Hearing Research 244 (2008) 51–65

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults

Ying Huang, Qiang Huang, Xun Chen, Tianshu Qu, Xihong Wu, Liang Li *

Department of Psychology, Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, 5 Yiheyuan Road, Beijing 100871, China

ARTICLE INFO

Article history:
Received 24 February 2008
Received in revised form 8 July 2008
Accepted 22 July 2008
Available online 30 July 2008

Keywords:
Auditory perception
Energetic masking
Informational masking
Perceptual integration
Precedence effect
Reverberant environment

ABSTRACT

This study evaluated unmasking functions of perceptual integration of target speech and simulated target-speech reflection, which were presented by two spatially separated loudspeakers. In both younger-adult listeners with normal hearing and older-adult listeners in the early stages of presbycusis, reducing the time interval between target speech and target-reflection simulation (inter-target interval, ITI) from 64 to 0 ms not only progressively enhanced perceptual integration of target-speech signals, but also progressively released target speech from either speech masking or noise masking. When the signal-to-noise ratio was low, the release from speech masking was significantly larger than the release from noise masking in both younger listeners and older listeners, but the longest ITI at which a significant release from speech masking occurred was significantly shorter in older listeners than in younger listeners. These results suggest that in reverberant environments with multi-talker speech, perceptual integration between the direct sound wave and correlated reflections, which facilitates perceptual segregation of various sources, is critical for unmasking attended speech. The age-related reduction of the ITI range for releasing speech from speech masking may be one of the causes for the speech-recognition difficulties experienced by older listeners in such adverse environments.

1. Introduction

In a noisy, reverberant room, listeners receive not only sound waves that directly emanate from various sources but also filtered and time-delayed reflections from surfaces at various locations. In such an environment, to perceptually segregate a target signal from other disruptive stimuli (which will not be as highly correlated with the target signal), the auditory system must not only integrate sound waves that directly come from the signal source with reflections of the signal source, but also at the same time, integrate sound waves that come from a disruptive source with reflections of the disruptive source. Otherwise the auditory scene will be cluttered and confusing.

Adults with normal hearing have the ability to perceptually integrate correlated sound waves. When the time interval between the direct wave coming from the source and a reflected wave of the source is sufficiently short, attributes of the delayed reflection are perceptually captured by the direct wave (Li et al., 2005), leading to a single fused image whose point of origin is perceived to be around the location of the leading source. This phenomenon is called the "precedence effect" (Wallach et al., 1949; Blauert, 1997; Litovsky et al, 1999; Li and Yue, 2002). Since a source is usually more correlated with its time-delayed reflections and less correlated (or uncorrelated) with other sources, the perceptual integration associated with the precedence effect facilitates perceived spatial segregation of various sound sources.

The importance of perceptual fusion of correlated sound waves for speech recognition under multiple-talker conditions has been experimentally demonstrated (e.g., Brungart et al., 2005; Freyman et al., 1999, 2001; Li et al., 2004; Rakerd et al., 2006; Wu et al., 2005, 2007). For example, based on the principle of the precedence effect (the correlated sound waves delivered from the two spatially separated loudspeakers are perceptually fused), when both the target speech and the masker (either speech masker or noise masker) are presented by a loudspeaker to the listener's left and another loudspeaker to the listener's right, the perceived location of the target and that of the masker can be manipulated by changing the delay between the two loudspeakers for the target signals and the masker signals (Li et al., 2004). If the masker is speech, recognizing target speech under the condition of perceived target-masker spatial separation is markedly better than that under the condition of perceived target-masker co-location, even though neither the masker energy at each ear nor the masker-image compactness/diffusiveness is substantially changed. However, when the masker is steady–state speech-spectrum noise, such a spatial separation leads to a relatively smaller (but significant) release.

* Corresponding author. Tel.: +86 905 569 4628; fax: +86 905 569 4850.
E-mail address: liangli@pku.edu.cn (L. Li).
URL: http://www.psy.pku.edu.cn/faculty/liliang_resume.htm (L. Li).

0378-5955/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2008.07.006

Because steady–state speech-spectrum noise only produces energetic masking and a speech masker produces both energetic masking and informational masking (Arbogast et al., 2002; Brungart et al., 2001; Durlach et al., 2003; Freyman et al., 1999; Kidd et al., 1994; Li et al., 2004), it appears that perceptual segregation between target speech and masking speech mainly reduces informational masking of target speech. The reduction of informational masking may be caused by the enhanced perceptual differences (i.e., in perceived spatial location) between target speech and masking speech, leading to improved selective attention to target speech (for a recent review see Schneider et al., 2007).

The advantage of perceptual integration can occur over a large range of lead–lag intervals. In a recent study by Rakerd et al. (2006), a two-talker-speech masker was presented by two spatially separated loudspeakers. One loudspeaker was located directly in front (at 0°) and the other one was 60° to the right of the listener. Both loudspeakers were 1.5 m away from the listener. The inter-loudspeaker time interval for the speech masker (inter-masker interval) was varied in a broad range from −64 to +64 ms. At the same time the target speech was presented only by the frontal loudspeaker. When the absolute value of inter-masker interval was 32 ms or shorter, there was consistent evidence of release from speech masking for target-speech recognition. However, when the inter-masker interval was either −64 or +64 ms, there was no evidence of release from masking. If the masker became speech-spectrum noise, significant release occurred only at a few short inter-masker intervals less than 4 ms. Thus the release of target speech from speech masking over a range of inter-masker intervals between 4 and 32 ms cannot be explained by a reduction in energetic masking, and the perceptual integration of the leading and lagging speech maskers must play a role in reducing informational masking of target speech. Interestingly, for the masker signals, even when the loudspeaker that delivered both the target and the masker led the loudspeaker that only delivered the masker by a time interval between 0 and 32 ms (when there was no perceived spatial separation between the target and the masker), the release was still evident, suggesting that in addition to introducing differences in perceived spatial location, introducing differences in auditory image (compactness/diffusiveness, timbre, and/or loudness) between target speech and masking speech can unmask target speech.

Brungart et al. (2005) digitally implemented head-related transfer functions (HRTFs) to generate headphone reproductions of the spatial auditory cues that occur in free-field listening. More specifically, they processed acoustic signals with HRTFs to simulate the 0° and 60° (to the right of the listener) source locations in the azimuth. Their HRTFs were derived from measurements that were made every 1° in azimuth in the horizontal plane with a compact sound source located 1 m away from a Knowles Electronic Manikin for Acoustic Research (KEMAR). Using the virtual synthesis techniques, a speech masker was simulated as being presented by two spatially separated loudspeakers: One loudspeaker was at 0° position (the frontal loudspeaker) and the other one was at 60° position to the right of the listener. The target speech was simulated as being presented only by the frontal loudspeaker. The inter-masker interval was also varied in a broad range from −64 to +64 ms. Their behavioral results show that when the masker was one- or two-talker-speech, a significant release from masking occurred across a broad range of the inter-masker intervals, and when the masker was speech-spectrum noise, a significant release occurred only at a few short inter-masker intervals.

To parse the auditory scene in a noisy, reverberant environment, perceptual integration occurs not only between correlated masking stimuli but also between the direct sound wave coming from the target source and the target reflections. Since listeners normally try to attend to target signals and ignore masking stimuli, the function of perceptually integrating target stimuli must be more important than that for masking stimuli. To our knowledge, the unmasking effect of perceptual integration of target speech with the target-reflection simulation has not been reported in the literature.

It has been well documented that recognizing speech in noisy, reverberant environments is more difficult for older-adult listeners than for younger-adult listeners (e.g., Nabelek and Robinson, 1982; Nabelek, 1988; Helfer and Wilber, 1990). As mentioned above, perceptual integration of correlated sounds in such environments is critical for perceptually segregating a target signal from other disruptive stimuli, and perceptual segregation between target speeches and masking speech mainly releases target speech from informational masking. Thus it is interesting to know whether the age-related difficulties in speech recognition under adverse conditions are related to an assumed age-related decline in the advantage of perceived spatial separation. To our knowledge, only two studies have addressed this issue (Li et al., 2004; Helfer and Freyman, 2008). Both studies have shown that speech-recognition performance in older participants was generally poorer than that in younger participants under either speech-masking or noise-masking conditions, but perceived spatial separation led to an equivalent release from same-sex speech masking between younger participants and older participants. However, it should be noted that in each of the two studies, the inter-masker interval was set only at very short values (4 ms in the Helfer and Freyman study; −3, 0, and +3 ms in the Li et al. study), and perceptual integration of leading and lagging masker signals was well established in all participants due to the short inter-masker intervals. It is not clear whether the release of speech from speech masking at longer lead–lag intervals is affected by aging. Although the age range of participants used in Rakerd et al.'s study (2006) was from 20 to 63 years, and the performance in the older participants appeared to be poorer than that in the younger participants, aging effects were not discussed in the report.

The present study investigated whether modulating the strength of the perceptual integration between the target speech and the simulated target-speech reflection affects target-speech recognition in younger listeners and older listeners, when either a speech masker or a noise masker is present. The strength of the perceptual integration of target signals was modulated by changing the time interval between the target speech and its spatially–separated single-reflection simulation (inter-target interval, ITI) over the same range (0–64 ms) that was used by Brungart et al. (2005) and Rakerd et al. (2006).

In the Chiang and Freyman study (1998), when a leading sound was delivered from a loudspeaker at 45° to the right of center and a lagging stimulus from 45° left, presenting background noise substantially reduced both the dominance of the leading sound on perceived location and the echo threshold for fusing the leading and lagging sounds. If background noise has a weakening effect on the source-reflection integration, the ITI-related modulation of the strength of perceptual integration of target signals may be influenced by the signal-to-noise ratio (SNR). Thus the present study also divided both younger participants and older participants into groups with different SNRs.

2. Methods

2.1. Participants

Thirty-six young university students (18–33 years old, mean age = 23.2 years, 23 females) and thirty-six older adults (60–75 years old, mean age = 65.9 years, 26 females) participated in speech-recognition testing in this study. Their first language was

Mandarin Chinese. None of the participants had any history of hearing disorders, and none used hearing aids. All participants gave their written informed consent to participate in the experiments and were paid a modest stipend for their participation.

The thirty-six young university students all had normal and symmetrical (no more than 15 dB difference between the two ears) pure-tone hearing thresholds (<25 dB HL) between 0.125 and 8 kHz. The thirty-six older adults had symmetrical and no more than 45 dB pure-tone hearing thresholds between 0.125 and 4 kHz. Some older adults had asymmetrical hearing thresholds exceeding 15 dB only for one frequency at 6 or 8 kHz. To introduce the between-subject factor SNR, both the thirty-six younger participants and the thirty-six older adults were randomly divided into three groups with twelve for each group. Different groups were assigned with different SNRs in speech-recognition testing. For younger participants, relative to the single-loudspeaker target sound pressure level, the sound pressure level of either two-talker-speech masker or noise masker delivered by the singleloudspeaker was adjusted to produce one of the three SNRs: −4 dB (for Younger Group 1), −6 dB (for Younger Group 2), and −8 dB (for Younger Group 3). These three SNRs occupied the middle section of younger listeners' psychometric functions that had been obtained in our previous studies (Li et al., 2004; Wu et al., 2005, 2007; Yang et al., 2007). For older participants, because our previous studies have shown that older listeners needed a higher SNR (which was less than 3 dB) than did younger listeners to achieve the younger listeners' levels of performance (Li et al., 2004), the three single-loudspeaker SNRs were −2 dB (for Older Group 1), −4 dB (for Older Group 2), and −6 dB (for Older Group 3). Fig. 1 presents average hearing levels for the six participant groups as a function of the testing-tone frequency.

