Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 54 (2012) 529–542

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Effects of aging on the ability to benefit from prior knowledge of message content in masked speech recognition

Meihong Wu, Huahui Li, Zhiling Hong, Xinchi Xian, Jingyu Li, Xihong Wu, Liang Li *

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

Received 13 June 2011; received in revised form 6 November 2011; accepted 16 November 2011
Available online 1 December 2011

Abstract

Under conditions in the presence of competing talkers, presenting the early part of a target sentence in quiet improves recognition of the last keyword of the sentence. This content-priming effect depends on a working-memory resource holding the information of the early presented part of the target speech (the content prime). Older adults usually exhibit declined working memory and experience more difficulties in speech recognition under "cocktail-party" conditions. This study investigated whether speech masking also affects recall of the content prime and whether the content-priming effect declines in older adults. The results show that in both younger adults and older adults, although the content prime was heard in quiet, recall of keywords in the prime was significantly affected by the signal-to-masker ratio of the target/masker presentation. The vulnerability of prime recall to speech masking was larger in older adults than that in younger adults. Also, the content-priming effect disappeared in older adults, even though older adults are able to use the content prime to determine the target speech in the presence of competing talkers. Thus, a speech masker affects not only recognition but also recall of speech, and there is an age-related decline in both content-priming-based unmasking of the target speech and recall of the prime.

Keywords: Auditory aging; "Cocktail-party" problem; Content priming; Energetic masking; Informational masking; Speech recognition; Working memory

1. Introduction

The cocktail-party problem, "How do we recognize what one person is saying when others are speaking at the same time?" proposed by Cherry (1953), has been an important issue in psychology, neurophysiology, signal processing, and computer engineering for half a century. It reflects humans' remarkable ability to detect, locate, discriminate, and identify individual speech sources in the presence of competing talkers.

To improve their recognition of the target speech in a noisy environment in the presence of competing talkers, listeners use some perceptual/cognitive cues available in the environment to facilitate selective attention to the target speech and/or to suppress influences of competing speech stimuli. When peripheral neural activity elicited by a signal is overwhelmed by that elicited by a masker, leading to a degraded or noisy neural representation of the signal, making it difficult for subsequent cognitive processes to extract the signal, this masker produces energetic masking (Freyman et al., 1999; Kidd et al., 1994, 1998; Leek et al., 1991). However, some of the perceptual/cognitive cues used by listeners do not (substantially) affect energetic masking. These cues include precedence-effect-induced spatial separation between the target image and masker image (Freyman et al., 1999, 2001; Huang et al., 2008, 2009; Li

* Corresponding author at: Department of Psychology, Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871, China. Tel.: +86 10 6275 6804; fax: +86 10 6276 1081.
E-mail address: liangli@pku.edu.cn (L. Li).

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.11.003

et al., 2004; Rakerd et al., 2006; Wu et al., 2005), prior knowledge about where and/or when the target speech will occur (Best et al., 2007, 2008; Kidd et al., 2005), knowledge/familiarity of the target-talker's voice (Brungart et al., 2001; Helfer and Freyman, 2009; Huang et al., 2010; Newman and Evers, 2007; Yang et al., 2007), prior knowledge about the topic of the target sentence (Helfer and Freyman, 2008), and viewing a speaker's movements of the speech articulators (Grant and Seitz, 2000; Helfer and Freyman, 2005; Rosenblum et al., 1996; Rudmann et al., 2003; Sumby and Pollack, 1954; Summerfield, 1979). It appears that many perceptual/cognitive cues, if they facilitate listeners' selective attention on the target speech and ignorance of competing speech, can improve recognition of the target speech against competing speech by reducing informational masking (for the concept of informational masking, see Arbogast et al., 2002; Agus et al., 2009; Freyman et al., 1999; Helfer and Freyman, 2009; Kidd et al., 1994, 1998, 2005; Leek et al., 1991; Schneider et al., 2007).

