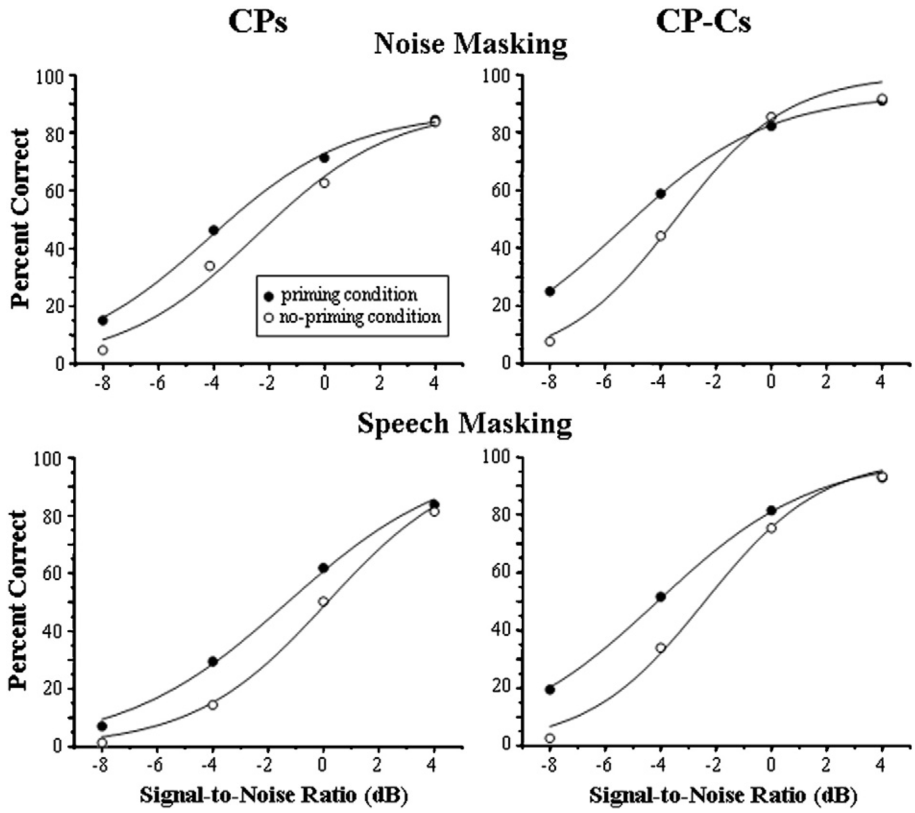

Fig. 5. Group-mean percent-correct recognition of the last keyword in target sentences as a function of SNR along with the group-mean best-fitting psychometric function (curve) in the chronic patient group and demographics-matched healthy control group under the no-priming (open circles) and priming conditions (solid circles), when the masker was either noise (upper panels) or speech (lower panels). For abbreviations see Fig. 2.

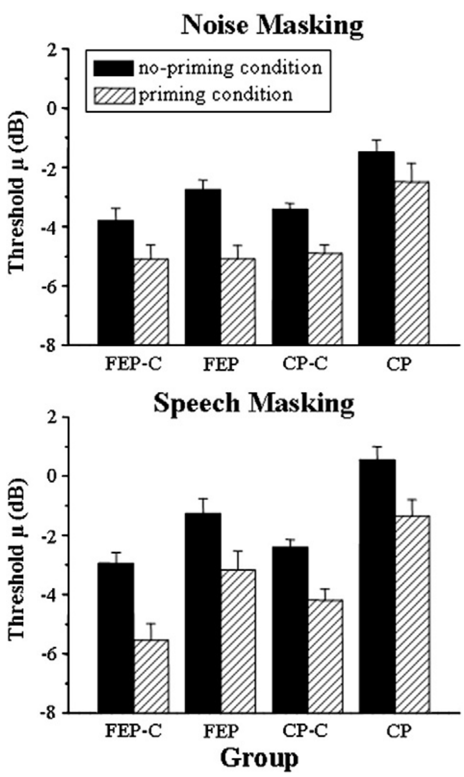

Fig. 6. Comparisons of group-mean thresholds (μ) for correct recognition of the last keyword in target sentences across participants groups and masking condition under either the no-priming condition or the priming condition. Error bars indicate the standard errors of the mean. For abbreviations see Fig. 2.

number, gender, education, severity of PANSS scores, and antipsychotic medications in the FEP group did not match those in the CP group, comparisons in masking of target speech between FEPs and CPs still need further investigation in the future.

