Brain & Language 135 (2014) 85–95

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Attentional modulation of the early cortical representation of speech signals in informational or energetic masking

Changxin Zhang, Lingxi Lu, Xihong Wu, Liang Li *

Department of Psychology, Speech and Hearing Research Center, McGovern Institute for Brain Research at PKU, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Accepted 5 June 2014
Available online 1 July 2014

Keywords:
Speech encoding
Speech recognition
Attention
Informational masking
Energetic masking
Event-related potentials
Perceptual separation
Precedence effect
Active listening
Passive listening

ABSTRACT

It is easier to recognize a masked speech when the speech and its masker are perceived as spatially segregated. Using event-related potentials, this study examined how the early cortical representation of speech is affected by different masker types and perceptual locations, when the listener is either passively or actively listening to the target speech syllable. The results showed that the two-talker-speech masker induced a much larger masking effect on the N1/P2 complex than either the steady-state-noise masker or the amplitude-modulated speech-spectrum-noise masker did. Also, a switch from the passive- to active-listening condition enhanced the N1/P2 complex only when the masker was speech. Moreover, under the active-listening condition, perceived separation between target and masker enhanced the N1/P2 complex only when the masker was speech. Thus, when a masker is present, the effect of selective attention to the target-speech signal on the early cortical representation of the speech signal is masker-type dependent.

1. Introduction

1.1. Energetic masking and informational masking of speech

Under noisy listening conditions (e.g., a cocktail-party environment; Cherry, 1953), listeners usually find it difficult to comprehend target speech and participate in conversations due to auditory masking (Miller, 1947). The mechanism underlying auditory masking is complicated and particularly influenced by the masker type. Any masker can simultaneously produce two categories of masking effects: *energetic masking* and *informational masking* (e.g., Arbogast, Mason, & Kidd, 2002; Brungart, 2001; Brungart & Simpson, 2002; Durlach et al., 2003; Ezzatian, Li, Pichora-Fuller, & Schneider, 2011; Freyman, Balakrishnan, & Helfer, 2001; Freyman, Helfer, McCall, & Clifton, 1999; Kidd, Mason, Deliwala, Woods, & Colburn, 1994; Kidd, Mason, Rohtla, & Deliwala, 1998; Li, Daneman, Qi, & Schneider, 2004; Wu et al., 2005; for a review see Schneider, Li, & Daneman, 2007). Energetic masking mainly occurs in the cochlea when the signal sound wave physically interacts with the masker sound wave in the same auditory filter, leading to a substantially degraded or noisy representation of the signal at the peripheral processing level. The effectiveness of energetic masking cannot be modulated by higher-level cognitive and attentional processes. Wideband noises with or without amplitude modulations have been generally used as maskers that mainly produce energetic masking of speech.

On the other hand, competing sound sources can also cause informational masking that interferes with the processing of the signal in addition to energetic masking. For example, although a speech masker induces energetic masking (due to the speech masker-elicited activities in the same or nearby regions on the basilar membrane that are processing the target speech at the same time), processing of the information in the speech masker interferes with processing of the target speech at both perceptual (e.g., phonemic identification) and cognitive (e.g., semantic processing) levels, making selective attention and segregation of target speech from masking speech difficult for listeners. Thus, when the spectrum of the speech masker overlaps with that of the target speech, a speech masker can produce both energetic and information masking of the target speech.

1.2. Perceptual/cognitive cues used for releasing target speech from masking

Listeners are able to use various perceptual/cognitive cues to release target speech from irrelevant-speech-induced informational masking. The cues include perceptual familiarity with the

* Corresponding author.
E-mail address: liangli@pku.edu.cn (L. Li).

http://dx.doi.org/10.1016/j.bandl.2014.06.002
0093-934X/

talker's voice (Brungart, 2001; Huang, Xu, Wu, & Li, 2010; Yang et al., 2007), prior knowledge about part of the target-sentence content (i.e., temporally pre-presented content prime, Freyman, Balakrishnan, & Helfer, 2004; Wu, Li, Gao, et al., 2012; Wu, Li, Hong, et al., 2012; Wu, Cao, et al., 2012; Wu, Li, et al., 2013; Yang et al., 2007), and viewing a speaker's movements of the speech articulators that are presented either at the same time with target speech (Helfer & Freyman, 2005) or temporally before target speech (Wu, Cao, Zhou, Wu, & Li, 2013; Wu, Li, et al., 2013), knowledge of a source's location (Kidd, Arbogast, Mason, & Gallun, 2005; Singh, Pichora-Fuller, & Schneider, 2008), and particularly, perceived spatial separation of target from masker (Freyman et al., 1999, 2001; Huang, Huang, Chen, Wu, & Li, 2009; Huang et al., 2008; Li, Kong, Wu, & Li, 2013; Li et al., 2004; Wu et al., 2005). Unmasking effects of all these cues are largely caused by introducing and/or facilitating listeners' selective attention to the target speech.

1.3. Precedence effect, perceived spatial separation, and facilitation of selective attention to target speech

What is perceived spatial separation? It is well known that masking of a target sound can be reduced if a spatial separation is introduced between the target and the masker. The spatial unmasking is caused by the combination of three effects: (1) the head-shadowing effect (which improves the signal-to-masker ratio (SMR) in sound-pressure level at the ear near the target), (2) the effect of interaural-time-difference (ITD) disparity (which enhances auditory neuron responses to the target sound), and (3) the perceptual effect (which facilitates both selective attention to the target and suppression of the masker). However, when the listening environment is reverberant, a sound source induces numerous reflections bouncing from surfaces, and both the unmasking effect of head shadowing and that of ITD disparity are limited or even abolished, but the perceptual unmasking caused by perceptual separation between the target and masker is still effective (Freyman et al., 1999; Kidd, Mason, Brughera, & Hartmann, 2005; Koehnke & Besing, 1996; Zurek, Freyman, & Balakrishnan, 2004). Thus, introducing a (simulated) reverberant listening condition can be used for isolating the perceptually unmasking effect. This unmasking effect is closely associated with the auditory precedence effect (see below).

What is the precedence effect and what is its role in noisy, reverberant environments? In a (simulated) reverberant environment, to distinguish signals from various sources and particularly recognize the target source, listeners need to not only perceptually integrate the direct wave with the reflections of the target source (Huang et al., 2008, 2009; Li et al. 2013) but also perceptually integrate the direct wave with the reflections of the masking source (Brungart, Simpson, & Freyman, 2005; Rakerd, Aaronson, & Hartmann, 2006). More specifically, when the delay between a leading sound (such as the direct wave from a sound source) and a correlated lagging sound (such as a reflection of the direct wave) is sufficiently short, attributes of the lagging sound are perceptually captured by the leading sound (Li, Qi, He, Alain, & Schneider, 2005), causing a perceptually fused sound that is perceived as coming from a location near the leading source (*the precedence effect*, Freyman, Clifton, & Litovsky 1991; Huang et al., 2011; Litovsky, Colburn, Yost, & Guzman, 1999; Wallach, Newman, & Rosenzweig, 1949; Zurek, 1980). Thus, this perceptual fusion (integration) is able to produce *perceptual separation* between uncorrelated sound sources. For example, when both the target and masker are presented by a loudspeaker to the listener's left and by another loudspeaker to the listener's right, the perceived location of the target and that of the masker can be manipulated by changing the inter-loudspeaker time interval for the target and that for the masker (Li et al., 2004). More specifically, for both the target and masker, when the sound onset of the right loudspeaker leads that of the left loudspeaker by a short time (e.g., 3 ms), both a single target image and a single masker image are perceived by human listeners as coming from the right loudspeaker. However, if the onset delay between the two loudspeakers is reversed only for the masker, the target is still perceived as coming from the right loudspeaker but the masker is perceived as coming from the left loudspeaker. The perceived co-location and perceived separation are based on perceptual integration of correlated sound waves delivered from each of the two loudspeakers. Note that when the two loudspeakers are symmetrical to the listener, a change between the perceived co-location and the perceived separation alters neither the SMR in sound pressure level at each ear nor the stimulus-image compactness/diffusiveness (Li et al., 2004). It has been confirmed that perceived target-masker spatial separation facilitates the listener's selective attention to target signals and significantly improves recognition of target signals (Freyman et al., 1999; Huang et al., 2008; Huang et al., 2009; Li et al., 2004; Li et al., 2013; Rakerd et al., 2006; Wu et al., 2005). Moreover, it has been known that the perceptual fusion can be induced by headphone simulation of the presentation of the direct and reflection waves (Brungart et al., 2005; Huang et al., 2011; also see a review by Litovsky et al., 1999).

