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

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

D a P , S a H a R a C , M I B a R a a PKU, K La a Ma P (M E a),
P U , B 100871, C a

A R T I C L E I N F O

A :
Accepted 5 June 2014
Available online 1 July 2014

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

A B S T R A C T

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. E a a a a a

Under noisy listening

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. P[illegible]

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 ([illegible], 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 [illegible] 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 [illegible]

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 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. E a

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 / / under each of the 12 conditions were analyzed across participants.(of)-4504.02.72762foa perceptualN1/P2dbk0 they had heard the probe syllable / /, 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. A [illegible]: ERP [illegible]: [illegible]

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.

Grand mean ERP waveforms recorded from the electrode site Cz across participants to the target syllable / / 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 / / 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 [$F(1,11) = 8.370$, < 0.05, partial $\eta^2 = 0.432$], a significant main effect of attention type [$F(1,11) = 7.358$, < 0.05, partial $\eta^2 = 0.401$], a significant main effect of masker type [$F(1,11) = 24.870$, < 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 [$F(2,22) = 4.479$, < 0.05, partial $\eta^2 = 0.289$]. However, the two-way interaction between masker type and perceptual location, the two-way inter-

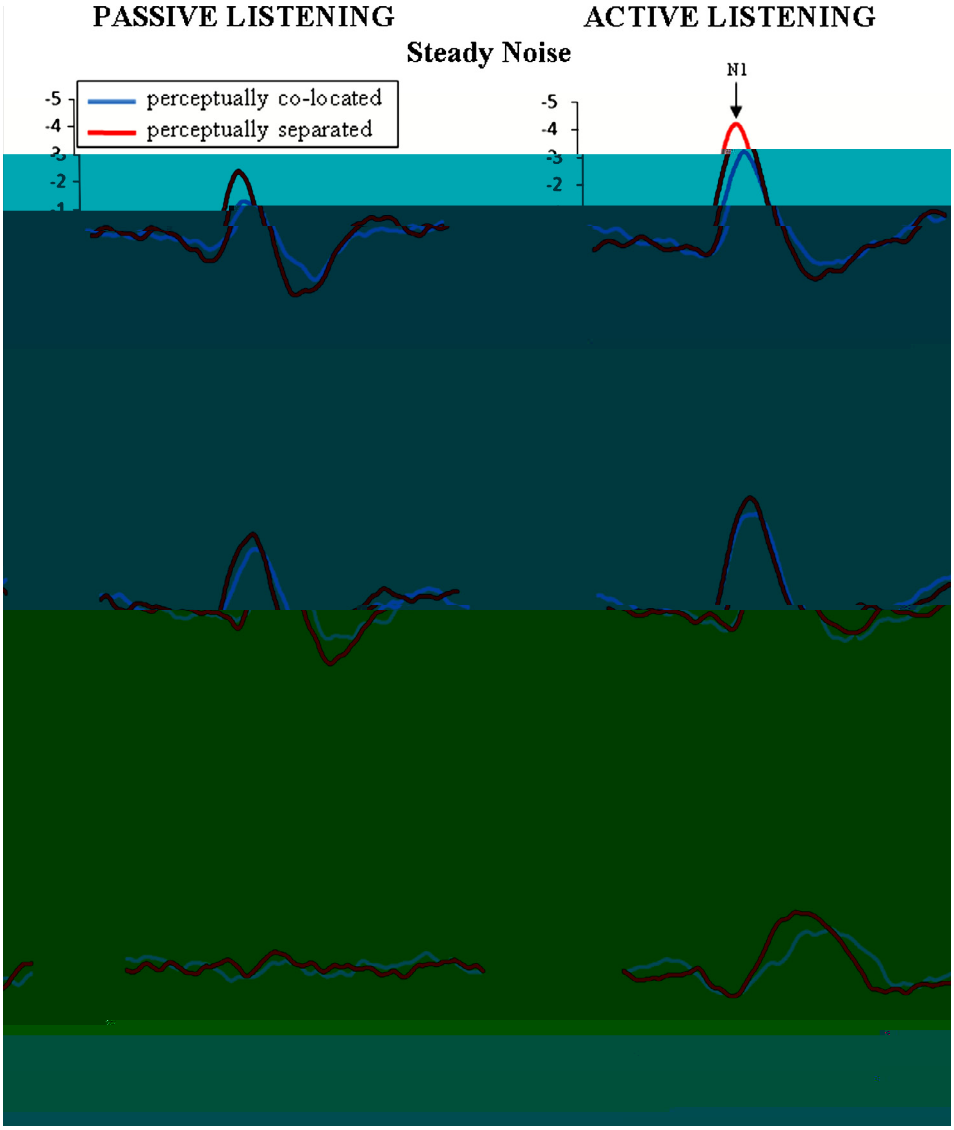

Fig. 3. Grand mean ERP waveforms recorded from the electrode site Cz across participants to the syllable / / under each of the 12 conditions. The target syllable / / evoked much larger N1/P2 complex when the masker was noise (either steady or modulated) than when the masker was speech, especially under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