In addition to the cues described above, prior knowledge (memory) of the early part of a target sentence (i.e., the content prime) improves listeners' recognition of speech in a masker. More specifically, when either a noise masker or speech masker is present, recognition of the last (third) keyword in a three-keyword sentence is improved if the content prime, an early segment of the same sentence (including the first two keywords), is presented in quiet (Ezzatian et al., 2011; Freyman et al., 2004; Yang et al., 2007). Since the target sentences used in these studies are meaningless ("nonsense"), listeners receive no contextual support from the content prime for recognizing the last keyword. Moreover, the priming benefit is much larger when the masker is speech than when the masker is noise (Ezzatian et al., 2011; Freyman et al., 2004; Yang et al., 2007). As suggested by Freyman et al. (2004), the content prime mainly helps listeners focus attention more quickly on the target, thereby facilitating recognition of the last keyword in the target stream against speech informational masking, "which is caused by confusion between the target and masker and/or uncertainty regarding the target" (Helfer and Freyman, 2009).

It should be emphasized that in humans the content-priming effect depends on a memory resource that holds the prime-content information during the target/masker presentation. However, working memory in humans, which is a system for temporary storage and processing of information during the performance of cognitive tasks (Baddeley, 1986), is vulnerable to disruptive influences. Thus, recall of the content prime may be affected by the presentation of the masker, particularly at low signal-to-masker ratios (SMRs). In previous human studies of the content-priming effects (Ezzatian et al., 2011; Freyman et al., 2004; Yang et al., 2007), the accuracy of recalling the prime is not reported. One of the purposes of this study is to investigate whether recall of keywords in the content prime is affected by speech masking.

Older adults often experience difficulties understanding speech under conditions with multiple people talking at the same time (e.g., Agus et al., 2009; Cheesman et al., 1995; Duquesnoy, 1983; Frisina and Frisina, 1997; Gelfand et al., 1988; Helfer and Freyman, 2008, 2009; Helfer and Wilber, 1990; Helfer et al., 2010; Huang et al., 2008, 2010; Humes and Roberts, 1990; Jerger et al., 1991; Rossi-Katz and Arehart, 2009; Schneider et al., 2000; Tun et al., 2002). The age-related difficulties may be due to both age-related bottom-up deficits at the sensory level (including reduced temporal and/or spectral sensitivities) and age-related top-down deficits at the cognitive level (including declines in selective attention, working memory, inhibitory control, and processing pace) (for reviews see Schneider, 1997; Schneider et al., 2007). Particularly, working memory generally declines in older adults (Salthouse, 1991; Verhaeghen et al., 1993). Previous studies have shown that in addition to the peripheral contribution to sound audibility, some cognitive factors such as working memory, attention, inhibitory control, and speed of processing contribute significantly to speech perception, particularly under noisy listening conditions (for reviews see Humes, 2007; Schneider et al., 2007). The age-related declines in cognitive function may also be associated with age-related impairment of speech recognition. Particularly related to this study, the inhibitory–deficit hypothesis (Hasher and Zacks, 1988) suggests that the age-related decline in working memory is a result of a decrease in the ability to inhibit irrelevant information in working memory. Decreased inhibitory mechanisms cannot prevent irrelevant information from both coming into working memory and occupying storage capacity/processing resources, leading to reduced working memory. Thus, because the presentation of the content prime in quiet is immediately followed by the target/masker complex, it is predicted that recall of keywords in the prime is vulnerable to speech masking, particularly for older adults. Also, if there is an age-related deficit in the memorial preservation of the prime signal, the content-priming effect would be reduced in older adults. However, a recent study by Ezzatian et al. (2011) shows that English-speaking older adults are equivalent to their age controls (younger adults) in the amount of benefit they gain from content priming, suggesting that older adults are as capable as younger adults in using the prime to facilitate parsing the auditory scene and recognizing words. This study also investigated whether recall of the prime content is more affected by the speech masker in older adults than in younger adults, and whether there is an age-related reduction of the benefit from content priming in Chinese-speaking old adults.

In previous studies of the content-priming effect (Ezzatian et al., 2011; Freyman et al., 2004; Yang et al., 2007), the content prime was not the only cue for segregating the target speech from competing (masking) speech. In a test trial of these studies, the target sentence was started about 1 s after the onset of the masker, and listeners were instructed to attend to the speech sentence with the delayed

onset and repeat the sentence after all the stimuli terminated. Thus, the masker/target onset delay was heavily used by listeners to reduce the target/masker confusion and quickly determine which stimulus stream was the target among the target/masker complex. If the masker/target onset delay is removed, the content prime will become the only semantic cue helping listeners attend to the target sentence (see Helfer and Freyman, 2009). The present study specifically investigated whether the content prime can be used to determine the target-speech stream in the presence of competing talkers when the onset delay cue is absent and whether an increase of the prime length from four syllables to eight syllables improves recognition of the last keyword in younger adults and older adults.