1.4. ERP recordings are useful for examining effects of attentional modulation

Event-related potentials (ERPs) offer a way to study the effects of masking on speech processing under both passive and active listening conditions (Alho, 1992; Bennett, Billings, Molis, & Leek, 2012; Billings, Bennett, Molis, & Leek, 2011; Martin & Stapells, 2005; Tremblay, Friesen, Martin, & Wright, 2003). This is in contrast to psychophysical studies of speech recognition that require the listener to attend to and repeat the target sentence immediately after the stimulus presentation (e.g., Freyman et al., 1999; Li et al., 2004). Thus, when a masker is present, using the ERP-recording method, both the effect of introducing attention to target speech (by shifting attention from irrelevant stimuli to target speech) and the effect of facilitating attention to target speech (by moving the masker image away from the attention focus on target speech) on cortical representations of the target speech signal can be studied.

It has been known since the Hillyard, Hink, Schwent, and Picton, (1973) that auditory ERPs can be enhanced by attention to the sound presentation (Nager, Estorf, & Münte, 2006; Snyder, Alain, & Picton, 2006; Woldorff & Hillyard, 1991; Woods, Alho, & Algazi, 1994). However, it is still not very clear (1) whether the enhancing effect of attention is predominantly on the primary and/or secondary auditory cortex or equally on all the auditory cortical regions (for reviews see Fritz, Elhilali, David, & Shamma, 2007; Muller-Gass & Campbell, 2002), and more importantly, (2) whether the attentional facilitation of auditory ERPs depends on listening conditions, particularly when a disrupting masker background is presented.

The N1/P2 ERP complex, a group of components of the early cortical auditory-evoked potentials, can be reliably elicited by speech stimuli (e.g. single syllables) even when a noise or a speech masker is co-presented (Billings et al., 2011; Martin, Kurtzberg, & Stapells, 1999; Martin, Sigal, Kurtzberg, & Stapells, 1997; Martin & Stapells, 2005; Muller-Gass, Marcoux, Logan, & Campbell, 2001; Polich, Howard, & Starr, 1985; Tremblay et al., 2003; Whiting, Martin, & Stapells, 1998). It has been recently reported that, relative to a steady-state noise masker, a four-talker speech masker with a SMR of −3 dB causes a larger masking effect on the N1 component to spoken syllables when listeners' attention was drawn away from the acoustic signals (the passive homogenous paradigm) (Billings

et al., 2011). Also, to examine whether attention affects ERPs under masking conditions, Billings et al. (2011) collapsed waveforms across the three masking conditions (continuous steady-state noise, interrupted noise, four-talker speech) and found that the N1 amplitude was significantly larger under the active paradigm than the passive paradigm, indicating a facilitating effect of introducing attention to the target sound on the ERP component. However, it is still not clear whether the effect of attentional introduction is masker-type dependent. More importantly, as mentioned above, perceived separation perceptually moves the masker image away from the target image, and consequently facilitates selective attention to the target. However, it is not clear whether introducing perceptual separation between the target-speech signal and the masker also affects ERPs to the target signal when the listening condition is either passive or active.

1.5. The goals of this study

To verify whether the unmasking effect of attentional modulation on cortical representations of speech signals is masker-type dependent, this study examined how scalp ERPs to a speech syllable under masking are modulated by attention and particularly whether the attentional modulation is different between noise- and speech-masking conditions. Specifically, ERPs to the speech syllable /*bi*/ were recorded under either a passive-listening condition (listeners attended to irrelevant video presentations) or an active-listening condition (listeners attended to the target syllable) when the masker was steady-state speech-spectrum noise, amplitude-modulated speech-spectrum noise, or speech. Since enhanced attention to target speech is mainly associated with a release of target speech from informational masking, it was predicted that a shift of the listening condition from the passive one to the active one would cause a larger enhancement of ERPs to the target syllable when the masker is speech than when the masker is noise.

More importantly, for each of the listening-condition/masker-type combinations, this study also examined whether ERPs to the target syllable are affected by changing the perceived location relationship between the target and the masker. Since a shift from perceived co-location to perceived separation between the target syllable and the masker improves the salience of the target syllable and consequently facilitates selective attention to the syllable, it was predicted that this co-location-to-separation shift would enhance the ERPs to the target syllable when masker is speech.

The masking strength of a speech masker depends on the number of talker voices contained in the masker (Carhart, Johnson, & Goodman, 1975; Freyman et al., 2004; Wu et al., 2007). For example, both Freyman et al. (2004) and Wu et al. (2007) have reported that when a speech masker contains two-talker voices, the degree of informational masking reaches the maximum level, and progressively reduces as the number of masking-talker voices increases. Since individual syllables within the two-talker masker are still resolved by listeners, linguistic information in the masker can interfere with processing the target speech more efficiently. With an increase of the number of masker-talker voice, the speech masker becomes more noise-like and consequently more difficult to resolve individual syllables and voices in the speech masker, leading to a decrease of the informational masking effect. Thus, in this study, to maximize the informational masking effect under the speech-masking condition, two-talker speech was used as the speech masker. Also, to minimize the informational masking effect under the energetic masking condition, the steady-state noise masker used in this study was created by summing speech sounds recited by 50 talkers (Yang et al., 2007).

2. Material and methods

2.1. Participants

Twelve young adults (8 males and 4 females) with the mean age of 22.3 years (range = 19–25 years) participated in this study. They were students recruited from Peking University and gave their written informed consent to participate in this study. All participants were right handed native Chinese speakers with normal (audiometric thresholds < 25 dB HL between 250 and 8000 Hz) and bilaterally balanced hearing (interaural threshold differences at each of the frequencies did not exceed 10 dB). The participants were paid a modest stipend for their participation.

2.2. Materials and apparatus

The speech signals were consonant–vowel syllable /*bi*/ (the target stimulus) and /*di*/ (the probe stimulus for maintaining participants' attention to acoustic stimuli). They were obtained from the standardized UCLA version of the Nonsense Syllable Test (Dubno & Schaefer 1992) and modified to be 474 ms in duration. The syllables were spoken by a female talker.