4.2. Working memory of the prime-content information under the priming condition

In this study, as indicated in Table 1, both the score of "total-response" and the scope of "total-correct" in the PASAT were lower in patients than in healthy controls, suggesting that patients had impaired short-term working memory (Tombaugh, 2006). Particularly, CPs also showed markedly lower scores in CRR and digital span than their healthy controls, indicating that further impairment in working memory occurred in CPs. Indeed, under the priming condition, recall of the prime keywords was not perfect in each of the four groups of participants who had normal hearing, even though the prime keywords were presented in quiet to the participants before the co-presentation of the target speech and the masker. Moreover, recall of the prime keywords in patients, especially CPs, was significantly worse than that in their healthy controls. Thus, the results suggest that compared to that in healthy controls, working memory of prime keywords in patients of schizophrenia is more vulnerable to masking, and the vulnerability may reflect working memory deficiencies (but not auditory deficiencies). The results are consistent with previous reports that patients with schizophrenia experience progressively augmented impairments in working memory (Ohrmann et al., 2007; Pantelis et al., 2009; Zanello et al., 2009; for a review see Forbes et al., 2009).

More importantly, in each of the four participant groups, recall of the prime keywords was a significant function of the SNR when the masker was speech but not noise. Also, separate statistical tests for the younger-participant groups (FEP and FEP-C) and older-participant groups (CP and CP-C) confirmed that the difference in recalling the prime keywords between patients and their healthy controls was larger when the masker was speech than when the masker was noise. The results suggest that working memory for holding the prime-content information is vulnerable to informational masking but not energetic masking, and the vulnerability is larger in patients, particularly CPs, than in healthy controls.

4.3. Using the prime to unmask the last keyword in target speech

In the present study, recognition of the last keyword was worse in patients than that in the matched healthy controls. Also, although there was no significant difference between the two healthy-control groups and the performance was worse in CPs than in FEPs, further investigation is still needed to clarify the illness-duration effect. However, for both patient groups and healthy-control groups, recognition of the last keyword was significantly improved by presenting the prime in quiet.

As mentioned in the Introduction, knowledge/familiarity of early part of the content of target speech and/or the target-talker's voice improve recognition of the last keyword in target speech when the masker is either noise or speech (Freyman et al., 2004; Yang et al., 2007; Huang et al., 2010; Ezzatian et al., 2011), because the prime helps listener attend to the target stream. Thus, although patients with schizophrenia exhibit many deficits related to speech recognition, they are still able to use the prime to facilitate their selective attention to the target stream among the target/masker complex and unmask target speech under either noise-masking or speech-masking conditions.

5. Conclusions

This study for the first time provides evidence to show that compared to that in their demographics-matched healthy listeners, recognition of speech in both FEPs and CPs is more vulnerable to masking, especially informational masking. Also, the vulnerability in CPs is larger than that in FEPs, showing the sectional effect of the illness. Moreover, similar to healthy controls, both FEPs and CPs are able to use the content/voice prime to release target speech from either noise masking or speech masking, even though the working memory in patients with schizophrenia is more vulnerable to informational masking than that in healthy controls.