2. Experiment 1: With the onset delay cue

2.1. Materials and methods

2.1.1. Participants

Twenty-four younger adults (15 females and 9 males, mean age = 24.0 yr between 20 and 27 yr) recruited from Peking University and 12 older adults (7 females and 5 males, mean age = 66.4 yr between 57 and 75 yr) recruited from the local community participated in Experiment 1 of this study. All the participants had symmetrical hearing (no more than a 15-dB difference between the two ears). Younger participants had pure-tone hearing thresholds no more than 25 dB HL between 0.125 and 8 kHz, and the older participants had pure-tone hearing thresholds no more than 35 dB HL between 0.125 and 1 kHz and no more than 65 dB HL between 2 and 8 kHz. Their first language was Mandarin Chinese. The participants gave their written informed consent to participate in the experiment and were paid a modest stipend for their participation.

As Fig. 1 shows, the thresholds of older participants were generally higher than those of younger participants, and the age difference in threshold increased with frequency. Particularly for frequencies of 4, 6, and 8 kHz, the thresholds of older adults exceeded 30 dB HL. Thus, the two groups of participants were different not only in age but also in hearing sensitivity. Although these older adults were clinically normal in hearing for their ages, they were best characterized as being in the early stages of presbycusis.

2.1.2. Apparatus and stimuli

The participant was seated at the center of an anechoic chamber (Beijing CA Acoustics Co. Ltd, Beijing, China), which was 560 cm in length, 400 cm in width, and 193 cm in height. All acoustic signals were digitized at a sampling rate of 22.05 kHz using a 24-bit Creative Sound Blaster PCI128 with a built-in anti-aliasing filter (Creative Technology, Ltd., Singapore) and were edited using Cooledit Pro 2.0, under the control of a computer with a Pentium IV processor (Intel Corporation, Santa Clara, California, USA). The acoustic analog outputs were delivered to a loudspeaker (Dynaudio Acoustics, BM6 A, Dynaudio, Risskov, Denmark) at 0° azimuth and elevation relative to the participant. The loudspeaker height was 106 cm, which was approximately ear level for a seated listener with average body height. The distance between the loudspeaker and the center of the participant's head (which was not fixed) was about 185 cm.

The 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, 2004), Li et al. (2004), and Ezzatian et al. (2011). The sentences have a subject–predicate–object structure and provide no contextual support for recognizing keywords. Each sentence has 12 characters (also 12 syllables) including the subject (first), predicate (second), and object (third) keywords with two characters (syllables) for each. For example, the English translation of one Chinese nonsense sentence is "This polyester will expel that stomach" (the keywords are underlined). The development of the Chinese nonsense sentences was described elsewhere (Yang et al., 2007).

In the present study, a large number of nonsense-sentence stimuli were required. To satisfy this requirement, and to guarantee both high quality and uniformity of the acoustical features of the stimuli, both target and priming speech were recited by three different artificially synthesized young female voices (see below). The speech masker was a 47-s loop of digitally combined continuous recordings of Chinese nonsense sentences (whose keywords did not appear in the target sentences) spoken by two young female talkers (Yang et al., 2007). Thus, during the target/masker presentation, the target sentence was presented in a two-talker background. It has been known that the two-talker masker was most effective in creating informational masking (e.g., Freyman et al., 2004; Wu et al., 2007). Each of the two masking talkers spoke different sentences and the sound pressure levels were the same across their speech sounds for each testing session. There was no regular relation in sentence phase between the two masking talkers' speech streams and the loop was started randomly at a point for a test trial.

Acoustic stimuli were calibrated using a B&K sound level meter (Type 2230) whose microphone was placed at the center position of the participant's head when the participate was absent, using a "slow"/"RMS" meter response. To minimize both floor and ceiling effects in older participants, the level of both prime and target sounds were set at 60 dBA, and the sound pressure level of the masker was adjusted to produce the SMR of −8, −4, 0, or 4 dB (Huang et al., 2008; Ezzatian et al., 2011).