As Fig. 1 shows, the thresholds of older participants were generally higher than those of younger participants, and the age difference in the threshold increased with frequency. Specifically, for frequency between 125 and 2000 Hz, the threshold difference between the younger-group mean and older-group mean was between 8 and 14 dB. For frequencies of 4000, 6000, and 8000 Hz, the differences between the younger-group mean and older-group mean were as high as 23, 26, and 35 dB, respectively. Thus younger participants and older participants were different not only in age but also in hearing sensitivity. Although these older adults were clinically normal in hearing, they were best characterized as being in the early stages of presbycusis.

2.2. Apparatus

The participant was seated at the center of an anechoic chamber (Beijing CA Acoustics), which was 560 cm in length, 400 cm in width, and 193 cm in height. Acoustic signals were digitized using the 24 bits Creative Sound Blaster PCI128 (which had a built-in anti-aliasing filter) and audio editing software (Cooledit Pro 2.0). The analog outputs were delivered to two loudspeakers (Dynaudio Acoustics, BM6 A) in the frontal azimuthal plane at the left and right 45° positions with respect to the median plane. This arrangement of loudspeaker orientation was consistent with that used in our previous studies (Li et al., 2004; Wu et al., 2005, 2007). The loudspeaker height was 140 cm, which was approximately ear level for a seated listener with average body height. Considering that the HRTFs for nearby sources (distance < 1 ms) are apparently more complicated than those for relatively distant sources (located 1 m or more from the listener) particularly for low frequencies, such as leading to enlarged binaural difference cues for auditory distance perception (e.g., Brungart and Rabinowitz, 1999; Brungart et al., 1999), in this study the distance between the loudspeaker and the center of the seated listener's head was set at 200 cm.

2.3. Stimuli

Speech stimuli were Chinese "nonsense" sentences, which are syntactically correct but not semantically meaningful. Direct English translations of the sentences are similar but not identical to the English nonsense sentences that were developed by Helfer (1997) and also used in studies by Freyman et al. (1999, 2001, 2004) and Li et al. (2004). For example, the English translation of one Chinese nonsense sentence is "*These war situations continually look into the workshop*". Each of the Chinese sentences has three key components: subject, predicate, and object, which are also the three keywords, with two characters for each (one syllable for each character). Note that the sentence frame cannot provide any contextual support for recognizing the keywords. The development of the Chinese nonsense sentences is described by Yang et al. (2007).

Target speech was spoken by a young female talker (Talker A). For the two-source target presentation, the same target sentences were presented from the two loudspeakers with the left loudspeaker either leading (in positive ITI values) or lagging behind (in negative ITI values) the right loudspeaker by the ITI of 0, 1, 2, 4, 8, 16, 32, or 64 ms for both younger groups and older groups. This range of ITIs was chosen to cover the whole range of inter-sound interval associated with the precedence effect for speech, from summing localization, to fusion, and to two separated images (Brungart et al., 2005; Rakerd et al., 2006). An additional ITI of 0.5 ms for both left- and right-loudspeaker leading conditions was used only for younger participants.

For the two-source masker presentation, which was associated with the two-source target presentation, two loudspeakers presented either (young female) two-talker-speech maskers (both talkers and contents were different between the two loudspeakers, see below) or speech-spectrum-noise maskers that were not correlated between the two loudspeakers. For the single-source masker presentation, which was associated with the single-source target presentation (by the right loudspeaker), only the right loudspeaker presented either two-talker-speech masker or noise masker.

The speech masker presented from the left loudspeaker was a 47 s loop of digitally-combined continuous recordings for Chinese nonsense sentences (whose keywords did not appear in target sentences) spoken by two different young female talkers (Talkers B and C). The speech masker presented from the right loudspeaker was also a 47 s loop of digitally-combined continuous recordings of Chinese nonsense sentences (whose keywords did not appear in target sentences also) spoken by a different pair of young female talkers (Talkers D and E). Each of the 4 masking talkers spoke different sentences and the sound pressure levels were the same across the four masking talkers' speech sounds within a testing session. In a trial, a speech masker started from a different point in the loop, therefore the loop for the left loudspeaker was not in synchrony with that for the right loudspeaker on a trial-by-trial basis.

A noise masker was a stream of steady–state speech-spectrum noise. Three hundred frequently occurring syllables were chosen from the database of *People's Daily* published for one year. One hundred and thirteen sentences, which both appeared in *People's Daily* and contained 317 syllables including all the 300 frequently occurring syllables, were selected as material for making speech-spectrum noise. The 113 different sentences were assigned to 50 Chinese young female speakers. Fifty-seven sentences were spoken by the 25 speakers and the other 56 sentences were spoken by another 25 speakers at a medium rate of speech. Recording of the sentences were stored digitally onto computer disks, sampled at 22.05 kHz and saved as 16 bits PCM wave files. All the female

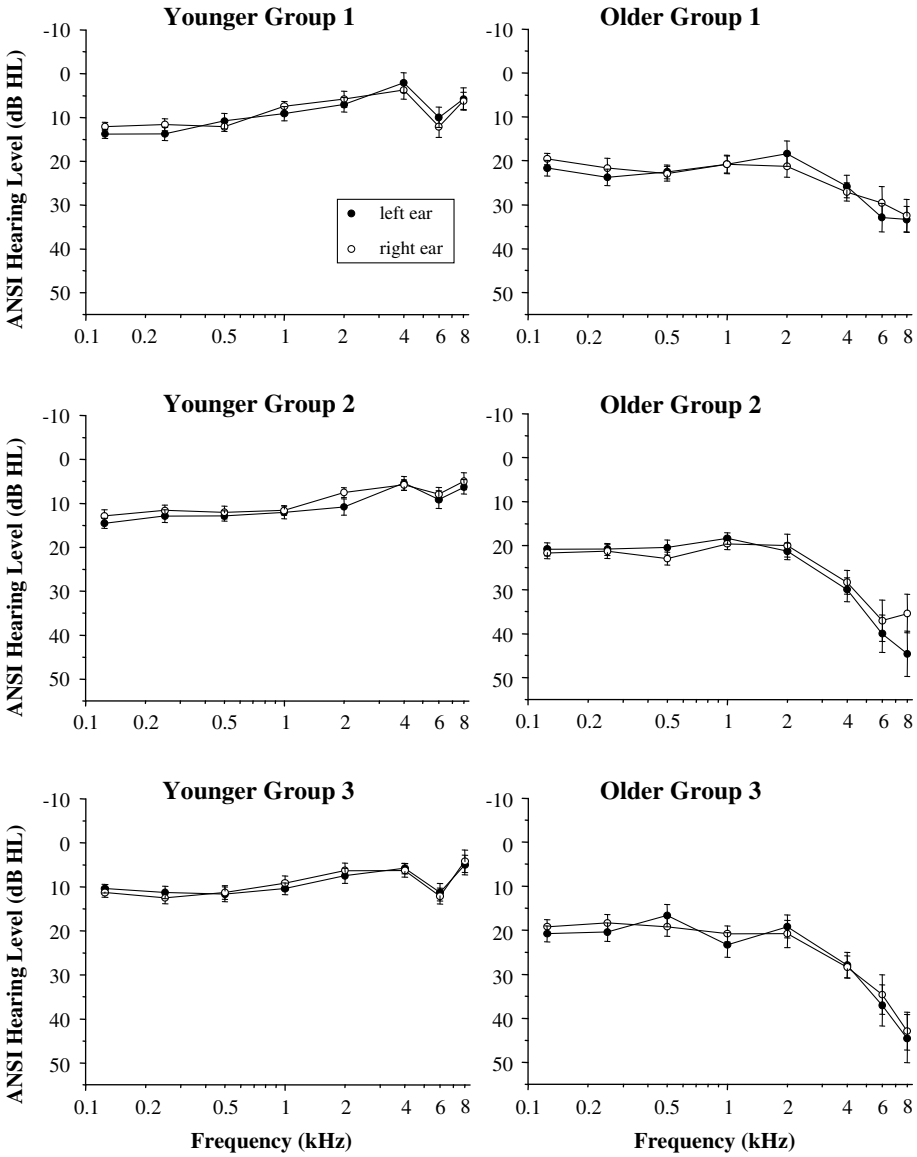

Fig. 1. Average hearing thresholds in the left ear (closed circles) and the right ear (open circles) for the three younger-participant groups (left panels) and those for the three older-participant groups (right panels) who participated in the speech-recognition tests under masking conditions. ANSI: American National Standards Institute (S3.6-1989). The error bars represent the standard errors of the mean.

speech sentences were mixed using Matlab programming and a stream of steady–state Chinese speech babble noise with the duration of 10 s was obtained. No individual talkers' voice characters could be detected in the noise.

All speech stimuli were recorded digitally onto computer disks, sampled at 22.05 kHz and saved as 16 bits PCM wave files. Speech from the loudspeaker were calibrated using a B&K sound level meter (Type 2230) whose microphone was placed at the location where the center of the listener's head would be when the listener was absent, using a "slow"/"RMS" meter response. During a session, target-speech sounds were presented at a constant level such that each loudspeaker, playing alone, would produce a sound pressure of 56 dBA. This target level allowed both younger participants and older participants to obtain near-perfect speech recognition at various ITIs when no maskers were presented (see Fig. 4) and made the masker levels at various SNRs within the comfortable range.

Fig. 2 shows the diagrams indicating the presentation configurations of target speech and maskers under different experiment conditions.

2.4. Design and procedures

For each of the six participant groups, there were two within-subject factors for the two-source-presentation condition: (1) masker type (speech masker, noise masker), and (2) ITI. A single-source condition was also used for each group. Eighteen target sentences (also 18 trials) were used in each condition. In each participant group, the order of presenting masker types was counterbalanced among twelve participants. The order of ITIs for two-source-presentation conditions was arranged in a random manner.

In each trial, the participant pressed a button on a response box to start the masker presentation (masking signals from the two loudspeakers began at the same time under two-source-presentation conditions). About 1 s later, a single target sentence was presented along with the masker signals, and then the target was gated off with the masker signals. The participant was instructed to loudly repeat the whole target sentence as best as he/she could immediately after all the sounds were completed. Performance for each participant was scored on the number of correctly identified syllables in keywords.