Role of funding source

Funding for this study was provided by the “973” National Basic Research Program of China (2011CB707805), the National Scientific and Technical Supporting Programs of China (2007BAI17B04), the Chinese Ministry of Education (20090001110050), and “985” grants from Peking University. They had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Contributors

Each of the authors, Chao Wu, Shuyang Cao, Fuchun Zhou, Chuanyue Wang, Xihong Wu, and Liang Li designed the study and wrote the protocol. The experiment program was written by Shuyang Cao. Experiments were conducted by Chao Wu and Fuchun Zhou. Chao Wu wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interests

Each of the authors has no conflict of interests to disclose.

Acknowledgments

This work was supported by the “973” National Basic Research Program of China (2011CB707805), the National Scientific and Technical Supporting Programs of China (2007BAI17B04), the Chinese Ministry of Education (20090001110050), and “985” grants from Peking University.

References

Alain, C., Bernstein, L.J., Cortese, F., Yu, H., Zipursky, R.B., 2002. Deficits in automatically detecting changes in conjunction of auditory features in patients with schizophrenia. Psychophysiology 39, 599–606.

Arbogast, T.L., Mason, C.R., Kidd, G.J.R., 2002. The effect of spatial separation on informational and energetic masking of speech. J. Acoust. Soc. Am. 112, 2086–2098.

Baddeley, A., 1981. The concept of working memory — a view of its current state and probable future-development. Cognition 10, 17–23.

Baddeley, A., 1992. Working memory. Science 255, 556–559.

Best, V., Ozmeral, E.J., Shinn-Cunningham, B.G., 2007. Visually-guided attention enhances target identification in a complex auditory scene. J. Assoc. Res. Otolaryngol. 8, 294–304.

Brenner, C.A., Krishnan, G.P., Vohs, J.L., Ahn, W.Y., Hetrick, W.P., Morzorati, S.L., et al., 2009. Steady state responses: electrophysiological assessment of sensory function in schizophrenia. Schizophr. Bull. 356, 1065–1077.

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 110, 2527–2538.

Bull, H.C., Venables, P.H., 1974. Speech perception in schizophrenia. Brit. J. Psychiatr. 125, 350–354.

Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979.

Condray, R., 2005. Language disorder in schizophrenia as a developmental learning disorder. Schizophr. Res. 73, 5–20.

Condray, R., Steinhauer, S.R., Kammen, D.P., Kasparek, A., 2002. The language system in schizophrenia: effects of capacity and linguistic structure. Schizophr. Bull. 28, 475–490.

DeLisi, L.E., 2001. Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. Schizophr. Bull. 27, 481–496.

DeLisi, L.E., 2008. The concept of progressive brain change in schizophrenia: implications for understanding schizophrenia. Schizophr. Bull. 34, 312–321.

Ezzatian, P., Li, L., Pichora-Fuller, K., Schneider, B., 2011. The effect of priming on release from informational masking is equivalent for younger and older adults. Ear. Hearing. 32, 84–96.

Federspiel, A., Begre, S., Kiefer, C., Schroth, G., Strik, W.K., Dierks, T., 2006. Alterations of white matter connectivity in first episode schizophrenia. Neurobiol. Dis. 22, 702–709.

First, M., Spitzer, R., Gibbon, M., Williams, J., 1996. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID), Clinician Version. American Psychiatric Press, Washington, DC.

Forbes, N.F., Carrick, L.A., McIntosh, A.M., Lawrie, S.M., 2009. Working memory in schizophrenia: a meta-analysis. Psychol. Med. 39, 889–905.

Freyman, R.L., Helfer, K.S., McCall, D.D., Clifton, R.K., 1999. The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. 10, 3578–3588.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2004. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J. Acoust. Soc. Am. 115, 2246–2256.

Friedman, J.I., Tang, C., Carpenter, D., Buchsbaum, M., Schmeidler, J., Flanagan, L., et al., 2008. Diffusion tensor imaging findings in first-episode and chronic schizophrenia patients. Am. J. Psychiatry 165, 1024–1032.

Gasparotti, R., Valsecchi, P., Carletti, F., Galluzzo, A., Liserre, R., Cesana, B., et al., 2009. Reduced fractional anisotropy of corpus callosum in first-contact, antipsychotic drug-naive patients with schizophrenia. Schizophr. Res. 108, 41–48.