To minimize any potential voice-priming effects on recognition of the Chinese target sentences (Huang et al., 2010; Yang et al., 2007), three different artificially synthesized young female voices were used for reciting both the prime and the target speech. In a trial, the voice reciting

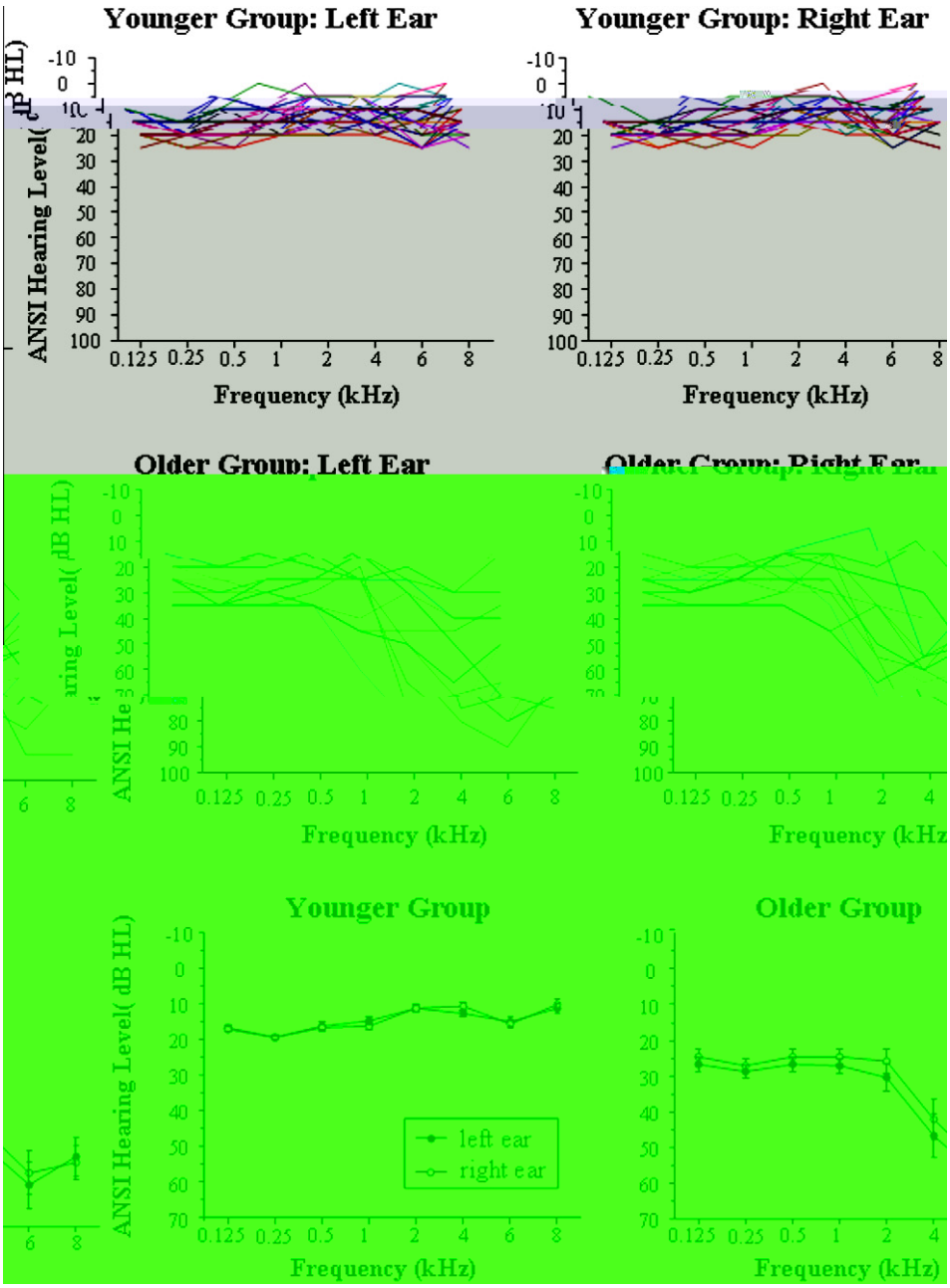

Fig. 1. Top panels: Hearing thresholds in the left ear (left panel) and the right ear (right panel) for individual younger participants who participated in Experiment 1. Middle panels: Hearing thresholds in the left ear (left panel) and the right ear (right panel) for individual older participants who participated in Experiment 1. Bottom panels: Average hearing thresholds in the left ear (filled circles) and the right ear (open circles) for the younger-participant group (left panel) and the older-participant group (right panel). ANSI: American National Standards Institute (S3.6-1989). The error bars represent the standard errors of the mean.

the prime was always different from that reciting the target. Thus, there was uncertainty regarding the prime- and target-talker voices from trial to trial.