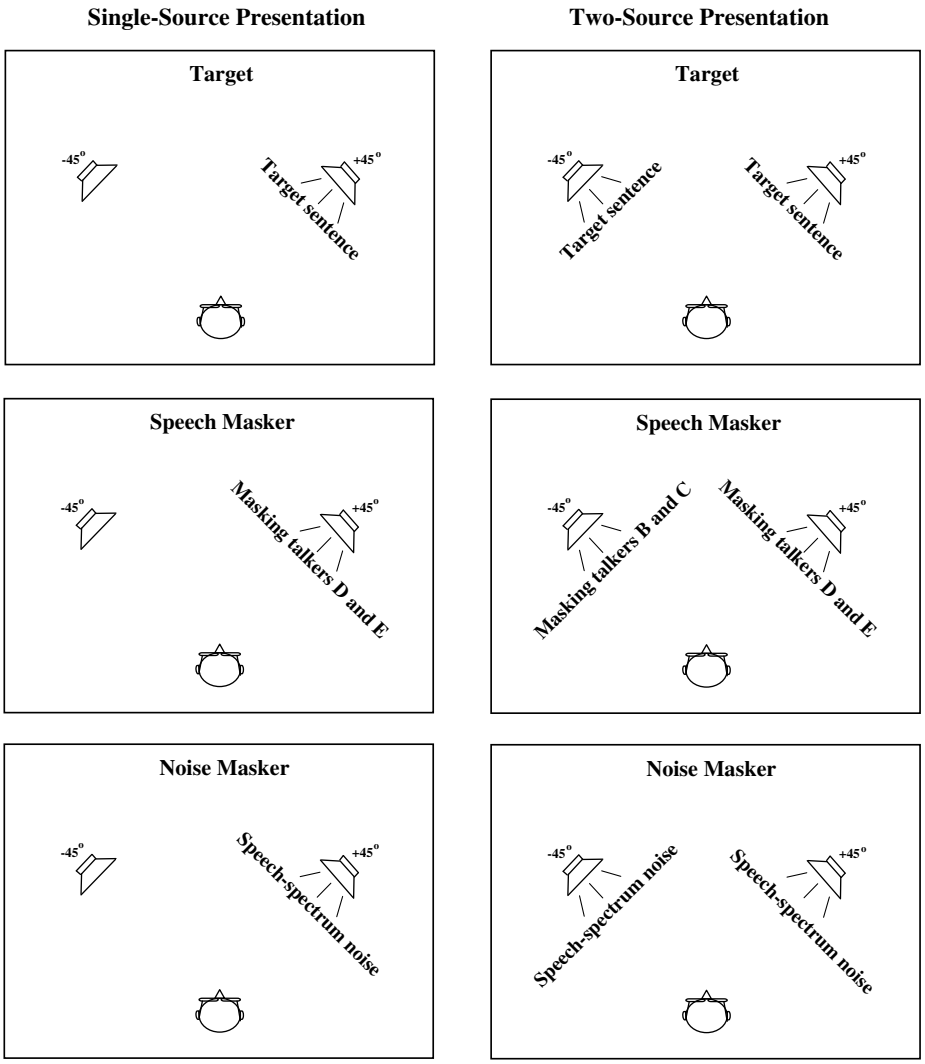

Fig. 2. Diagrams showing the presentation configurations of target speech and maskers under different experiment conditions. Under single-source-presentation conditions (left panels), target speech was presented only from the right loudspeaker, and the speech masker or the noise masker was also presented only from the right loudspeaker. Under two-source-presentation conditions (right panels), two identical target speech sounds were presented from the right and the left loudspeakers with an ITI between the two loudspeakers, and two different speech maskers (two-talker-speech masker A and two-talker-speech masker B) or two uncorrelated noise maskers (speech-spectrum-noise masker A and speech-spectrum-noise masker B) were presented from the right and the left loudspeaker simultaneously.

For all the six participant groups before formal testing, there was one training session with speech masking at the same SNR as the testing sessions. This training session consisted of 18 trials under the two-source-presentation condition with the ITI of 0 ms, and nonsense sentences, which were not used in the formal experiments, were used as target speech stimuli. No feedback was provided to participants.

Twelve younger participants who were randomly selected from the three younger-participant groups and all the twelve participants from Older Group 3 also participated in additional speech-recognition experiments under the condition without masker presentation. Seven of these younger participants, five other normal-hearing younger adults who did not participate in any speech-recognition tests in this study, and all the twelve participants from Older Group 3 were asked to describe their subjective perceptions about number, location, and compactness/diffusiveness (compact or broad) of the image(s) of the target speech at various ITIs when the masker was absent.

3. Results

3.1. Perceptual representation of the target speech at various ITIs

Twelve younger participants and twelve older participants were asked to describe the number, compactness/diffusiveness, and location(s) of the target-speech image(s) at various ITIs. Their descriptions are summarized in Fig. 3.

At the longest ITI (64 ms), all the participants perceived two target images (twilled bars in Fig. 3), one being near the location of the left loudspeaker and the other one being near the location of the right loudspeaker. At the ITI of 32 ms, eleven young participants reported that they perceived two target images, but eleven old participants reported that they perceived only one diffuse target images (black bars). When the ITI was reduced to 16 ms, the majority of the younger participants and the older participants perceived one broad image as coming from the semi-field with the leading loudspeaker, but two younger participants still perceived two target images. When the ITI was in the range between

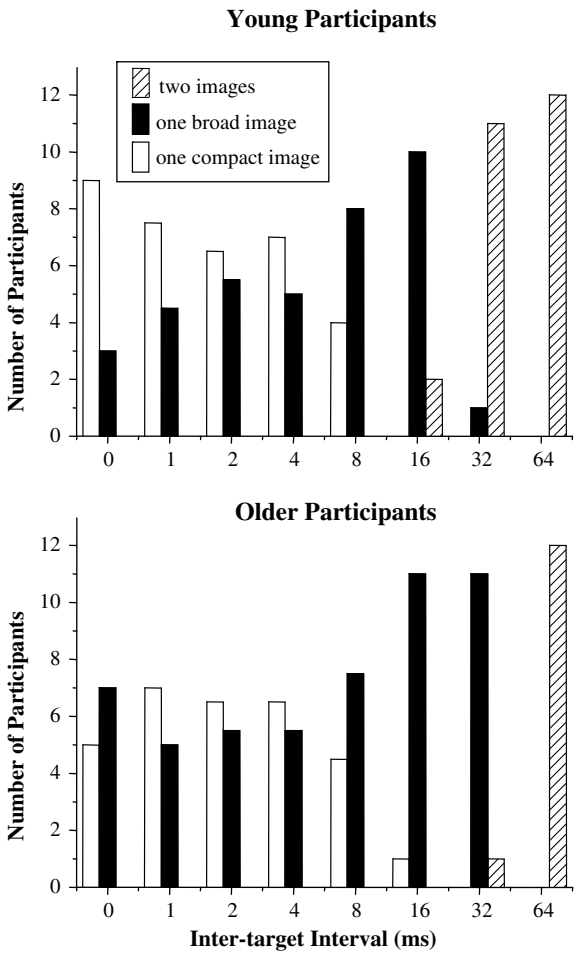

Fig. 3. The number of participants reporting their perceptions of the target-speech image at various inter-target intervals (ITIs) in younger participants (top panel) and older participants (bottom panel). Twilled bars, two separate images; black bars, a single broad image; white bars, a single compact image.

0 and 8 ms all the participants perceived only one target image as coming from either the semi-field with the leading loudspeaker or the frontal field. However, with the further reduction of the ITI, the number of participants who perceived one compact image (white bars) increased in younger participants but not in older participants.

3.2. Effects of changing ITI on speech recognition under the condition without masker

Twelve younger participants and twelve older participants participated in the speech-recognition testing when no masker was presented. The single-source-target-presentation condition was also used, in which only the right loudspeaker was used to present target stimuli. The results indicate that when the masker was absent, both younger participants and older participants obtained near-perfect speech recognition at various ITIs (Fig. 4). One-way within-subject ANOVAs show that the ITI had no significant effect on speech recognition for either younger participants ($p = 0.670$) or older participants ($p = 0.301$). In addition, under the single-source-target-presentation condition, the percent-correct recognition of target speech for the older participants (91.5%) was lower than that for the younger participants (97.6%). A one-way between-subject ANOVA indicates that the difference between the age groups was significant ($F_{1,22} = 7.918$, $p = 0.010$).

3.3. Effect of changing the ITI on releasing speech from masking in younger participants

The top panel of Fig. 5 shows percent-correct recognition of keyword syllables under the single-source-presentation condition for the three younger-participant groups when the masker was speech (black columns) or speech-spectrum noise (hatched columns). Obviously both SNR and masker type affected target-speech recognition. With the increase of SNR from −8 to −4 dB, speech recognition improved across the three younger groups under either the speech-masking condition or the noise-masking condition. One-way between-subject ANOVAs show that the effect of SNR on recognition of single-source target speech was significant for both the speech-masking condition ($F_{2,33} = 66.592$, $p < 0.001$) and the noise-masking condition ($F_{2,33} = 154.864$, $p < 0.001$). Moreover, at each SNR (for each group) the noise masker caused a larger masking effect than the speech masker (Group 1: $F_{1,11} = 25.426$, $p < 0.001$; Group 2: $F_{1,11} = 16.447$, $p = 0.002$; and Group 3: $F_{1,11} = 25.001$, $p < 0.001$).

Fig. 6 shows percent-correct recognition of keyword syllables as a function of the ITI for the three younger-participant groups when the masker was speech (left panels) or speech-spectrum noise

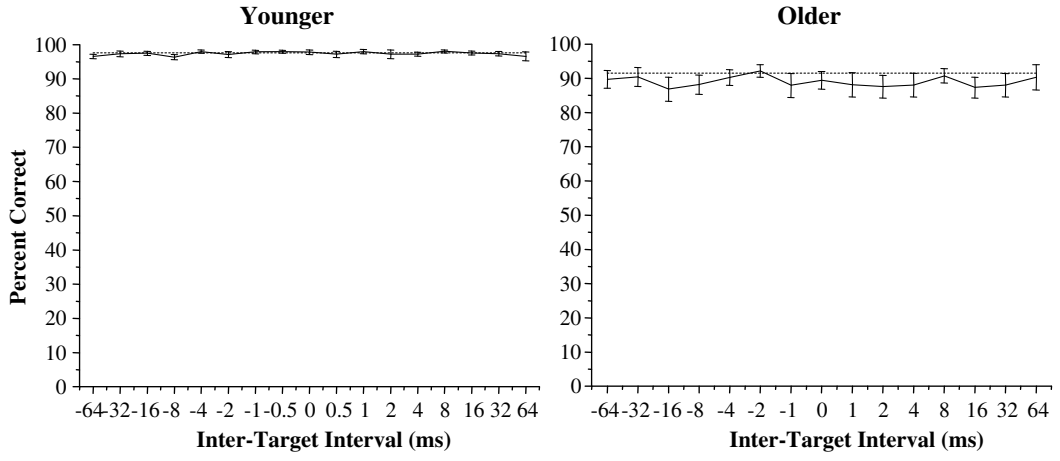

Fig. 4. Mean percent-correct recognition of keyword syllables as a function of the inter-target interval (ITI) for 12 randomly selected younger participants (left panel) and for the 12 older participants from Group 3 (right panel) when the masker was not presented. Percent-correct recognition under the single-source-presentation condition (right loudspeaker only) is shown as the broken line in each panel. The error bars represent the standard errors of the mean.