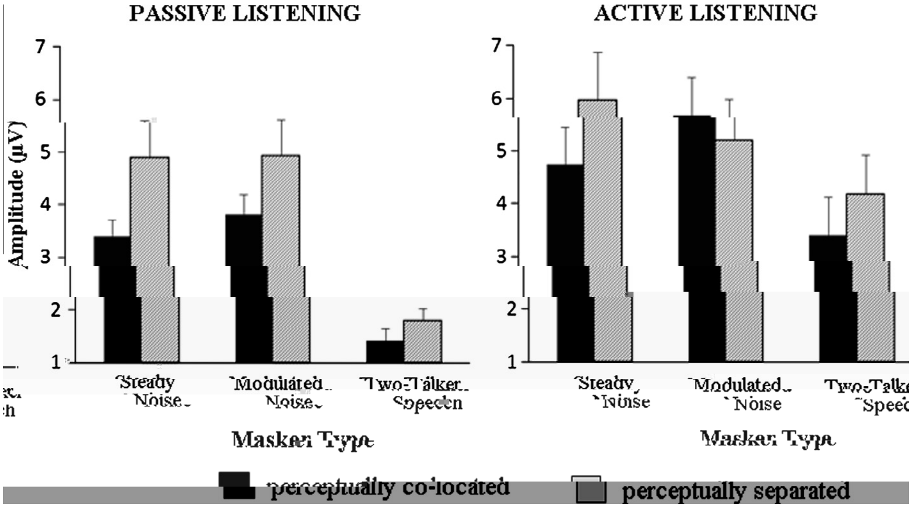

Fig. 4. Average values of N1/P2 peak-to-peak amplitudes to the target syllable / / across participants under each of the 12 conditions.

3.1.3. P a a a a N1/P2

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 [$F(2,22) = 35.850$, < 0.01, partial $\eta^2 = 0.765$] and perceptual location [$F(1,11) = 10.347$, < 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 (< 0.01).

Under the active-listening condition, the ANOVA revealed a marginally significant two-way interaction between masker type and perceptual location [$F(2,22) = 3.162$, $= 0.06$, 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 (< 0.05), but not under either the steady-noise-masking or the modulated-noise-masking condition (both > 0.05).

3.2. La ERP a a

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 [$F(2,22) = 5.575$, < 0.05, partial $\eta^2 = 0.336$], and the two-way interaction between listening condition and masker type was significant [$F(2,22) = 17.985$, < 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 > 0.05). For the P2 component, a 3 by 2 by 2 repeated-measures ANOVA showed that the three-way interaction was significant [$F(2,22) = 13.390$, $< .001$, partial $\eta^2 = 0.549$].

3.2.1. P a a a N1 a P2 a a -

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 [$F(2,22) = 3.711$, < 0.05, partial η^2