Fig. 1 shows the spectrum of each of the three types of maskers used in this study: steady-state speech-spectrum noise ("steady noise" for short), speech-envelope-modulated noise ("modulated noise" for short), and two-talker speech. The noise maskers mainly produced the energetic masking effect, and the speech masker produced both the energetic and informational masking effects. The steady-state noise masker was a 10-s continuous noise loop created by mixing in total 113 Chinese sentences voiced by 50 different speakers (each of the speakers spoke different sentences) (Yang et al., 2007). The speech masker was a 10-min loop of digitally combined continuous recordings of two different streams of Chinese nonsense sentences (e.g., the English translation of one of the sentences is "*One appreciation could retire his ocean.*"), spoken by two female talkers, respectively. These two masking talkers spoke different nonsense sentences (321 sentences in total, without repetition) whose waveforms were mixed with equal root-mean-square (RMS) levels from the two sources (Yang et al., 2007). The average fundamental frequencies of the two masking voices were 229 and 240 Hz, respectively. Thus, during the target/masker presentation, the target syllable (with the fundamental frequency of 254 Hz) was presented against the two-talker speech background. Since the duration of the nonsense sentence varied between 3.3 and 3.6 s and each of the two masking talkers spoke different sentences with different speeds, there was no regular relation in sentence phase between the two masking talkers' speech streams. Also, the loop was started randomly at a point for a test trial. The modulated noise masker was speech-spectrum noise modulated by the envelope of the two-talker speech with the low-pass frequency of 50 Hz and the high-cutoff frequency of 150 Hz. The envelope was extracted with the Hilbert transform (Oppenheim, Schafer, & Buck, 1989) as used by previous investigators (Smith, Delgutte, & Oxenham, 2002; Zeng et al., 2005).

Using a computer with a Pentium IV processor, all acoustic signals were digitized at the sampling rate of 22.05 kHz with the 24-bit amplitude quantization. These signals were transferred using a Creative Extigy sound blaster and presented to participants by two tube ear inserts used for ERP recordings (Neuroscan, El Paso, TX, USA).

The target syllable was presented with the right ear leading the left ear by 3 ms. Thus, participants always perceived the target image as coming from the right ear across trials. For the perceptual co-location condition, both the masker and the target syllable were presented with the right ear leading the left ear by 3 ms and

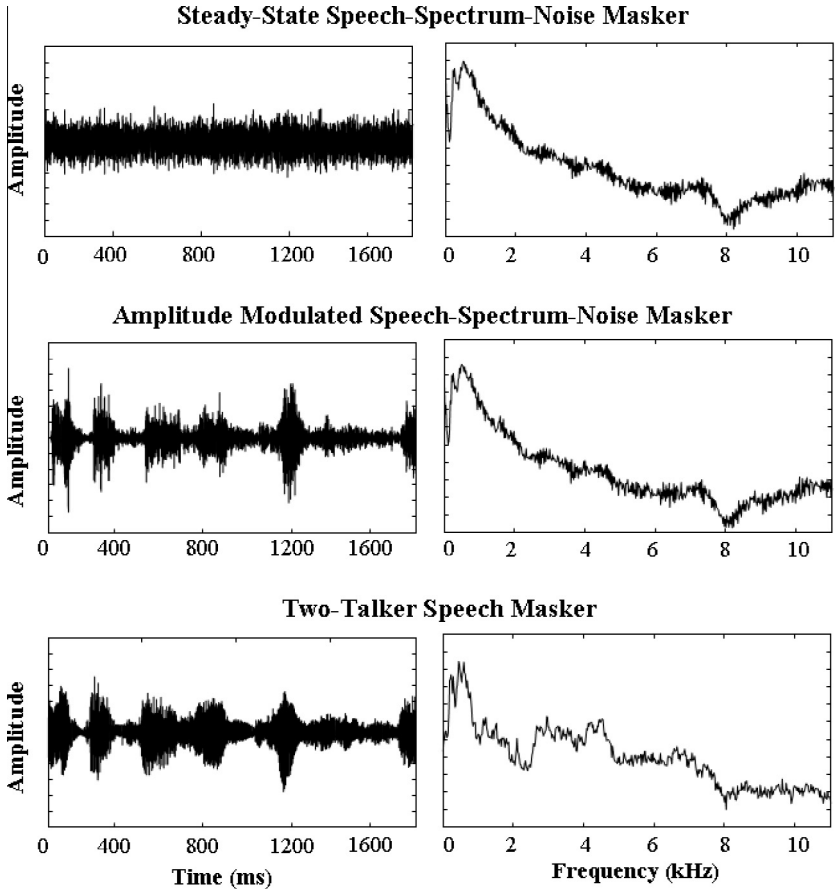

Fig. 1. The waveforms (left panels) and spectra (right panels) of a section of the three maskers.

perceived as coming from the right ear. On the contrary, for the perceptual separation condition, the masker was presented with the left ear leading the right ear by 3 ms. Note that a shift between the perceptual co-location condition and the perceptual separation condition did not alter either the SMR or the compactness/diffuseness of sound images.

2.3. Electrophysiological recordings

ERP recordings were conducted in a dim double-walled sound-attenuating booth (EMI Shielded Audiometric Examination Acoustic Suite) that was equipped with a 64-channel NeuroScan SynAmps system (Compumedics Limited, Victoria, Australia). The participant was seated 1 m in front of a 12-inch Lenovo monitor.

Electroencephalogram (EEG) signals were recorded by the NeuroScan system with a sample rate of 1000 Hz and the reference electrode located on the nose. EEG signals were on-line amplified 500 times and band-pass filtered between 0 and 200 Hz. Waveforms were then off-line band-pass filtered between 1 and 30 Hz (Billings et al., 2011). Eye movements and eye blinks were recorded from electrodes located superiorly and inferiorly to the left eye and at the outer canthi of the two eyes. Ocular artifacts exceeding ±70 μV were rejected before averaging. A recording period including 100 ms before (served as the baseline) and 500 ms after the target-syllable onset was used for data analyses.

The averaged ERPs evoked by the target syllable /*bi*/ under each of the 12 conditions were analyzed across participants. Both the N1/P2 peak-to-peak amplitudes and the latencies of the N1 and P2 components were statistically analyzed.

2.4. Procedures

To examine the effects of each of the following factors: (1) masker type, (2) listening condition, and (3) perceptual location relation, 12 recording blocks were used to encompass all the possible 12 combinations of these 3 factors (3 masker types: steady noise, modulated noise, two-talker speech; 2 listening conditions: passive, active; 2 perceptual locations of masker related to target: perceptually co-located, perceptually separated). Each block contained 300 trials with the duration of 1800 ms for each, including 240 trials presenting the syllable /*bi*/, and 60 trials presenting the deviant syllable /*di*/. The order of the 12 recordings blocks was counterbalanced across participants.

The target syllable was presented at the sound pressure level of 60 dBA at each ear, and the SMR was −4 dB for each of the masker types. Calibration of stimuli was completed by a Larson Davis Audiometer Calibration and Electroacoustic Testing System (Audit and System 824, Larson Davis, USA).

Under the passive-listening conditions, participants were asked to watch a silent cartoon movie and ignore sounds presented from the earphones during ERP recordings. A trial was started with the masker, and then the target was presented within a 1000–1200 ms window after the onset using a different, randomized starting time for each block. The trial interval was 1200 ms. It took about 12 min to complete one block under the passive-listening condition.

Under the active-listening condition, both the stimuli and procedures were identical to those under the passive-listening condition except that to maintain participants' attention to the acoustic stimuli, participants were instructed to attend to sounds presented from the earphones and press a button after a trial if

they had heard the probe syllable /*di*/, whose fundamental frequency was 258 Hz. To limit eye movements, participants were also asked to watch a cross in the centre of the monitor. The interval between trials was 2000 ms. Due to the time for button-pressing responses, it took longer time (about 15 min) to complete one recording block under the active condition.