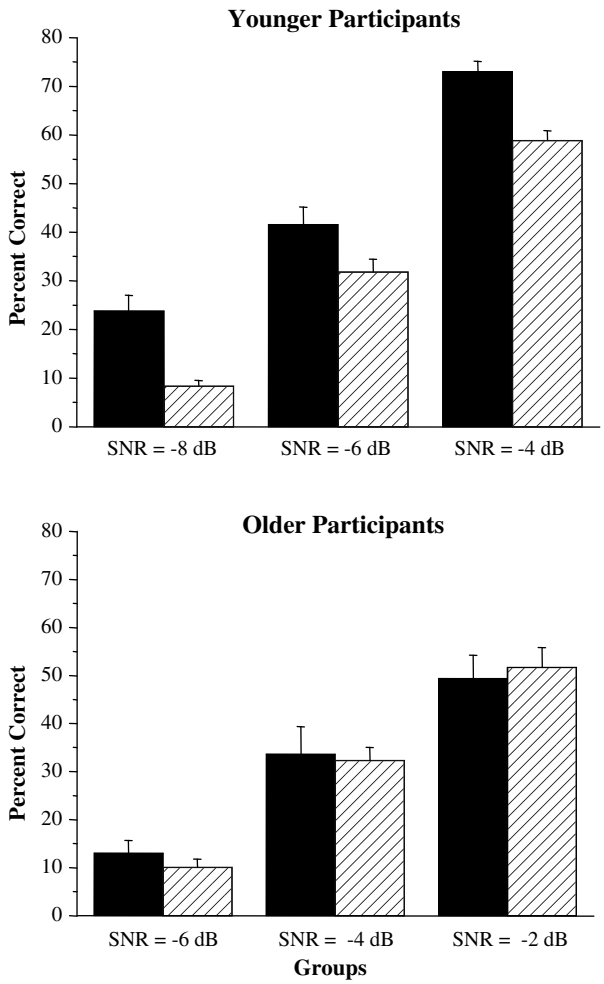

Fig. 5. Percent-correct recognition of keyword syllables in younger participants (upper panel) and older participants (lower panel), when both the target and masker were presented only by the right loudspeaker. Black columns, speech masking; hatched columns, noise masking. The error bars represent the standard errors of the mean.

(right panels). SNR, masker type, and ITI had marked influences on target-speech recognition.

When the masker was either speech or noise, with the decrease of the SNR across the three groups, the percent-correct speech recognition reduced. Under two-source-presentation conditions (when the target speech and masker were presented by the two loudspeakers), participants generally recognized more keyword syllables under short ITIs than under long ITIs. With the change of the absolute ITI value from 64 to 0 ms (left-loudspeaker or right-loudspeaker leading), the percent-correct speech recognition increased progressively. This ITI effect was also affected by both masker type and SNR: A larger ITI effect occurred when the masker was speech than when the masker was noise. And when the masker was speech, a larger ITI effect occurred at the SNR of −8 dB or −6 dB than at the SNR of −4 dB. The smaller ITI effect at the SNR of −4 dB suggests that ceiling effects reduced the ITI effect.

A mixed 3 (SNR group) by 2 (masker type) by 17 (ITI) ANOVA confirms that the main effect of SNR group was significant ($F_{2,33} = 181.960$, $p < 0.001$), the main effect of masker type was significant ($F_{1,33} = 285.811$, $p < 0.001$), and the main effect of ITI was significant ($F_{16,528} = 85.697$, $p < 0.001$). However, the three factors were highly interacted (SNR group by masker type by ITI: $F_{32,528} = 5.137$, $p < 0.001$; SNR group by masker type: $F_{2,33} = 7.360$, $p = 0.002$; SNR group by ITI: $F_{32,528} = 2.620$, $p < 0.001$; masker type by ITI: $F_{16,528} = 11.123$, $p < 0.001$).

As shown in Fig. 6, the mean percent-correct recognition of target speech was very similar between left-loudspeaker leading and right-loudspeaker leading conditions for each masker-type/SNR-group combination. Indeed, for each participant group, one-way within-subject ANOVAs and Post hoc analyses show that there was no significant difference between the two loudspeaker-leading directions at each of the ITIs for either noise-masking or speech-masking conditions. Thus data for the two leading directions were combined for analyzing the ITI-induced releasing effect.

Because speech recognition continually improved with the reduction of the ITI from 64 to 0 ms under both speech-masking conditions and noise-masking conditions, it is reasonable to use the poorest performance at the longest ITI (64 ms) as the baseline performance for two-source-presentation conditions. Based on the averaged left-right percent-correct recognition, the release of speech from masking at an ITI is defined as the difference between the percentage of correct speech recognition at the particular ITI and the percentage of correct speech recognition at the ITI of 64 ms. Fig. 7 shows the release of target speech as a function of the absolute value of ITI for each of the three younger participant groups under the speech-masking condition (filled circles) and that under the noise-masking condition (open circles).

For all the three younger-participant groups, under either the speech-masking condition or the noise-masking condition, the release obviously increased with the decrease of the absolute value of ITI, but larger releases generally occurred under the speech-masking condition. Also, the release was influenced by the SNR.

For Younger Group 1 (SNR = −4 dB), a two-way within-subject ANOVA indicates that the interaction between masker type and ITI was not significant ($F_{8,88} = 1.351$, $p = 0.230$), the main effect of masker type was not significant ($F_{1,11} = 4.152$, $p = 0.066$), and the main effect of ITI was significant ($F_{8,88} = 53.426$, $p < 0.001$). Thus the ITI-induced release of speech from speech masking and that from noise masking were similar. Follow-up *t*-tests revealed that at the level of 0.00625 (0.05/8, with a Bonferroni adjustment) the release from speech masking was significant when the ITI was 16 ms or shorter, and the release from noise masking was significant when the ITI was 32 ms or shorter.

For Younger Group 2 (SNR = −6 dB), the ITI-induced release was markedly larger under the speech-masking condition than that under the noise-masking condition. A two-way within-subject ANOVA shows that the interaction between masker type and ITI was significant ($F_{8,88} = 6.738$, $p < 0.001$), the main effect of masker type was significant ($F_{1,11} = 10.802$, $p = 0.007$), and the main effect of ITI was significant ($F_{8,88} = 45.189$, $p < 0.001$). Separate one-way within-subject ANOVAs show that the release from speech masking was significantly larger than that from noise masking at ITIs from 0 to 16 ms ($p < 0.050$), except for the ITI of 8 ms ($p = 0.070$), and the ITI effect was significant for both speech masking ($F_{8,88} = 39.982$, $p < 0.001$) and noise masking ($F_{8,88} = 19.862$, $p < 0.001$). Follow-up *t*-tests revealed that at the level of 0.00625 (with a Bonferroni adjustment) the release from speech masking was significant when the ITI was 16 ms or shorter, and the release from noise masking was significant when the ITI was 32 ms or shorter.

For Younger Group 3 (SNR = −8 dB), the ITI-induced release was also markedly larger under the speech-masking condition than that under the noise-masking condition. A two-way within-subject ANOVA shows that the interaction between masker type and ITI was significant ($F_{8,88} = 29.891$, $p < 0.001$), the main effect of masker type was significant ($F_{1,11} = 119.904$, $p < 0.001$), and the main effect of ITI was significant ($F_{8,88} = 71.935$, $p < 0.001$). Separate one-way within-subject ANOVAs show that the release from speech masking was significantly larger than that from noise masking at ITIs

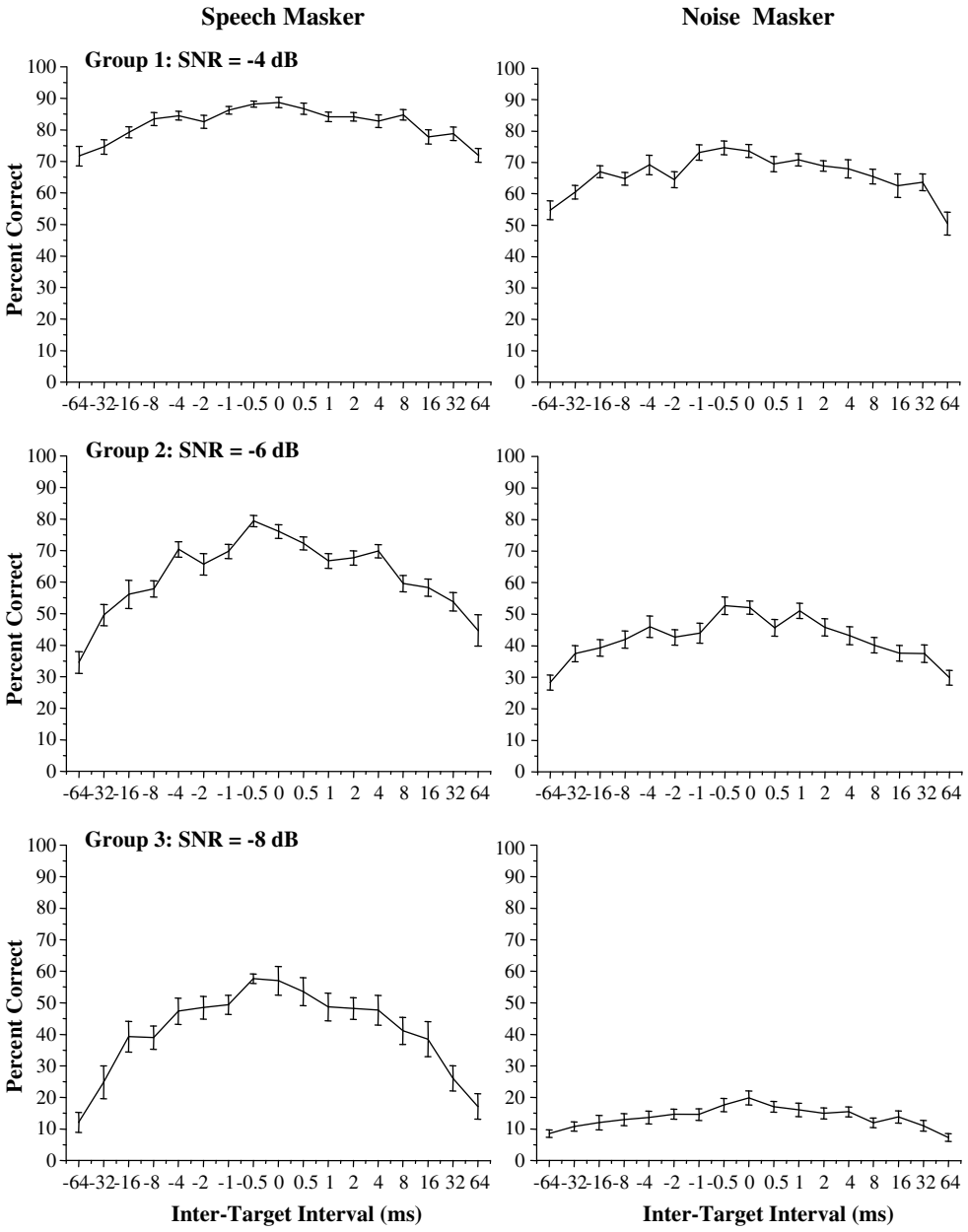

Fig. 6. Percent-correct recognition of keyword syllables as a function of the ITI for the three younger-participant groups when the masker was speech (left panels) or speech-spectrum noise (right panels) under two-source-presentation conditions. Positive ITI values indicate conditions in which the left loudspeaker led the right loudspeaker for target presentation, while negative ITI values indicate conditions in which the right-loudspeaker led the left loudspeaker for target presentation. The error bars represent the standard errors of the mean.

from 0 to 32 ms ($p < 0.010$), and the ITI effect was significant for both speech masking ($F_{8,88} = 58.945$, $p < 0.001$) and noise masking ($F_{8,88} = 17.617$, $p < 0.001$). Follow-up *t*-tests revealed that at the level of 0.00625 (with a Bonferroni adjustment) the release from speech masking was significant when the ITI was 32 ms or shorter, and the release from noise masking was significant when the ITI was 32 ms or shorter.