Grant, K.W., Seitz, P.F., 2000. The use of visible speech cues for improving auditory detection of spoken sentences. J. Acoust. Soc. Am. 108, 1197–1208.

Green, M.F., Nuechterlein, K.H., Breitmeyer, B., Mintz, J., 1999. Backward masking in unmedicated schizophrenic patients in psychotic remission: possible reflection of aberrant cortical oscillation. Am. J. Psychiatry 156, 1367–1373.

Helfer, K.S., 1997. Auditory and auditory–visual perception of clear and conversational speech. J. Speech. Lang. Hear. R. 40, 432–443.

Helfer, K.S., Freyman, R.L., 2005. The role of visual speech cues in reducing energetic and informational masking. J. Acoust. Soc. Am. 117, 842–849.

Helfer, K.S., Freyman, R.L., 2009. Lexical and indexical cues in masking by competing speech. J. Acoust. Soc. Am. 125, 447–456.

Hoffman, R.E., Rapaport, J., Mazure, C.M., Quinlan, D.M., 1999. Selective speech perception alterations in schizophrenic patients reporting hallucinated “voices”. Am. J. Psychiatry 156, 393–399.

Huang, Y., Huang, Q., Chen, X., Qu, T.S., Wu, X.H., 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. Res. 244, 51–65.

Huang, Y., Huang, Q., Chen, X., Wu, X.H., Li, L., 2009. Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. J. Exp. Psychol. Hum. Percept. Perform. 35, 1618–1628.

Huang, Y., Xu, L.J., Wu, X.H., Li, L., 2010. The effect of voice cuing on releasing speech from informational masking disappears in older adults. Ear. Hearing. 31, 579–583.

Javitt, D.C., Strous, R.D., Grochowski, S., Ritter, W., Cowan, N., 1997. Impaired precision, but normal retention, of auditory sensory (“echoic”) memory information in schizophrenia. J. Abnorm. Psychol. 106, 315–324.

Javitt, D.C., Shelley, A.M., Silipo, G., Lieberman, J.A., 2000. Deficits in auditory and visual context-dependent processing in schizophrenia — defining the pattern. Arch. Gen. Psychiatry 57, 1131–1137.

Källstrand, J., Montnémery, P., Nielzén, S., Olsson, O., 2002. Auditory masking experiments in schizophrenia. Psychiatr. Res. 113, 115–125.

Kanahara, N., Shimizu, E., Sekine, Y., et al., 2009. Does hypofrontality expand to global brain area in progression of schizophrenia?: a cross-sectional study between first-episode and chronic schizophrenia. Biol. Psychiatry 33, 410–415.