3. Results

3.1. Amplitudes of ERPs to the target speech syllable

Fig. 2 shows average ERP waveforms at each of the electrode sites across the 6 passive-listening conditions (associated with 6 masker-type/perceptual-location combinations, Panel A) and those across the 6 active conditions (Panel B). The N1/P2 complex was salient at the fronto-central electrode sites, and did not exhibit obvious differences between the left and right hemispheres. Since the N1/P2 complex at the center site (Cz) was the most salient (also see Martin et al., 1997, 1999; Martin & Stapells, 2005; Tremblay et al., 2003), both the N1/P2 peak-to-peak amplitude and the latencies of the N1 and P2 components recorded from the site Cz were selected for statistical analyses.

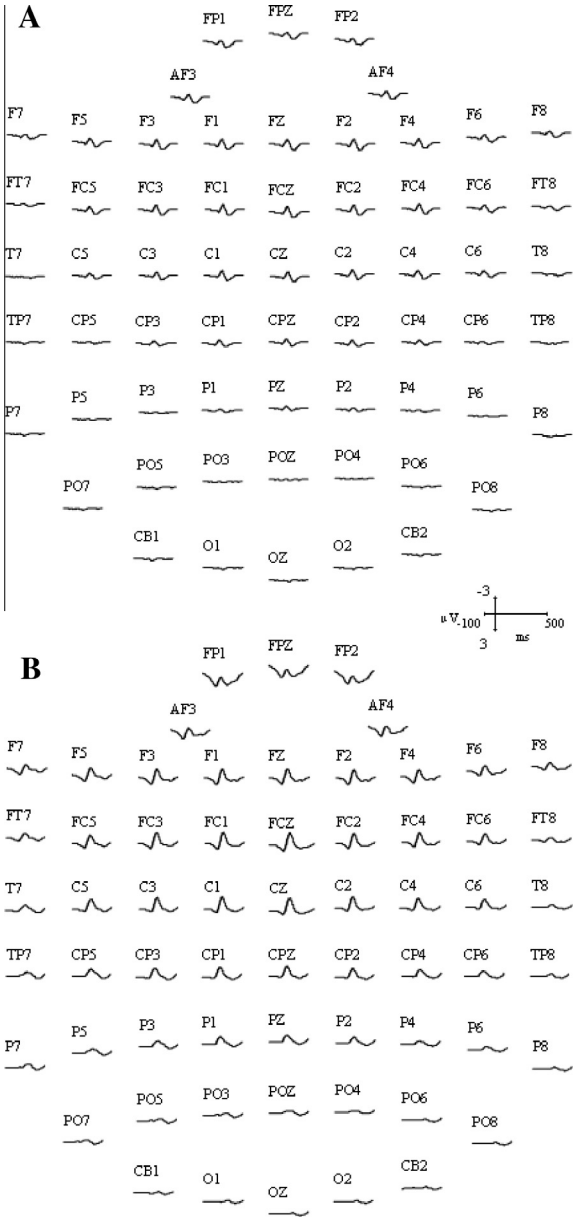

Fig. 2. Average waveforms at each of the electrode sites across the 6 passive-listening conditions (Panel A) and those across the 6 active-listening conditions (Panel B). Note that for the electrode sites surrounding the site Cz, the average amplitude to the syllable /*bi*/ was larger under the active-listening condition than that under the passive-listening condition.

Grand mean ERP waveforms recorded from the electrode site Cz across participants to the target syllable /*bi*/ under each of the 12 conditions are shown in Fig. 3. Obviously, the syllable evoked a much larger N1/P2 complex when the masker was noise (either steady or modulated) than when the masker was speech, especially under the passive-listening condition. Also, the N1/P2 complex amplitude was generally larger when the target and masker were perceptually separated than when they were co-located under the passive-listening condition when the masker was noise and under the active-listening condition when the masker was speech. Furthermore, a shift from the passive-listening condition to the active-listening condition markedly enhanced the N1/P2 complex, especially when the masker was speech.

The average values of N1/P2 peak-to-peak amplitudes to syllable /*bi*/ across participants under each of the 12 conditions are displayed in Fig. 4. A 3 (masker type: steady noise, modulated noise, speech) by 2 (listening condition: passive, active) by 2 (perceptual location: perceived co-location, perceived separation) repeated-measures analysis of variance (ANOVA) showed a significant main effect of relative location [$F(1,11) = 8.370$, $p < 0.05$, partial $\eta^2 = 0.432$], a significant main effect of attention type [$F(1,11) = 7.358$, $p < 0.05$, partial $\eta^2 = 0.401$], a significant main effect of masker type [$F(1,11) = 24.870$, $p < 0.001$, partial $\eta^2 = 0.693$], and a significant two-way interaction on the N1/P2 peak-to-peak amplitude between masker type and listening condition [$F(2,22) = 4.479$, $p < 0.05$, partial $\eta^2 = 0.289$]. However, the two-way interaction between masker type and perceptual location, the two-way interaction between listening condition and perceptual location, and the three-way interaction were not significant (all $p > 0.05$). To further examine the effects of each of the three factors on the N1/P2 complex, the following analyses were conducted.

3.1.1. Masker-type effects on the amplitude of the N1/P2 complex

Since the two-way interaction between masker type and listening condition was significant, the masker-type effect was examined separately under the passive-listening condition and the active-listening condition.

Under the passive-listening conditions, Bonferroni post hoc comparisons showed that the N1/P2 peak-to-peak amplitude evoked by syllable /*bi*/ was significantly smaller under the speech-masking condition than that under either the steady-noise-masking or modulated-noise-masking condition (both $p < 0.001$). However, there was no significant difference between the two noise-masking conditions ($p > 0.05$).

Under the active-listening condition, Bonferroni post hoc comparisons showed that the N1/P2 peak-to-peak amplitude to syllable /*bi*/ was also significantly smaller under the speech-masking condition than that under either the steady-noise-masking or modulated-noise-masking condition (both $p < 0.05$). There was no significant difference between the two noise-masking conditions ($p > 0.05$).

3.1.2. Listening-condition effect on the amplitude of the N1/P2 complex

The listening-condition effect was examined for each of the three types of maskers. When the masker was two-talker speech, Bonferroni post hoc comparisons showed that the N1/P2 peak-to-peak amplitude was significantly larger under the active-listening condition than under the passive-listening condition ($p < 0.05$). When the masker was either steady noise or modulated noise, the N1/P2 peak-to-peak amplitude was not significantly different between the two listening conditions (both $p > 0.05$).