3.4. Effect of changing the ITI on releasing speech from masking in older participants

The bottom panel of Fig. 5 shows percent-correct recognition of keyword syllables under the single-source-presentation condition for the three older-participant groups when the masker was speech (black columns) or speech-spectrum noise (hatched columns). Obviously, target-speech recognition was different across the three groups under either the speech-masking condition or the noise-masking condition.

Under the single-source-presentation condition, as the SNR was reduced from −2 dB (Group 1) to −4 dB (Group 2), and to −6 dB (Group 3), the percent-correct recognition decreased. One-way between-subject ANOVAs show that the SNR effect on recognition of single-source target speech was significant for both the speech-masking condition ($F_{2,33} = 15.716$, $p < 0.001$) and the noise-masking condition ($F_{2,33} = 49.347$, $p < 0.001$). Interestingly, in each of the older-participant groups under the single-source-presentation condition, there was no significant difference in speech recognition between the speech-masking condition and the noise-masking condition (the group with the SNR of −2 dB: $F_{1,11} = 0.378$, $p = 0.551$; the group with the SNR of −4 dB: $F_{1,11} = 0.104$,

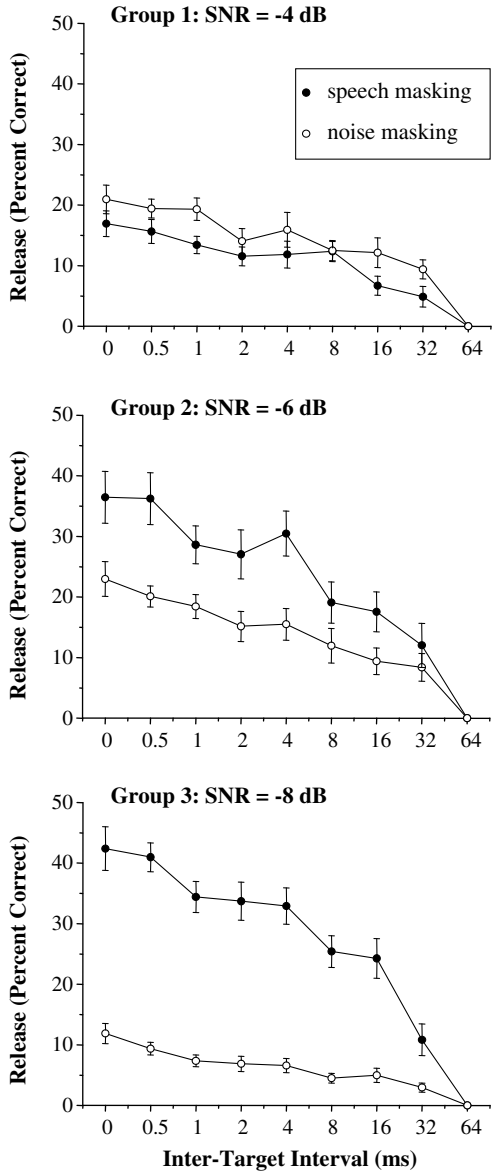

Fig. 7. The release of target speech as a function of the absolute value of ITI for each of the three younger-participant groups under the speech-masking condition (closed circles) and under the noise-masking condition (open circles). Note that data for the two leading directions were combined at each absolute value of ITI. The release of speech from masking at each absolute value of ITI is defined as the difference between the percent speech recognition at that particular ITI and the percent recognition at the ITI of 64 ms. The error bars represent the standard errors of the mean.

$p = 0.754$; the group with the SNR of −6 dB: $F_{1,11} = 1.587$, $p = 0.234$).

Percent-correct recognition for older participants was lower than that for younger participants. Under the single-source speech-masking condition, when the SNR was −6 dB, the mean percent-correct recognition was 13.0% for Older Group 3, which was smaller than that (41.6%) for Younger Group 2. Under the single-source noise-masking condition, the percent-correct recognition was about 10.0% for Older Group 3, which was also lower than that (31.8%) for Younger Group 2. One-way between-subject ANOVAs show that the difference in speech recognition between Older Group 3 and the Younger Group 2 was significant under both the single-source speech-masking condition ($F_{1,22} = 40.934$, $p < 0.001$) and the single-source noise-masking condition ($F_{1,22} = 46.011$, $p < 0.001$).

When the SNR was −4 dB, under the single-source speech-masking condition the mean percent-correct recognition was 33.6% for Older Group 2, which was lower than that (73.0%) for Younger Group 1. Under the single-source noise-masking condition, the percent-correct recognition was 32.3% for Older Group 2, which was also lower than that (58.8%) for Younger Group 1. One-way between-subject ANOVAs show that the difference in speech recognition between Older Group 2 and Younger Group 1 was significant under both the single-source speech-masking condition ($F_{1,22} = 41.608$, $p < 0.001$) and the single-source noise-masking condition ($F_{1,22} = 62.826$, $p < 0.001$).

Fig. 8 shows percent-correct recognition of keyword syllables as a function of the ITI under two-source-presentation conditions for the older-participant groups when the masker was speech (left panels) or noise (right panels). Similar to target-speech recognition in younger groups, target-speech recognition in older groups was also influenced by SNR, masker type, and ITI. Generally, with the increase of SNR from −6 to −2 dB, speech recognition improved across the three younger groups under either the speech-masking condition or the noise-masking condition. Also, with the reduction of the ITI, percent-correct recognition increased progressively, and the improvement was larger under speech-masking conditions than under noise-masking conditions. A mixed 3 (SNR group) by 2 (masker type) by 15 (ITI) ANOVA shows that the main effect of SNR group was significant ($F_{2,33} = 47.365$, $p < 0.001$), the main effect of masker type was significant ($F_{1,33} = 32.963$, $p < 0.001$), and the main effect of ITI was significant ($F_{14,462} = 63.612$, $p < 0.001$). The three-way interaction among the three factors was not significant (SNR group by masker type by ITI: $F_{28,462} = 1.121$, $p = 0.308$), and the interaction between SNR group and masker type was not significant ($F_{2,33} = 1.725$, $p = 0.194$). However, the interaction between SNR group and ITI was significant ($F_{28,462} = 1.824$, $p = 0.007$) and the interaction between masker type and ITI was significant ($F_{14,462} = 17.997$, $p < 0.001$).

Also as shown in Fig. 8, percent-correct recognition under the left-loudspeaker leading condition and that under the right-loudspeaker leading condition was comparable. One-way within-subject ANOVAs and Post hoc analyses show that at each of the ITIs there was no significant difference between the two leading directions for both the noise-masking condition and the speech-masking condition. Thus data for the two leading directions were also combined for analyzing the ITI-induced releasing effect. The ITI-induced release of speech from masking for older-participant groups is defined in the same way as that for younger-participant groups.

Based on the averaged left-right percent-correct recognition, Fig. 9 shows the release of target speech as a function of the absolute value of ITI for each of the three older-participant groups under the speech-masking condition (filled circles) and the noise-masking condition (open circles). The release obviously increased with the decrease of the absolute value of ITI, but larger releases occurred under the speech-masking condition. However, the release was less influenced by the SNR.

For Older Group 1 (SNR = −2 dB), a two-way within-subject ANOVA indicates that the interaction between masker type and ITI was significant ($F_{7,77} = 4.132$, $p = 0.001$), the main effect of masker type was not significant ($F_{1,11} = 3.473$, $p = 0.089$), and the main effect of ITI was significant ($F_{7,77} = 27.829$, $p < 0.001$). Separate one-way within-subject ANOVAs show that the release from speech masking was significantly larger than that from noise masking only at the ITI of 4 ms ($p = 0.004$). Thus the ITI-induced release of speech from speech masking and that from noise masking were generally similar. Also, the ITI effect was significant for both speech masking ($F_{7,77} = 21.610$, $p < 0.001$) and noise masking ($F_{7,77} = 11.211$, $p < 0.001$). Follow-up t-tests revealed that at the level of 0.00714 (0.05/7, with a Bonferroni adjustment) the release from speech masking was significant when the ITI was 16 ms or shorter, and