Kasai, K., Nakagome, K., Itoh, K., Koshida, I., Hata, A., Iwanami, A., et al., 2002a. Impaired cortical network for preattentive detection of change in speech sounds in schizophrenia: a high-resolution event-related potential study. Am. J. Psychiatry 159, 546–555.
Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, K., Fukuda, M., et al., 2002b. Neuromagnetic correlates of impaired automatic categorical perception of speech sounds in schizophrenia. Schizophr. Res. 59, 159–172.
Kidd, G., Arbogast, T.L., Mason, C.R., Gallun, F.J., 2005. The advantage of knowing where to listen. J. Acoust. Soc. Am. 118, 3804–3815.
Lee, S.H., Chung, Y.C., Yang, J.C., Kim, Y.K., Suh, K.Y., 2004. Abnormal speech perception in schizophrenia with auditory hallucinations. Acta Neuropsychiatrica. 163, 154–159.
Li, L., Daneman, M., Qi, J.G., Schneider, B.A., 2004. Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? J. Exp. Psychol. Hum. Percept. Perform. 30, 1077–1091.
Michie, P.T., 2001. What has MMN revealed about the auditory system in schizophrenia? Int. J. Psychophysiol. 42, 177–194.
Newman, R.S., Evers, S., 2007. The effect of talker familiarity on stream segregation. J. Phon. 35, 85–103.
Ohrmann, P., Siegmund, A., Suslow, T., et al., 2007. Cognitive impairment and in vivo metabolites in first-episode neuroleptic-naive and chronic medicated schizophrenic patients: a proton magnetic resonance spectroscopy study. J. Psychiatr. Res. 41, 625–634.
Oie, M., Rund, B.R., Sundet, K., Bryhn, G., 1998. Auditory laterality and selective attention: normal performance in patients with early-onset schizophrenia. Schizophr. Bull. 24, 643–652.
Okugawa, G., Sedvall, G.C., Agartz, I., 2002. Reduced grey and white matter volumes in the temporal lobe of male patients with chronic schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. 252, 120–123.
Pantelis, C., Wood, S.J., Proffitt, T.M., et al., 2009. Attentional set-shifting ability in first-episode and established schizophrenia: relationship to working memory. Schizophr. Res. 112, 104–113.
Rabinowicz, E.F., Silipo, G., Goldman, R., et al., 2000. Auditory sensory dysfunction in schizophrenia—imprecision or distractibility? Arch. Gen. Psychiatry 57, 1149–1155.
Rakerd, B., Aaronson, N.L., Hartmann, W.M., 2006. Release from speech-on-speech masking by adding a delayed masker at a different location. J. Acoust. Soc. Am. 119, 1597–1605.
Rosenblum, L.D., Johnson, J.A., Saldana, H.M., 1996. Point-light facial displays enhance comprehension of speech in noise. J. Speech. Hear. Res. 39, 1159–1170.
Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C., Foxe, J.J., 2007. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophr. Res. 97, 173–183.
Rudmann, D.S., McCarley, J.S., Kramer, A.F., 2003. Bimodal displays improve speech comprehension in environments with multiple speakers. Hum. Factors 45, 329–336.
Schneider, B.A., Li, L., Daneman, M., 2007. How competing speech interferes with speech comprehension in everyday listening situations. J. Am. Acad. Audiol. 18, 559–572.
Shi, C., Yu, X., Wu, Z.Y., Heaton, R.K., Jin, H., Marcotte, T.D., et al., 2005. Neuropsychological feasibility study among HIV+/AIDS subjects in China. Chin. Ment. Health J. 19, 343–346.
Si, T.M., Yang, J.Z., Shu, L., Wang, X.L., Kong, Q.M., Zhou, M., et al., 2004. The reliability, validity of PANSS (Chinese version), and its Implication. Chin. Ment. Health J. 18, 45–47.
Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 26, 212–215.
Summerfield, A.Q., 1979. Use of visual information for phonetic processing. Phonetica 36, 314–331.
Titone, D., Levy, D.L., 2004. Lexical competition and spoken word identification in schizophrenia. Schizophr. Res. 68, 75–85.
Tombaugh, T.N., 2006. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). Arch. Clin. Neuropsychol. 21, 53–76.
Wolfram, S., 1991. Mathematica: A System for Doing Mathematics by Computer. Addison-Welsey, New York.
Woods, S.W., 2003. Chlorpromazine equivalent doses for the newer atypical antipsychotics. J. Clin. Psychiatry 64, 663–667.
Wu, X.H., Wang, C., Chen, J., Qu, H.W., Li, W.R., Wu, Y.H., et al., 2005. The effect of perceived spatial separation on informational masking of Chinese speech. Hearing. Res. 199, 1–10.
Yang, Z.G., Chen, J., Wu, X.H., Wu, Y.H., Schneider, B.A., Li, L., 2007. The effect of voice cuing on releasing Chinese speech from informational masking. Speech Commun. 49, 892–904.
Zanello, A., Curtis, L., Ba, M.B., et al., 2009. Working memory impairments in first-episode psychosis and chronic schizophrenia. Psychiatr. Res. 165, 10–18.