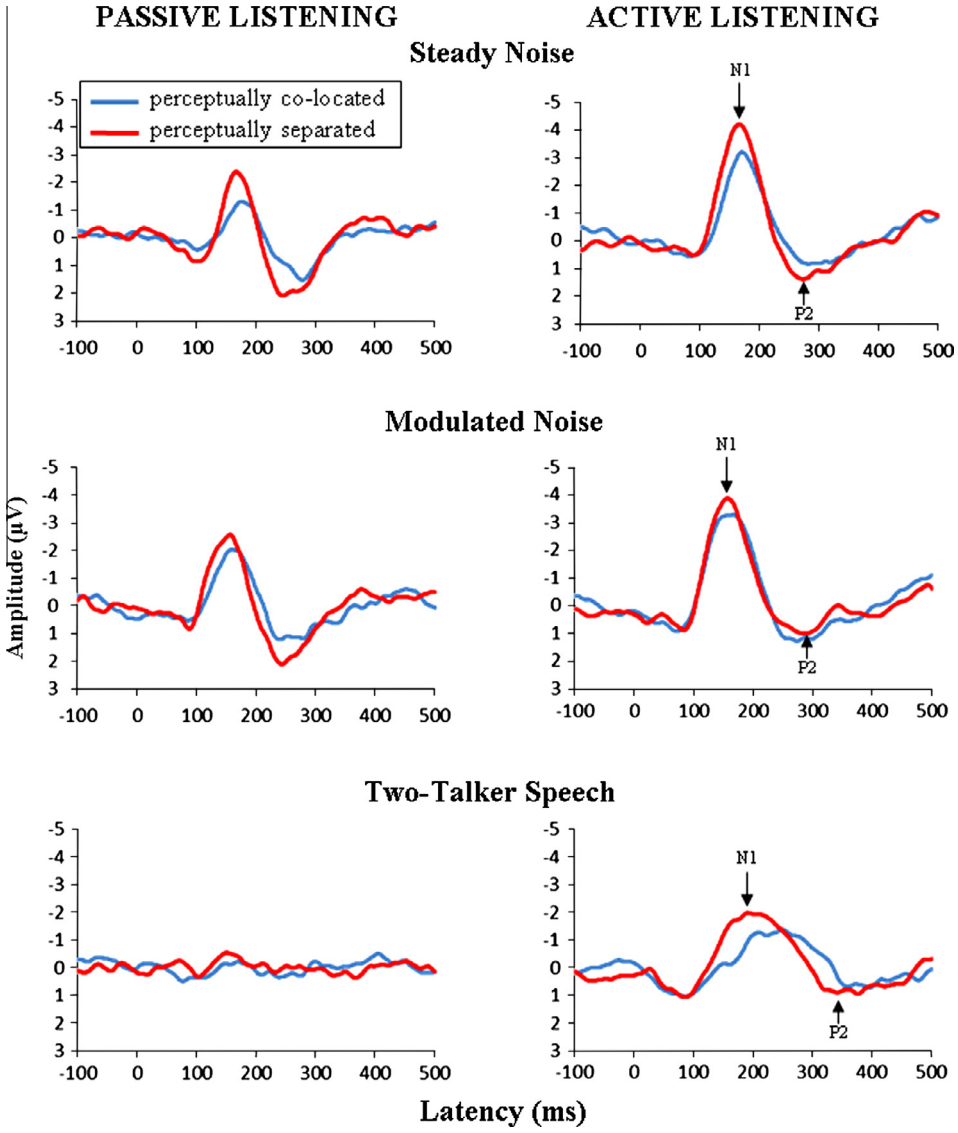

Fig. 3. Grand mean ERP waveforms recorded from the electrode site Cz across participants to the syllable /*bi*/ under each of the 12 conditions. The target syllable /*bi*/ evoked much larger N1/P2 complex when the masker was noise (either steady or modulated) than when the masker was speech, especially under the passive-listening condition.

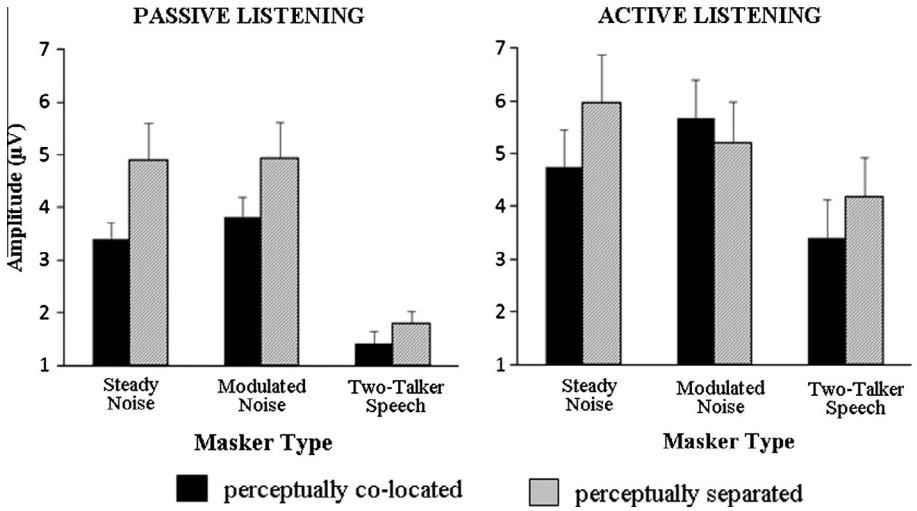

Fig. 4. Average values of N1/P2 peak-to-peak amplitudes to the target syllable /*bi*/ across participants under each of the 12 conditions.

3.1.3. Perceptual separation effects on the amplitude of the N1/P2 complex

To further examine the difference in N1/P2 peak-to-peak amplitude between the perceptual co-location condition and the perceptual separation condition for each of the three masker types, a 3 (masker type: steady noise, modulated noise, speech) by 2 (perceptual location: perceived co-location, perceived separation) two-way repeated measures ANOVA was conducted under each of the two listening conditions.

Under the passive-listening condition, the ANOVA revealed a significant main effect for both masker type [$F(2,22) = 35.850$, $p < 0.01$, partial $\eta^2 = 0.765$] and perceptual location [$F(1,11) = 10.347$, $p < 0.01$, partial $\eta^2 = 0.485$]. The two-way interaction was not significant. Post-hoc tests revealed that the N1/P2 peak-to-peak amplitude was significantly larger when the target and masker were perceptually separated than that when the target and masker were perceptually co-located ($p < 0.01$).

Under the active-listening condition, the ANOVA revealed a marginally significant two-way interaction between masker type and perceptual location [$F(2,22) = 3.162$, $p = 0.06$, partial $\eta^2 = 0.485$]. The Bonferroni post hoc comparisons showed that the N1/P2 peak-to-peak amplitude was significantly larger when the target and masker were perceptually separated than that when the target and masker were perceptually co-located only under the speech-masking condition ($p < 0.05$), but not under either the steady-noise-masking or the modulated-noise-masking condition (both $p > 0.05$).

3.2. Latencies of ERPs to the target speech syllable

Fig. 5 shows the mean values of N1 and P2 latencies across participants for each of the masker types under either the passive-listening condition (left panels) or the active-listening condition (right panels). As can be seen in Fig. 5, perceptual separation particularly shortened the N1 and P2 latencies only when the masker was speech under the speech-masking condition. The low-right panel of Fig. 3 also shows that under the active-listening conditioning, a shift from the perceptual co-location to perceptual separation shortened the N1 and P2 latencies when the masker was speech. Interestingly, a shift from the passive-listening condition to the active-listening condition increased the N1 and P2 latencies when the masker was speech.

For the N1 component, a 3 (masker type) by 2 (listening condition) by 2 (perceptual location) repeated-measures ANOVA showed that the two-way interaction between perceptual location and masker type was significant [$F(2,22) = 5.575$, $p < 0.05$, partial $\eta^2 = 0.336$], and the two-way interaction between listening condition and masker type was significant [$F(2,22) = 17.985$, $p < 0.001$, partial $\eta^2 = 0.620$]. However, neither the two-way interaction between perceptual location and listening condition nor the three-way interaction was significant (both $p > 0.05$). For the P2 component, a 3 by 2 by 2 repeated-measures ANOVA showed that the three-way interaction was significant [$F(2,22) = 13.390$, $p < .001$, partial $\eta^2 = 0.549$].

3.2.1. Perceptual separation effect on the N1 and P2 latencies under the passive-listening condition

For the N1 component, under the passive-listening condition, a 3 (masker type) by 2 (perceptual location) repeated-measures ANOVA confirmed a significant two-way interaction [$F(2,22) = 3.711$, $p < 0.05$, partial $\eta^2 = 0.252$]. Bonferroni post hoc comparisons showed that the N1 latency was significantly shorter when the target and masker were perceptually separated than when target and masker were perceptually co-located only under the steady-noise masking condition ($p = 0.001$).