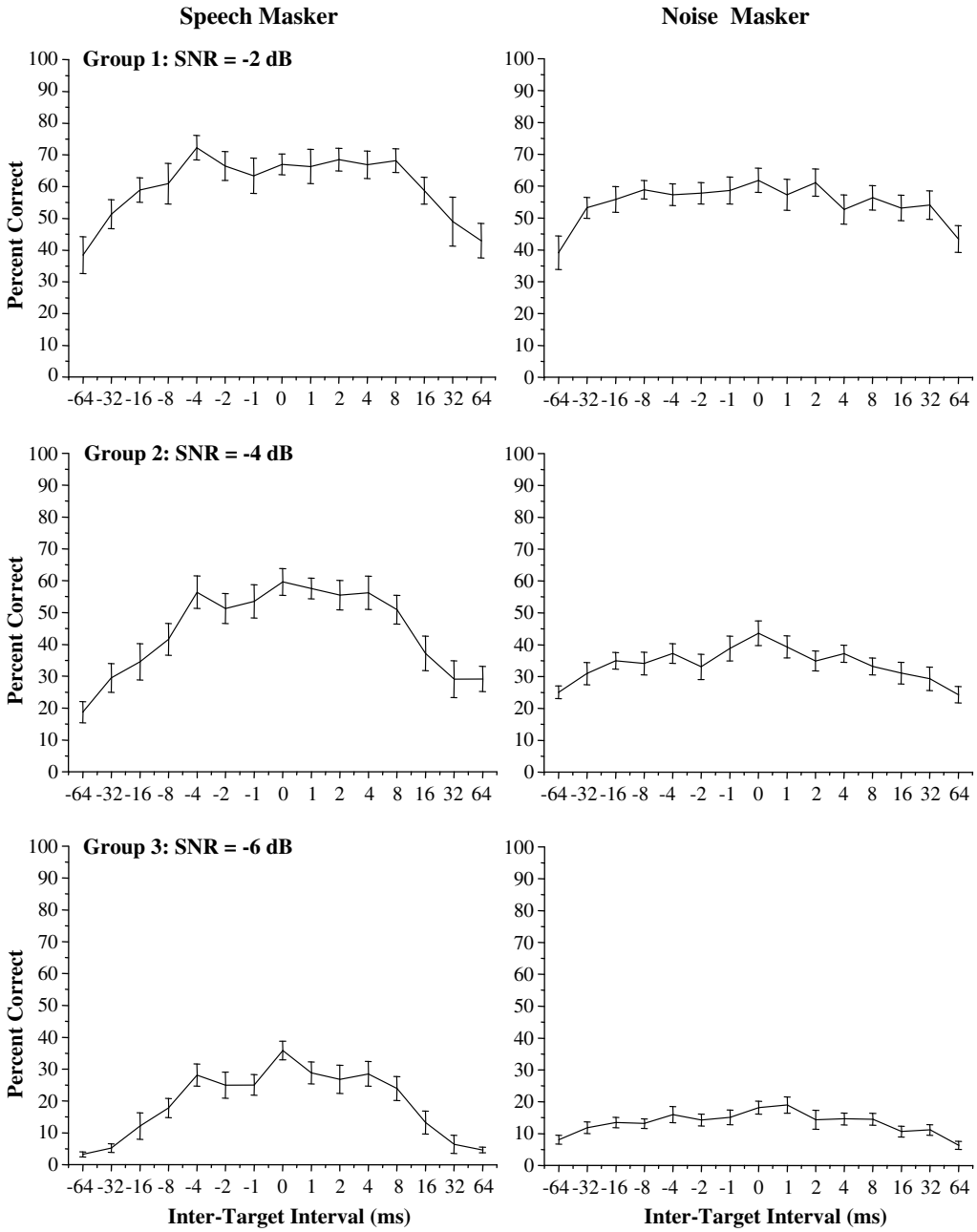

Fig. 8. Percent-correct recognition of keyword syllables as a function of the ITI for the three older-participant groups when the masker was speech (left panels) or speech-spectrum noise (right panels) under two-source-presentation conditions. Positive ITI values indicate conditions in which the left loudspeaker led the right loudspeaker for target presentation, while negative ITI values indicate conditions in which the right-loudspeaker led the left loudspeaker for target presentation. The error bars represent the standard errors of the mean.

the release from noise masking was significant when the ITI was 32 ms or shorter.

For Older Group 2 (SNR = −4 dB), the ITI-induced release was markedly larger under the speech-masking condition than that under the noise-masking condition. A two-way within-subject ANOVA shows that the interaction between masker type and ITI was significant ($F_{7,77} = 9.193$, $p < 0.001$), the main effect of masker type was significant ($F_{1,11} = 23.656$, $p < 0.001$), and the main effect of ITI was significant ($F_{7,77} = 38.390$, $p < 0.001$). Separate one-way within-subject ANOVAs show that the release from speech masking was significantly larger than that from noise masking at ITIs from 0 to 8 ms ($p < 0.010$), and the ITI effect was significant for both speech masking ($F_{7,77} = 28.085$, $p < 0.001$) and noise masking ($F_{7,77} = 14.535$, $p < 0.001$). Follow-up *t*-tests revealed that at the level of 0.00714 (with a Bonferroni adjustment) the release from speech masking was significant when the ITI was 8 ms or shorter, and the release from noise masking was significant when the ITI was 16 ms or shorter.

For Older Group 3 (SNR = −6 dB), the ITI-induced release was also markedly larger under the speech-masking condition than that under the noise-masking condition. A two-way within-subject ANOVA shows that the interaction between masker type and ITI was significant ($F_{7,77} = 21.428$, $p < 0.001$), the main effect of masker type was significant ($F_{1,11} = 34.122$, $p < 0.001$), and the main effect of ITI was significant ($F_{7,77} = 47.174$, $p < 0.001$). Separate one-way within-subject ANOVAs show that the release from speech masking was significantly larger than that from noise masking at ITIs from 0 to 8 ms ($p < 0.010$), and the ITI effect was significant for

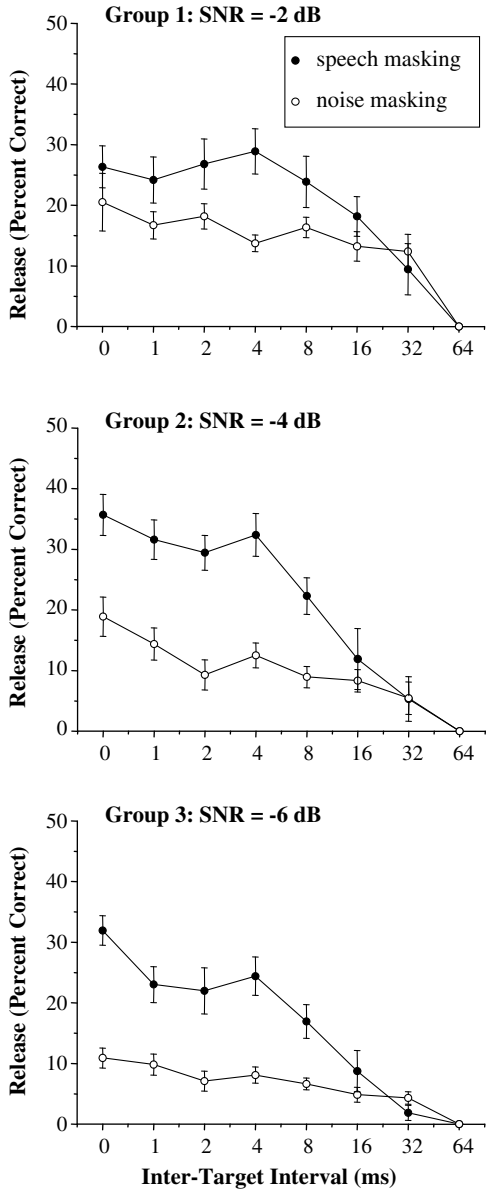

Fig. 9. The percent release of target speech as a function of the absolute value of ITI for each of the three older-participant groups under the speech-masking condition (closed circles) and under the noise-masking condition (open circles). Note that data for the two leading directions were combined at each absolute value of ITI. The release of speech from masking at each absolute value of ITI is defined as the difference between the percent speech recognition at that particular ITI and the percent recognition at the ITI of 64 ms. The error bars represent the standard errors of the mean.

both speech masking ($F_{7,77} = 43.287$, $p < 0.001$) and noise masking ($F_{7,77} = 13.404$, $p < 0.001$). Follow-up t-tests revealed that at the level of 0.00714 (with a Bonferroni adjustment) the release from speech masking was significant when the ITI was 8 ms or shorter, and the release from noise masking was significant when the ITI was 32 ms or shorter.

4. Discussion

4.1. The ITI affects the perceptual integration of target-speech signals

The results of the present study show that when target speech was presented by two spatially separated loudspeakers, perception of the target speech was strongly modulated by the ITI in both younger participants and older participants. Generally, with the reduction of the ITI from 64 to 0 ms (left or right-loudspeaker leading), the target-speech percept changed from two spatially separated images to one diffused image located in the semi-field with the leading loudspeaker, one less diffused image located in the semi-field with the leading loudspeaker, and one image as coming from the frontal field. These changes in perception indicate that the strength of perceptual integration between the two speech-sound waves is progressively enhanced with the reduction of the ITI.

Age-related effects on the precedence effect for long-duration speech have not been reported in the literature (see Lister and Roberts, 2005; Roberts and Lister, 2004; Roberts et al., 2002; Schneider et al., 1994). Using 4 ms bursts of white noise as stimuli, Roberts et al. (2002) reported that listeners with hearing loss (mean age = 68 years) had longer but not shorter echo thresholds than listeners with normal hearing (mean age = 29 years). Their later studies (Roberts and Lister, 2004) have shown that for 4 ms bursts of white noise, there was no effect of aging or hearing loss on the echo threshold under the dichotic or anechoic condition. However, under the reverberant condition, older adults with normal-hearing sensitivity (ONH) exhibited the highest thresholds, followed by those for younger adults with normal-hearing sensitivity (YNH) and older adults with impaired-hearing sensitivity (OIH). The mean echo thresholds for the ONH group were significantly higher than those of the OIH group, but the thresholds of the YNH group were not significantly different from those of the ONH group or the OIH group for the reverberant condition, showing no aging effects. Using 1/4-octave-wide noise with the center frequency of 1000, 2000, or 3000 Hz and the duration between 250 and 350 ms, Lister and Roberts (2005) did not find any significant effects of aging or hearing loss on the echo threshold. In this study, when the ITI was 32 ms, most of the younger participants reported that they perceived two speech images, but most of the older participants reported that they perceived one broad speech image. Moreover, with the ITD was decreased from 8 to 0 ms, the number of younger participants reporting one single broad image reduced but the number of younger participants reporting one single compact speech image increased. However, the number of older participants reporting either type of single speech image did not markedly change. Further studies are needed to clarify whether age-related changes in hearing sensitivity and/or temporal-resolution affect the echo threshold and image compactness for speech sounds presented under free-field conditions.

In both younger and older participants used in this study, when no masking stimuli were presented, although the change of the ITI in the range of 0–64 ms (either left or eight loudspeaker led) markedly modulated the speech image, it had no effects on target-speech recognition. The results suggest that in a reverberant environment the effect of perceptually integrating target speech with its reflections on speech recognition lead to certain perceptually compensative effects of overcoming any potential distortive influence from the reflections (e.g., Watkins, 2005). However, it should also be noted that in this study, target speech was set at a sufficiently high level which allowed participants to be able to recognize speech almost perfectly in quiet. It is not clear whether the change of the ITI can affect speech recognition in quiet when the target-speech level is relatively lower.

4.2. Target-reflection integration releases speech from masking

In this study, the manipulation of the ITI led to a marked change not only in the strength of the perceptual integration of target signals but also target-speech recognition under either speech- or noise-masking conditions. The target-reflection integration must cause certain perceptual differences (e.g., in spatial location,

compactness, and/or loudness) between the target image and the background masker to help participants selectively attend to the target signals, leading to a release of the target speech from the masker. Since maskers delivered from the two loudspeakers were not perceptually fused, the set up used in the present study is particularly useful for studying the function of the perceptual integration of target signals.

When the SNR was at the level of −4 dB for younger participants and −2 dB for older participants, reducing the ITI (absolute value) from 64 to 0 ms led to an equivalent improvement of target-speech recognition for both the speech-masking condition and the noise-masking condition, suggesting a ceiling effect.

When the SNR was reduced to −6 or −8 dB for younger participants and to −4 or −6 dB for older participants, the improvement of speech recognition with the reduction of the ITI was markedly larger under the speech-masking condition than under the noise-masking condition. Thus when the masker is speech and the SNR is sufficiently low, the perceptual integration of target speech and target-speech reflection particularly releases speech from the informational component of speech masking in both younger listeners with normal hearing and older listeners in the early stages of presbycusis.