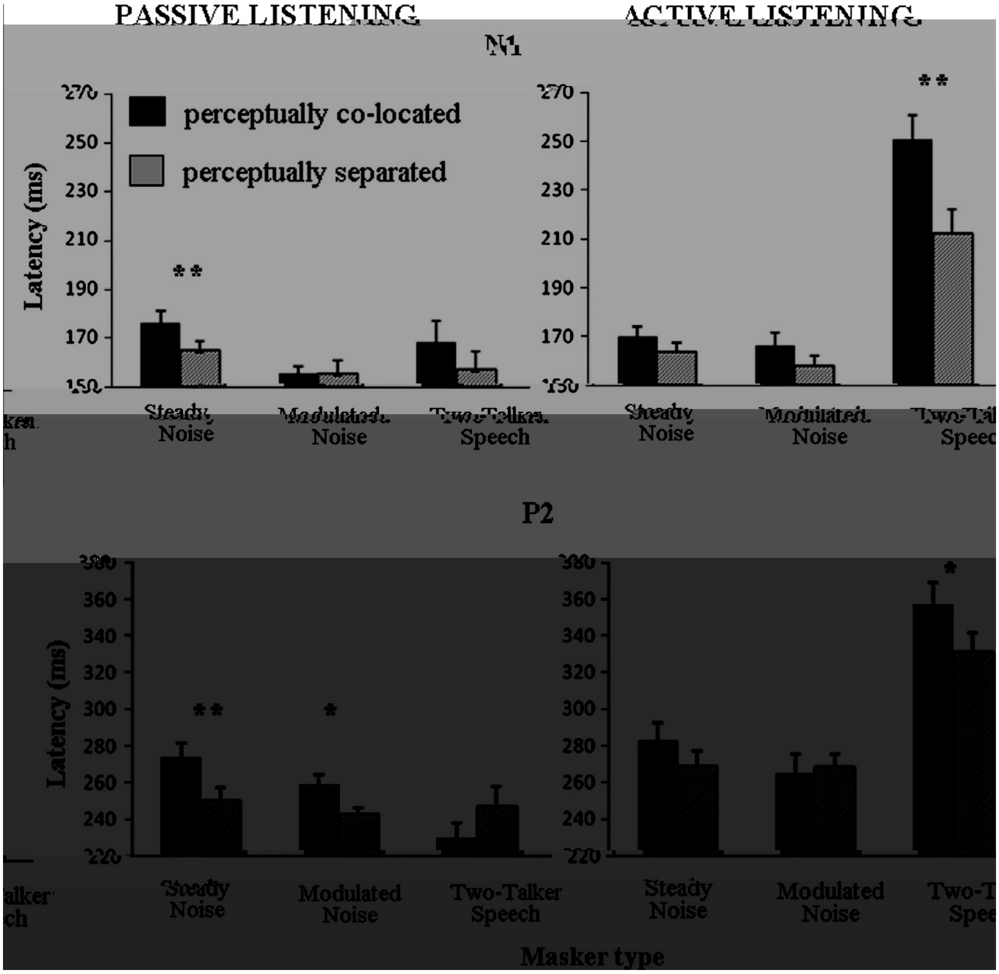

Fig. 5. The mean values of N1 and P2 latencies across participants for each of the masker types under either the passive-listening condition (left panels) or the active-listening condition (right panels). Perceptual separation particularly shortened the N1 and P2 latencies when the masker was steady 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 / i/ 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

Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. a A a A a, 112, 2086–2098.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. Na , 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. Ea a H a , 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. Ea a H a , 32, 53–60.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. a A a A a, 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. a A a A a, 112, 664–676.

Brungart, D. S., Simpson, B. D., & Freyman, R. L. (2005). Precedence-based speech segregation in a virtual auditory environment. a A a A a, 118, 3241–3251.

Callaway, E., & Halliday, R. (1982). The effect of attentional effort on visual evoked potential N1 latency. P a R a , 7, 299–308.

Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. T a A a A a, 58, 35.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. a A a A a, 25, 975–979.

Du, Y., Huang, Q., Wu, X., Galbraith, G. C., & Li, L. (2009). Binaural unmasking of frequency-following responses in rat amygdala. a N , 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". N a B a R w , 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. E a a N , 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. a A a A a, 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. a A a A a, 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. Ea a H a , 32, 84–96.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. a A a A a, 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. A a A a a, 115, 2246–2256.

Freyman, R. L., Clifton, R. K., & Litovsky, R. Y. (1991). Dynamic processes in the precedence effect. a A a A a, 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. a A a A a, 106, 3578–3588.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention – Focusing the searchlight on sound. C O N , 17, 437–455.

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. a A a A a, 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. S , 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. H a R a , 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. a E a P : H a P a P a , 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. a C N , 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. Ea a H a , 31, 579–583.

Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. a A a A a, 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. A a A a U A a, 91, 526–536.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. a A a A a, 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. a A a A a, 104, 422–431.

Koehnke, J., & Besing, J. M. (1996). A procedure for testing speech intelligibility in a virtual listening environment. Ea a H a , 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? a E a P : H a P a P a , 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. P S 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. H a R a , 202, 235–247.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. a A a A a, 106, 1633–1654.

Luck, S. (2005). A - a a . 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/. a S , La a , a H a R a , 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/. a A a A a, 101, 1585–1599.

Martin, B. A., & Stapells, D. R. (2005). Effects of low-pass noise masking on auditory event-related potentials to speech. Ea a H a , 26, 195–213.

Miller, G. A. (1947). The masking of speech. P a B , 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. I a a a P , 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. N L , 299, 197–200.

Nager, W., Estorf, K., & Münte, T. F. (2006). Crossmodal attention effects on brain responses to different stimulus classes. BMC N , 7, 31.

Nelken, I., Rotman, Y., & Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. Na , 397(6715), 154–157.

Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1989). D - a . 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. B a P , 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. a A a A a, 119, 1597–1605.

Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. A a O -La a, 117, 34–38.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. a A a A a A , 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. a A a A a, 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. a A a A a, 124, 1294–1305.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. Na , 416, 87–90.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. a C N , 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. E a a N , 30, 1636–1642.

Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. T a N , 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. Ea a H a , 24, 225–232.

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. T A a a P , 62, 315–336.

Warren, J. D. (1999). Variations on the musical brain. a R a a M , 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/. Ea a H a , 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.