For the P2 component, under the active-listening condition, a 3 (masker type) by 2 (perceptual location) two-way repeated-measures ANOVA confirmed a significant two-way interaction [$F(2,22) = 8.551$, $p < 0.01$, partial $\eta^2 = 0.437$]. Bonferroni post hoc comparisons showed that the P2 latency was significantly shorter when the target and masker were perceptually separated than when target and masker were perceptually co-located under either the steady- or modulated-noise masking condition (both $p < 0.05$), but not under the speech-masking condition ($p > 0.05$).

3.2.2. Perceptual separation effect on the N1 and P2 latencies under the active-listening condition

For the N1 component, under the active-listening condition, a 3 (masker type) by 2 (perceptual location) repeated-measures ANOVA confirmed a significant two-way interaction between the two factors [$F(2,22) = 10.333$, $p < 0.01$, partial $\eta^2 = 0.484$]. Bonferroni post hoc comparisons showed that when the masker was speech, the N1 latency was significantly shorter when the target and masker were perceptually separated than when the target and masker were perceptually co-located ($p = 0.001$). However, this perceptual location effect did not occur when the masker was either steady or modulated noise (both $p > 0.05$).

For the P2 component, under the active-listening condition, a 3 (masker type) by 2 (perceptual location) two-way repeated-measures ANOVA confirmed a significant two-way interaction on P2 latency [$F(2,22) = 3.821$, $p < 0.05$, partial $\eta^2 = 0.258$]. Bonferroni post hoc comparisons showed that the P2 latency was significantly shorter when the target and masker were perceptually separated than when target and masker were perceptually co-located under the speech-masking condition ($p < 0.05$), but not under either the steady-noise-masking or the modulated-noise-masking condition (both $p > 0.05$).

3.2.3. Listening-condition effects on N1 and P2 latencies

As shown by Fig. 5, since a shift from the passive-listening condition to the active-listening condition increased the N1 and P2 latencies when the masker was speech, the listening-condition effects on N1 and P2 latencies were examined statistically.

For the N1 component, since the three-way ANOVA showed a significant interaction between masker type and listening condition, the effect of listening condition was examined under each of the three masker types: When the masker was either steady or modulated noise, post hoc comparisons showed no significant difference in N1 latency between the passive- and active- listening conditions (both $p > 0.05$). However, when the masker was speech, a post hoc comparison showed that the N1 latency under the active-listening condition was significantly longer than that under the passive-listening condition ($p < 0.01$).

To investigate the listening-condition effect on P2 latency for each of the three masker types, a 2 (listening condition) by 3 (masker type) repeated measures ANOVA showed a significant interaction between the two factors [$F(2,22) = 72.788$, $p < 0.001$, partial $\eta^2 = 0.869$]. Post-hoc comparisons showed that that the P2 latency under the active-listening condition was significantly longer than that under the passive-listening condition when the masker was speech ($p < 0.001$), but not when the masker types was either steady or modulated noise (both $p > 0.05$).

4. Discussion

4.1. Effects of masker type

The results of the present study suggest that the early cortical processing in the primary auditory cortex is involved in differentiating the speech masking and the noise masking of speech signals.

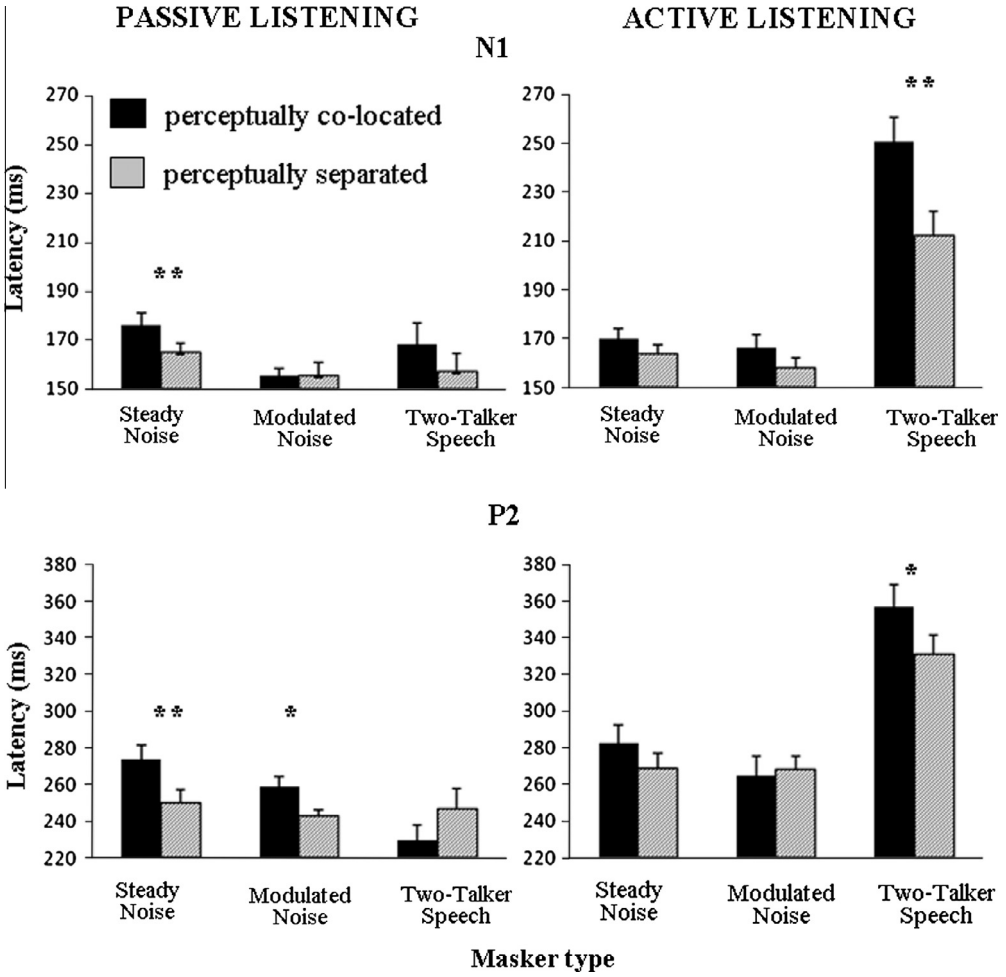

Fig. 5. The mean values of N1 and P2 latencies across participants for each of the masker types under either the passive-listening condition (left panels) or the active-listening condition (right panels). Perceptual separation particularly shortened the N1 and P2 latencies when the masker was steady noise under the passive-listening condition and when the masker was speech under the active-listening condition. A shift from the passive-listening condition to the active-listening condition prolonged the N1 and P2 latencies only when the masker was speech.

Regardless of whether the listening condition was passive or active, the peak-to-peak amplitude of the N1/P2 complex evoked by the syllable /*bi*/ was smaller under the speech-masking condition than that under either the steady-noise-masking or modulated-noise-masking condition, particularly when the target and masker were perceptually co-located. The results suggest that the two-talker speech masker caused a heavier masking effect on the early cortical representation of the target syllable than the noise maskers (also see Bennett et al., 2011). Since all three masking conditions had the same long-term SMR, the differences in masking potency between the maskers (particularly under the passive-listening condition) suggest that in addition to the energetic masking effect, irrelevant-speech-induced informational masking of speech signals occurs at early cortical processing stages. The results are generally in agreement with previous studies showing that the speech masker caused a larger masking effect on the N1 component of the ERPs to a syllable than the steady-state noise masker (Billings et al., 2011).