Note that a reduction of the ITI did not substantially increase the long-term average power of the target-speech signal. To estimate the effect of changing the ITI on acoustics in the ear canal, a KEMAR was located at the position of a participant in the anechoic chamber. Sound waves delivered by two loudspeakers were recorded using the KEMAR. Sound stimuli included (1) white noise (which was not used in the present psychophysical study), (2) speech-spectrum noise, and (3) target speech. Fig. 10 shows the

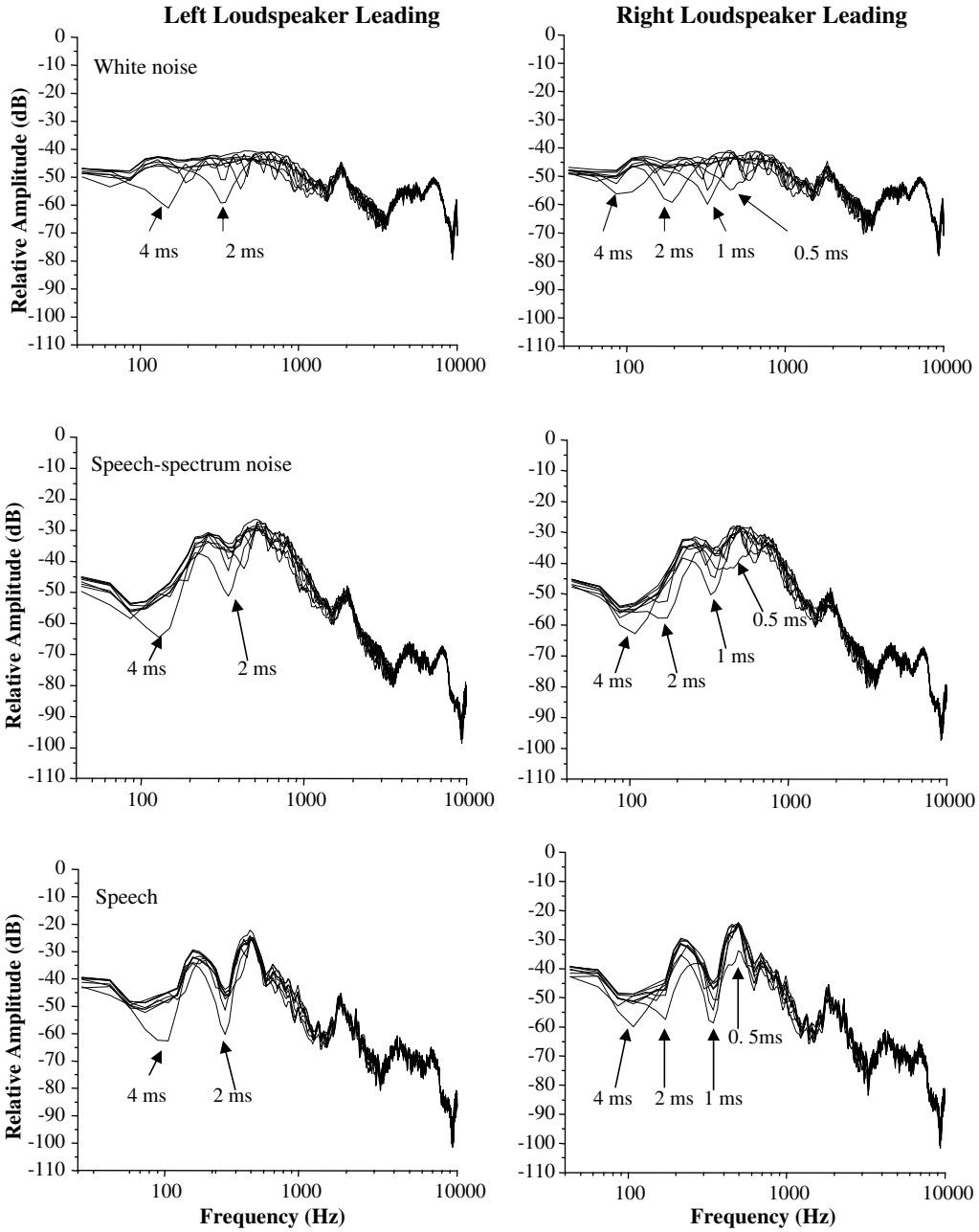

Fig. 10. The spectra of white noise (top panels), speech-spectrum noise (middle panels), and target speech (bottom panels), which were recorded at the right ear of the Knowles Electronic Manikin for Acoustic Research (KEMAR) placed at the participant's position in an anechoic chamber. Each curve in each panel represents the spectrum of one of the 3 types of sounds (white noise, speech-spectrum noise, speech) at a particular inter-loudspeaker delay between two identical stimuli.

spectrum shapes of white noise (top panels), steady–state speech-spectrum noise (middle panels), and target speech (bottom panels) recorded at the right ear of the KEMAR. In each panel, each curve represents the spectrum shape of the testing sound (white noise, speech-spectrum noise, or speech) at a particular inter-loudspeaker delay in the range between 0 and 64 ms.

As shown in Fig. 10, for each of the three types of stimuli, the change in inter-loudspeaker interval did not cause marked changes in the shape of the spectrum curve. However, due to the comb filtering effect (Narins et al., 1979) and head-related transfer functions induced by filtering effects of the head, pinnae, and torso on sound waves, energy for certain frequencies were modified, particularly at some short inter-loudspeaker intervals. For example, energy at certain low frequencies dropped down at some short ITIs. Also, with the change in the inter-loudspeaker interval, no substantial changes were found in the modulation frequency of the speech-waveform envelope. Thus the facilitating effect of reducing the ITI on target-speech recognition cannot be exclusively explained by an increase of the SNR at the ear, and higher-order central processing must be involved (also see Rakerd et al., 2006).

Previous studies have suggested that under the free-field stimulation condition, two of the features of the precedence effect, localization dominance and echo threshold, are weakened by the addition of a broadband background noise (Chiang and Freyman, 1998; Leakey and Cherry, 1957). However, the results of the present study show that the release of target speech from speech masking did not decline with the increase of the masker level, even though the release of target speech from noise masking was lowest when the SNR was at the lowest level used in the present study (−8 dB for younger participants; −6 dB for older participants). It is speculated that although a masker can affect the echo threshold, it cannot eliminate the perceptual integration of a speech sound with the reflection of the speech sound. Particularly, when a speech masker is present, the perceptual integration of the target speech with the target-speech reflection can still increase perceived differences between the target speech and the masking speech, leading to a release of target speech from masking.

4.3. Perceptual integration of target signals vs. perceptual integration of masker signals

Previous studies have shown that in a simulated reverberant environment with multiple-voice talking, the central auditory system not only integrates the target with the target-reflection simulation but also integrates the masker with the masker-reflection simulation at the same time (Freyman et al., 1999; Li et al., 2004; Wu et al., 2005, 2007). However, because listeners normally try to selectively attend to the target and ignore the masker, the target-reflection integration should be more important for parsing the auditory scene. In two previous studies (Brungart et al., 2005; Rakerd et al., 2006), a significant release occurred across a broad range of the inter-masker intervals (up to 32 ms) when the masker was speech, but a significant release occurred only at a few short inter-masker intervals (less than 4 ms) when the masker was noise. In the present study, when the masker was speech, a significant release occurred when the ITI was 32 ms or shorter in younger participants. For the same participants, when the masker was noise, a significant release occurred when the ITI was 0–32 ms. Thus the temporal dynamic pattern of the release from noise masking is different between the present study and the two previous studies (Brungart et al., 2005; Rakerd et al., 2006). In addition to differences in language (Chinese vs. English), pattern of speech material (open-set nonsense sentences vs. close-set coordinate response measure speech sentences), loudspeaker location (symmetrical arrangement vs. asymmetrical arrangement) and distance (200 cm vs. 150 cm or 100 cm), the difference between the results of this study and those of Rakerd et al. and Brungart et al. may be particularly due to differences in signal manipulation (i.e., manipulation of ITI vs. manipulation of inter-masker interval) and perceptual integration (target integration vs. masker integration). Obviously, in future studies, it will be necessary to investigate the co-operation of target-signal integration and masker-signal integration in releasing from masking for target-speech recognition.

4.4. Age group differences in the release of target speech from masking

The results of this study show that under the single-source-presentation condition, when the SNR was the same for older participants and younger participants, percent-correct recognition of target speech in older participants was significantly lower than that in younger participants under either the speech-masking condition or the noise-masking condition. The poorer hearing sensitivity in older participants might be account for the reduction in speech recognition. The results are consistent with the view that older participants were more vulnerable to masking than younger participants (for a review see Schneider et al., 2002).

When the masker was speech, due to the amplitude fluctuation in the speech masker, listeners can take the advantage of large troughs of the temporal envelope (i.e., long-duration drops of energy) in the speech masker to listen to the target speech (Bronkhorst and Plomp, 1992; Howardjones and Rosen, 1993; Gustafsson and Arlinger, 1994; Nelson et al., 2003; Summers and Molis, 2004). In the present study, a shift from the noise masker to the speech masker under the single-source-presentation condition improved speech recognition in younger participants but not in older participants. In each of the older-participant groups the noise masker caused a similar masking effect as the speech masker under the single-source-presentation condition. Thus age-related reduction of temporal processing (for a recent review, see Wingfield et al., 2005) may limit the "listening-through-gap" advantage in older listeners.

Although older participants were more vulnerable than younger participants to noise (energetic) masking, they still could take the advantage of the reduction of the ITI to improve their speech recognition under either the speech-masking condition or the noise-masking condition, indicating that the function of the source-reflection integration is still available in older listeners. Also, similar to that in younger participants, when the SNR was relatively low, with the reduction of the ITI, the target-speech intelligibility in older participants improved more markedly and reached higher levels under the speech-masking condition than under the noise-masking condition. These results support the views that older listeners are able to use certain perceptual cues for releasing speech from masking.

When the listening environment is reverberant, some of the perceptual cues for facilitating detection/discrimination of the target are limited or even abolished by reflections of sound waves (Freyman et al., 1999; Kidd et al., 2005; Koehnke and Besing, 1996; Zurek et al., 2004). Thus, the speech-recognition difficulties induced by interfering stimuli are substantially greater in reverberant environments, suggesting that central auditory operations are critical for segregating target speech from maskers in such environments. Results of the present study suggest that when the salience of a cue is reduced, the aging effect becomes apparent. Fig. 11 summarizes the longest ITI at which *t*-tests (with a Bonferroni adjustment) indicate that the release from speech masking or noise masking was significant for each of the six participant groups used in this study. As indicated in Fig. 11, for the older group with the SNR of −4 dB and the older group with the SNR of −6 dB, the release (relative to the performance at the ITI of 64 ms) was significant only when the ITI was 8 ms or shorter under the

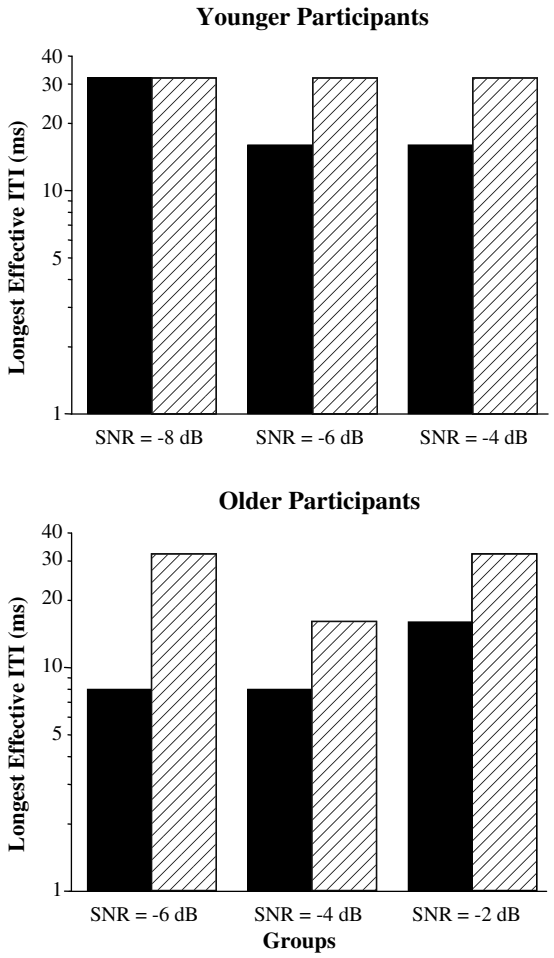

Fig. 11. Summary of the longest effective ITI, which caused significant release of target speech from making, at each of the stimulus conditions. Black columns, speech masking; hatched columns, noise masking.

speech-masking condition. However, for the younger group with the SNR of 6 dB, the release from speech masking was significant when the ITI was 16 ms or shorter, and for the younger group with the SNR of −8 dB, the release was significant when the ITI was 32 ms or shorter.

