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.

© 2014 Elsevier Inc. All rights reserved.

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/© 2014 Elsevier Inc. All rights reserved.

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

Download English Version:

https://daneshyari.com/en/article/925281

Download Persian Version:

https://daneshyari.com/article/925281

Daneshyari.com