ERPs are summated voltages of postsynaptic potentials of neurons which are activated at approximately the same time (Luck, 2005). Since a sound with a particular feature evokes a particular group of neurons in the auditory cortex (Bendor & Wang, 2005; Nelken, Rotman, & Yosef, 1999; Rauschecker, 1997; Theunissen, Sen, & Doupe, 2000), the speech signal and speech masker, due to their similar acoustic structure, may activate neuron groups that overlap to a considerable extent, leading to a larger masking effect on activity of cortical neurons encoding speech signals. On the other hand, since both the steady-state speech-spectrum noise and the speech-envelope modulated speech-spectrum noise do not contain the specific acoustic structures of speech sounds, they do not evoke the neuronal activation patterns that are specifically evoked by speech sounds.

As mentioned in the Introduction, informational masking of target speech occurs at both perceptual (e.g., phonemic identification) and cognitive (e.g., semantic processing) levels, interfering with the psychological segregation of target speech from masking speech (e.g., Arbogast et al., 2002; Brungart, 2001; Brungart & Simpson, 2002; Durlach et al., 2003; Ezzatian et al., 2011; Freyman et al., 1999; Freyman et al., 2001; Kidd et al., 1994; Kidd et al., 1998; Li et al., 2004; Schneider et al., 2007; Wu et al., 2005). Since the speech masker causes a much larger masking effect on ERPs to the target syllable than a steady-state or amplitude-modulated noise masker even under the passive-listening condition, informational masking of speech can also occur at the level of early cortical processes, perhaps at pre-attentional stages.

Note that some previous studies, such as the Scott, Rosen, Wickham, and Wise (2004), did not provide firm evidence for an involvement of the primary auditory cortex in informational or energetic masking, but showed that different masking contexts for speech perception recruit different neural systems beyond the primary auditory cortex. Specifically, under the speech-in-noise listening condition, regions in the rostral and dorsolateral prefrontal

cortex and posterior parietal cortex are recruited; under the speech-in-speech listening condition, the bilateral superior temporal gyri and superior temporal sulci are recruited. Clearly, further brain imaging studies are needed to verify whether speech signals represented in the primary auditory cortex have different vulnerabilities to energetic masking and informational masking.

4.2. Effects of listening condition

As mentioned in the Introduction, ERP recordings make it possible to examine how shifting attention between the target signal and the irrelevant signal affects auditory processing of the target signal. The ERP study of Tervaniemi et al. (2009) has shown that musicians display enhanced MMN and N2b to speech sounds than non-musicians under the attentive-listening condition but not the passive-listening condition, suggesting that certain musical training can induce top-down modulation of the cortical processing of speech signals only when the speech signals are attended (also see Warren, 1999). Moreover, Billings et al. (2011) collapsed the waveforms across the various masking conditions and examined whether attention affects ERPs. They reported that the N1 amplitude was larger under their active paradigm than the passive paradigm, suggesting that introducing attention facilitates the early cortical ERP component to target speech signals.

In the present study, one of the striking results is that shifting listener's attention from the irrelevant visual stimulus to the acoustic target stimulus increased both the N1/P2 complex amplitude and the N1/P2 complex latency to the target syllable when the masker was two-talker speech but not when the masker was steady noise (under the either separation or co-location condition) or modulated noise (under the separation condition). The results suggest that particularly under speech-on-speech masking situations, although the early cortical representation of the target syllable is suppressed, it is still retained at early auditory processing levels, and introducing selective attention to the target syllable largely reduces the suppression of the early cortical representation of the target. In other words, the attentional release of cortical representation of the speech signal is masker-type specific.

Moreover, previous visual ERP studies have shown that the latency of the Nl component increases with the effort expended in processing stimuli (Callaway & Halliday, 1982). Thus, the increase of the N1/P2 complex latency to the target syllable, which was caused by the shift from the passive-listening condition to the active-listening condition, may reflect an increased attentional effort at attending to the target syllable against the speech masker.

4.3. Effects of perceptual separation

The main purpose of this study was to investigate whether introducing perceptual separation between the target syllable and the masker, which alters neither the SMR nor the compactness/diffuseness of the sound images, can facilitate ERPs to the target syllable. Previous human psychoacoustic studies have confirmed that the perceptual separation between the target speech and a masker (especially a speech masker), which are presented with either loudspeakers or headphones, can facilitate the listener's selective attention to target speech and improve recognition of target speech (Brungart et al., 2005; Freyman et al., 1999; Huang et al., 2008; Huang et al., 2009; Li et al., 2004; Li et al., 2013; Rakerd et al., 2006; Wu et al., 2005). In agreement with the results of the previous psychophysical studies mentioned above, our study provides the first evidence that when listeners attend to a target syllable, the early cortical ERP component (N1/P2 complex) to the speech syllable is enhanced, but only when the masker is also speech and not when it is a noise masker. This speech-specific effect presumably arises through the perceptual separation of the target syllable from the speech masker. Moreover, in this study, also under the active-listening condition, both the N1 and P2 latencies became shorter when the target and the speech masker, but not the noise masker, were perceptually separated. The results suggests that when listeners attend to the target speech signal under speech (informational) masking conditions, removing the speech-masker image away from the target image further facilitates selective attention to the target, thereby enhancing the early cortical representation of the target syllable on the basis of the facilitating effect of active listening.

Surprisingly, when participants attended to irrelevant visual stimuli under the passive-listening condition, introducing the difference in leading ear between the target syllable and the noise masker also enhanced the N1/P2 peak-to-peak amplitude and decreased the N1/P2 latencies. Since participants did not attend to the target syllable, this release from noise masking should not be associated with attentional processes but could be explained as a neurophysiological process of binaural unmasking which has been demonstrated even in anesthetized laboratory animals (e.g., Du, Huang, Wu, Galbraith, & Li, 2009; Du, Kong, Wang, Wu, & Li, 2011; Du, Ma, Wang, Wu, & Li, 2009).

5. Summary

(1) Under either the active-listening condition or the passive-listening condition, the two-talker-speech masker induced a much larger masking effect than either the steady-state speech-spectrum noise masker or the speech-envelope modulated speech-spectrum noise masker on the N1/P2 complex to the target syllable, suggesting that irrelevant-speech-induced informational masking of speech signals occurs at early cortical processing stages.

(2) A shift from the passive-listening condition to the active-listening condition enhances the N1/P2 complex to the target syllable, particularly when the masker is speech. Thus, informational masking suppresses but not abolishes the early cortical representation of the target syllable, and the retained cortical representation of the target signal can be released from informational masking by selective attention to the target.

(3) More importantly, perceptual separation between the attended target syllable and the speech masker (but not any of the noise maskers) further enhances the early cortical representation of the target signal by promoting the selective attention.

Significances

This study for the first time shows that under the informational masking condition, but not the energetic masking condition, facilitating selective attention to the target-speech signal unmasks the early cortical representation of the target-speech signal.

Acknowledgments

This work was supported by the Research Special Fund for Public Welfare Industry of Health (201202001), the National Natural Science Foundation of China (31170985, 31200759), the "973" National Basic Research Program of China (2011CB707805), and "985" grants from Peking University. Mr. Sheng-Chuang Feng assisted in many aspects of this work.

References

Alho, K. (1992). Selective attention in auditory processing as reflected by event-related brain potentials. *Psychophysiology, 29*, 247–263.

Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America, 112*, 2086–2098.
Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature, 436*, 1161–1165.
Bennett, K. O. C., Billings, C. J., Molis, M. R., & Leek, M. R. (2012). Neural encoding and perception of speech signals in informational masking. *Ear and Hearing, 32*, 231–238.
Billings, C. J., Bennett, K. O., Molis, M. R., & Leek, M. R. (2011). Cortical encoding of signals in noise: Effects of stimulus type and recording paradigm. *Ear and Hearing, 32*, 53–60.
Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109*, 1101–1109.
Brungart, D. S., & Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America, 112*, 664–676.
Brungart, D. S., Simpson, B. D., & Freyman, R. L. (2005). Precedence-based speech segregation in a virtual auditory environment. *Journal of the Acoustical Society of America, 118*, 3241–3251.
Callaway, E., & Halliday, R. (1982). The effect of attentional effort on visual evoked potential N1 latency. *Psychiatry Research, 7*, 299–308.
Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *The Journal of the Acoustical Society of America, 58*, 35.
Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 25*, 975–979.
Du, Y., Huang, Q., Wu, X., Galbraith, G. C., & Li, L. (2009). Binaural unmasking of frequency-following responses in rat amygdala. *Journal of Neurophysiology, 101*, 1647–1659.
Du, Y., Kong, L., Wang, Q., Wu, X., & Li, L. (2011). Auditory frequency-following response: A neurophysiological measure for studying the "cocktail-party problem". *Neuroscience and Biobehavioral Reviews, 35*, 2046–2057.
Du, Y., Ma, T., Wang, Q., Wu, X., & Li, L. (2009). Two crossed axonal projections contribute to binaural unmasking of frequency-following responses in rat inferior colliculus. *European Journal of Neuroscience, 30*, 1779–1789.
Dubno, J. R., & Schaefer, A. B. (1992). Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *Journal of the Acoustical Society of America, 91*, 2110.
Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America, 114*, 368–379.
Ezzatian, P., Li, L., Pichora-Fuller, K., & Schneider, B. A. (2011). The effect of priming on release from informational masking is equivalent for younger and older adults. *Ear and Hearing, 32*, 84–96.
Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America, 109*, 2112–2122.
Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Acoustical Society of America Journal, 115*, 2246–2256.
Freyman, R. L., Clifton, R. K., & Litovsky, R. Y. (1991). Dynamic processes in the precedence effect. *Journal of the Acoustical Society of America, 90*, 874–884.
Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America, 106*, 3578–3588.
Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention – Focusing the searchlight on sound. *Current Opinion in Neurobiology, 17*, 437–455.
Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America, 117*, 842–849.
Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*, 177–180.
Huang, Y., Huang, Q., Chen, X., Qu, T., Wu, X., & Li, L. (2008). Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults. *Hearing Research, 244*, 51–65.
Huang, Y., Huang, Q., Chen, X., Wu, X., & Li, L. (2009). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1618–1628.
Huang, Y., Li, J. Y., Zou, X. F., Qu, T. S., Wu, X. H., Mao, L. H., et al. (2011). Perceptual fusion tendency of speech sounds. *Journal of Cognitive Neuroscience, 23*, 1003–1014.
Huang, Y., Xu, L., Wu, X., & Li, L. (2010). The effect of voice cuing on releasing speech from informational masking disappears in older adults. *Ear and Hearing, 31*, 579–583.
Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America, 118*, 3804–3815.
Kidd, G., Mason, C. R., Brughera, A., & Hartmann, W. M. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica United Acustica, 91*, 526–536.
Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America, 95*, 3475–3480.
Kidd, G., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998). Releas from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America, 104*, 422–431.
Koehnke, J., & Besing, J. M. (1996). A procedure for testing speech intelligibility in a virtual listening environment. *Ear and Hearing, 17*, 211–217.
Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (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, 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*(4), e63106.
Li, L., Qi, J. G., He, Y., Alain, C., & Schneider, B. (2005). Attribute capture in the precedence effect for long-duration noise sounds. *Hear Research, 202*, 235–247.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *Journal of the Acoustical Society of America, 106*, 1633–1654.
Luck, S. (2005). *An introduction to the event-related potential technique*. Cambridge: The MIT Press.
Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *Journal of Speech, Language, and Hearing Research, 42*, 271–286.
Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Journal of the Acoustical Society of America, 101*, 1585–1599.
Martin, B. A., & Stapells, D. R. (2005). Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear and Hearing, 26*, 195–213.
Miller, G. A. (1947). The masking of speech. *Psychological Bulletin, 44*, 105–129.
Muller-Gass, A., & Campbell, K. (2002). Event-related potential measures of the inhibition of information processing: I. Selective attention in the waking state. *International Journal of Psychophysiology, 46*, 177–195.
Muller-Gass, A., Marcoux, A., Logan, J., & Campbell, K. B. (2001). The intensity of masking noise affects the mismatch negativity to speech sounds in human participants. *Neuroscience Letters, 299*, 197–200.
Nager, W., Estorf, K., & Münte, T. F. (2006). Crossmodal attention effects on brain responses to different stimulus classes. *BMC Neuroscience, 7*, 31.
Nelken, I., Rotman, Y., & Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature, 397*(6715), 154–157.
Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1989). *Discrete-time signal processing*. New Jersey, USA: Prentice-Hall Press.
Polich, J., Howard, L., & Starr, A. (1985). Stimulus frequency and masking as determinants of P300 latency in event-related potentials from auditory stimuli. *Biological Psychology, 21*, 309–318.
Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *Journal of the Acoustical Society of America, 119*, 1597–1605.
Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Oto-Laryngologica, 117*, 34–38.
Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*, 559–572.
Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. (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.
Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *Journal of the Acoustical Society of America, 124*, 1294–1305.
Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*, 87–90.
Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18*, 1–13.
Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, A. D. (2009). Top-down modulation of auditory processing: Effects of sound context, musical expertise and attentional focus. *European Journal of Neuroscience, 30*, 1636–1642.
Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *The Journal of Neuroscience, 20*, 2315–2331.
Tremblay, K. L., Friesen, L., Martin, B. A., & Wright, R. (2003). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and Hearing, 24*, 225–232.
Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. *The American Journal of Psychology, 62*, 315–336.
Warren, J. D. (1999). Variations on the musical brain. *Journal of the Royal Society of Medicine, 92*, 571.
Whiting, K. A., Martin, B. A., & Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear and Hearing, 19*, 218–231.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology, 79*, 170–191.

Woods, D. L., Alho, K., & Algazi, A. (1994). Stages of auditory feature conjunction: An event-related brain potential study. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 81.

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, C., Cao, S., Zhou Wu, X., & Li, L. (2013). Temporally pre-presented lipreading cues release speech from informational masking. *Journal of the Acoustical Society of America, 133*, EL281–EL285.

Wu, X., Chen, J., Yang, Z., Huang, Q., Wang, M., & Li, L. (2007). Effect of number of masking talkers on speech-on-speech masking in Chinese. *Interspeech*, 390–393.

Wu, M., Li, H., Gao, Y., Lei, M., Teng, X., Wu, X., et al. (2012). Adding irrelevant information to the content prime reduces the prime-induced unmasking effect on speech recognition. *Hearing Research, 283*, 136–143.

Wu, M., Li, H., Hong, Z., Xian, X., Li, J., Wu, X., et al. (2012). Effects of aging on the ability to benefit from prior knowledge of message content in masked speech recognition. *Speech Communication, 54*, 529–542.

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

Wu, X., Wang, C., Chen, J., Qu, H., Li, W., Wu, Y., et al. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hearing Research, 199*, 1–10.

Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B. A., et al. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication, 49*, 892–904.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 2293–2298.

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

Zurek, P. M., Freyman, R. L., & Balakrishnan, U. (2004). Auditory target detection in reverberation. *Journal of the Acoustical Society of America, 115*, 1609.