The results indicate that the ITI range in which significant release of target speech from speech masking is significantly longer in younger listeners than in older listeners. Moreover, when the SNR was low (−6 or −8 dB for younger participants; −4 or −6 dB for older participants), the release of target speech from speech masking was significantly larger than that from noise masking over a broader range of ITI in younger participants (0–16 or 32 ms) than in older participants (0–8 ms). This age-related difference in the unmasking function of perceptual integration at long target-reflection delays may be related to the speech-recognition difficulties experienced by older listeners in noisy, multi-talker, reverberant environments.

To examine whether the speech-recognition performance under either the single-source-presentation condition or each of the two-source-presentation conditions was related to audiometric thresholds, the pure-tone thresholds were assigned into two categories: low frequency (125, 250, 500, and 1000 Hz) and high frequency (2000, 4000, 6000, and 8000 Hz). Mean pure-tone thresholds across the two ears were then calculated for the low-frequency pure-tone average (LFPTA) and the high-frequency pure-tone average (HFPTA). For both younger participants and older participants, speech recognition at any conditions was not significantly correlated with either LFPTA or HFPTA. Although the lack of correlation suggests that age effects on monaural hearing may not be the main reason for causing the age difference in speech recognition under masking conditions, the role of the age-related hearing loss should not be completely ruled out.

It should be noted that according to the study by Chiang and Freyman (1998), noise-induced reductions in the echo threshold are not due to the lowering of sensation level. Thus the reduced functional range of the ITI in older participants may not be associated with the hearing loss at high frequencies. On the other hand, Roberts et al. (2002) have shown that listeners with impaired hearing have higher echo thresholds than for the listeners with normal hearing, and for both listeners with impaired hearing and listeners with normal hearing, echo thresholds at the lower stimulus level (10 dB SPL) are significantly higher than echo thresholds measured at the higher stimulus level (40 dB SPL). Thus, the effects of interplays between aging, hearing loss, and noise masking on the precedence effect need further investigations in the future.

5. Summary and conclusions

(1) When either the speech masker or noise masker is present, with the reduction of the absolute ITI value from 64 to 0 ms, recognition of target speech progressively improves in both younger listeners and older listeners, indicating that the advantage of perceptual integration of target signals occurs over a large ITI range.
(2) When the SNR is relatively lower, this improvement in speech recognition is larger under the speech-masking condition than that under the noise-masking condition, indicating that perceptual integration of target signals mainly plays a role in releasing target speech from informational masking.
(3) Although older listeners are more vulnerable than younger listeners to speech masking and noise masking, their ability to use the perceptual cues provided by the reduction of ITI for improving speech recognition is well retained. However, the ITI range for a significant release from informational masking in older listeners is significantly shorter than in younger listeners. The age-related reduction of the effective ITI range may contribute to the speech-recognition difficulties experienced by older listeners under noisy, reverberant environments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (30711120563; 30670704; 60605016; 60535030; 60435010), the National High Technology Research and Development Program of China (2006AA01Z196; 2006AA010103), the Trans-Century Training Program Foundation for the Talents by the State Education Commission, and "985" grants from Peking University.

References

Arbogast, T.L., Mason, C.R., Kidd, G.J.R., 2002. The effect of spatial separation on informational and energetic masking of speech. J. Acoust. Soc. Am. 112, 2086–2098.
Blauert, J., 1997. Spatial Hearing. MIT Press, Cambridge, MA.
Bronkhorst, A.W., Plomp, R., 1992. Effect of multiple speech-like maskers on binaural speech recognition in normal and impaired hearing. J. Acoust. Soc. Am. 92, 3132–3139.
Brungart, D.S., Rabinowitz, W.M., 1999. Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am. 106, 1465–1479.
Brungart, D.S., Durlach, N.I., Rabinowitz, W.M., 1999. Auditory localization of nearby sources. II. Localization of a broadband source. J. Acoust. Soc. Am. 106, 1956–1968.

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 110, 2527–2538.
Brungart, D.S., Simpson, B.D., Freyman, R.L., 2005. Precedence-based speech segregation in a virtual auditory environment. J. Acoust. Soc. Am. 118, 3241–3251.
Chiang, Y.C., Freyman, R.L., 1998. The influence of broadband noise on the precedence effect. J. Acoust. Soc. Am. 104, 3039–3047.
Durlach, N.I., Mason, C.R., Shinn-Cunningham, B.G., Arbogast, T.L., Colburn, H.S., Kidd, G.J.R., 2003. Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J. Acoust. Soc. Am. 114, 368–379.
Freyman, R.L., Helfer, K.S., McCall, D.D., Clifton, R.K., 1999. The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. 106, 3578–3588.
Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. 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. J. Acoust. Soc. Am. 115, 2246–2256.
Gustafsson, H.A., Arlinger, S.D., 1994. Masking of speech by amplitude–modulated noise. J. Acoust. Soc. Am. 95, 518–529.
Helfer, K.S., 1997. Auditory and auditory-visual perception of clear and conversational speech. J. Speech Lang. Hear. Res. 40, 432–443.
Helfer, K.S., Freyman, R.L., 2008. Aging and speech-on-speech masking. Ear Hearing 29, 87–98.
Helfer, K.S., Wilber, L.A., 1990. Hearing-loss, aging, and speech-perception in reverberation and noise. J. Speech Hear. Res. 33, 149–155.
Howardjones, P.A., Rosen, S., 1993. The perception of speech in fluctuating noise. Acustica 78, 258–272.
Kidd, G.J.R., Mason, C.R., Deliwala, P.S., Woods, W.S., Colburn, H.S., 1994. Reducing informational masking by sound segregation. J. Acoust. Soc. Am. 95, 3475–3480.
Kidd, G.J.R., 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. Acust. Acta. Acust. 91, 526–536.
Koehnke, J., Besing, J.M., 1996. A procedure for testing speech intelligibility in a virtual listening environment. Ear Hearing 17, 211–217.
Leakey, D.M., Cherry, E.C., 1957. Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems. J. Acoust. Soc. Am. 29, 284–286.
Li, L., Yue, Q., 2002. Auditory gating processes and binaural inhibition in the inferior colliculus. Hear. Res. 168, 113–124.
Li, L., Daneman, M., Qi, J.G., Schneider, B.A., 2004. Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? J. Exp. Psychol Human 30, 1077–1091.
Li, L., Qi, J.G., He, Y., Alain, C., Schneider, B., 2005. Attribute capture in the precedence effect for long-duration noise sounds. Hear. Res. 202, 235–247.
Lister, J.J., Roberts, R.A., 2005. Effects of age and hearing loss on gap detection and the precedence effect: Narrow-band stimuli. J. Speech Lang. Hear. Res. 48, 482–493.
Litovsky, R.Y., Colburn, H.S., Yost, W.A., Guzman, S.J., 1999. The precedence effect. J. Acoust. Soc. Am. 106, 1633–1654.
Nabelek, A.K., Robinson, P.K., 1982. Monaural and binaural perception in reverberation for listeners of various ages. J. Acoust. Soc. Am. 71, 1242–1248.
Nabelek, A.K., 1988. Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. J. Acoust. Soc. Am. 84, 476–484.
Narins, P.M., Evans, E.F., Pick, G.F., Wilson, J.P., 1979. A comb-filtered noise generator for use in auditory neurophysiological and psychophysical experiments. IEEE Trans. Biomed. Eng. 26, 43–47.
Nelson, P.B., Jin, S.H., Carney, A.E., Nelson, D.A., 2003. Understanding speech in modulated interference. Cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 113, 961–968.
Rakerd, B., Aaronson, N.L., Hartmann, W.M., 2006. Release from speech-on-speech masking by adding a delayed masker at a different location. J. Acoust. Soc. Am. 119, 1597–1605.
Roberts, R.A., Besing, J., Koehnke, J., 2002. Effects of hearing loss on echo thresholds. Ear Hear. 23, 349–357.
Roberts, R.A., Lister, J.J., 2004. Effects of age and hearing loss on gap detection and the precedence effect: Broadband stimuli. J. Speech Lang. Hear. Res. 47, 965–978.
Schneider, B.A., Pichora-Fuller, M.K., Kowalchuk, D., Lamb, M., 1994. Gap detection and the precedence effect in young and old adults. J. Acoust. Soc. Am. 95, 980–991.
Schneider, B.A., Daneman, M., Pichora-Fuller, M.K., 2002. Listening in aging adults: From discourse comprehension to psychoacoustics. Can. J. Exp. Psych. 56, 139–152.
Schneider, B.A., Li, L., Daneman, M., 2007. How noise interferes with speech comprehension in everyday listening situations? J. Am. Acad. Audiol. 18, 578–591.
Summers, V., Molis, M.R., 2004. Speech recognition in fluctuating and continuous maskers: effects of hearing loss and presentation level. J. Speech Lang. Hear. Res. 47, 245–256.
Watkins, A.J., 2005. Perceptual compensation for effects of reverberation in speech identification. J. Acoust. Soc. Am. 118, 249–262.
Wallach, H., Newman, E.B., Rosenzweig, M.R., 1949. The precedence effect in sound localization. Am. J. Psych. 62, 315–336.
Wingfield, A., Tun, P.A., McCoy, S.L., 2005. Hearing loss in older adulthood – What it is and how it interacts with cognitive performance. Curr. Dir. Psych. Sci. 14, 144–148.
Wu, X.H., Wang, C., Chen, J., Qu, H.W., Li, W.R., Wu, Y.H., Schneider, B.A., Li, L., 2005. The effect of perceived spatial separation on informational masking of Chinese speech. Hear. Res. 199, 1–10.
Wu, X.H., Chen, J., Yang, Z.G., Huang, Q., Wang, M., Li, L., 2007. Effect of number of masking talkers on speech-on-speech masking in Chinese. Interspeech, 390–393.
Yang, Z.G., Chen, J., Wu, X.H., Wu, Y.H., Schneider, B.A., Li, L., 2007. The effect of voice cuing on releasing Chinese speech from informational masking. Speech Commun. 49, 892–904.
Zurek, P.M., Freyman, R.L., Balakrishnan, U., 2004. Auditory target detection in reverberation. J. Acoust. Soc. Am. 115, 1609–1620.