Speech synthesis based on the Hidden Markov Model (HMM) has been successfully used for text-to-speech transformation (converting written text into audible speech) (e.g., Yoshimura et al., 1999; Cao et al., 2011). Free software "HTS" is also available (Zen et al., 2007a,b). Since some speech acoustical parameters can be modeled and modulated using the HMM, voice characteristics can also be added into the artificially synthesized speech signals. In this study, acoustic signals of both the target speech and priming speech for the three prime/target voices were generated using an HMM-based system. A Chinese corpus including 6000 sentences with a news-broadcast style, which was both phonetically and prosodically rich, was downloaded from the website (see King and Karaiskos, 2009) and sampled for model training with a sampling rate of 16 kHz. Using the method developed by Zen et al. (2007a,b), some critical parameters of speech

features (including the mel-cepstrum, log F_0, and band aperiodicity measures) were extracted and the five-state left-to-right HMM structure (with no skip) was adopted. Then a five-dimensional multivariate Gaussian distribution was incorporated to model the distribution of the state duration probability. The context-dependent HMMs for each stream were constructed using the decision-tree-based context-clustering method with the minimum-description length (MDL) criterion developed by Shinoda and Watanabe (1997). At the synthesis stage, the speech-parameter sequence for each sentence stimulus was generated from the corresponding HMMs under dynamic feature constraints. Then, using the method developed by Fukada et al. (1992), a speech waveform was synthesized using the algorithm of the Mel Log Spectrum Approximation Filter with the generated parameters. Finally, the initial acoustical model was established by a training procedure using the speech corpus with the voice of a selected female Talker (Talker O).

Speech samples (about 600 sentences and lasting 40 min) of each of the three young female speakers (t_1, t_2, and t_3) were added into the initial acoustic model to obtain the acoustical model for each of the three prime/target voices through the model adaptive procedure. Consequently, for each of the prime/target voices, using the resultant target-voice acoustical models, written nonsense sentences were transformed into signals with the speaker's vocal characteristics. Finally, to equalize speech rates across the three synthesized voices, rate and other temporal information from Talker O were used to modulate the prime/target-voice models of the three prime/target voices, resulting in speech rates that were identical across the different synthesized speech samples.

Six native Chinese-speaking listeners (6 females) with the mean age of 24.6 (between 23 and 26) yr were invited for evaluating the quality of the artificially synthesized voices. They were presented sentences recited separately by six different voices, among which three were artificial ones and another three were natural ones by three young females. The listeners were informed both the number of artificial voices (i.e., 3) and the number of natural voices (i.e., 3), and asked to identify the voice type of each of the speech presentations. Interestingly, without any training the listeners felt it difficult to describe certain artificial traces in the synthesized speech and made some voice-type judgments that did not match the reality. More specifically, at the group level, when the natural-voice sentences were presented, about 27% of their responses were "artificial voice"; when the artificial-voice sentences were presented, about 32% of their responses were "natural voice".

Moreover, another 6 listeners (3 females and 3 males) with the mean age of 21.7 (between 20 and 23) yr were invited for evaluating differences in vulnerability to speech masking between sentences with an artificial voice and sentences with the voice of a young female talker. The results show that under the masking conditions used in the experiments of this study, there were no significant differences in vulnerability to speech masking (e.g., no significant difference in threshold μ, $F(1,5) = 3.507$, $p > 0.05$) between sentences recited by the artificially synthesized voice and sentences recited by the natural voice.

2.1.3. Procedures

For a testing session, which was associated to a particular priming-type/SMR combination, participants were informed of the type of priming condition (priming or no-priming condition). Before the testing, a visually aided presentation of the instructions (using Microsoft Office PowerPoint 2003) to participants was also provided to allow each of the participants become familiar to the experimental procedures as the following. The participant pressed a button on a response box to start a trial.

Under the no-priming (baseline) condition, the stimulus presentation had three temporal stages: (i) The masker started immediately after the button press; (ii) about 1 s (0.8–1.2 s) after the masker onset, a target sentence started; (iii) the masker and target ended simultaneously. Thus, for a session of the no-priming condition, participants were informed that after he/she started a trial by pressing the button, no prime was presented before the presentation of the masker and target. Under the priming condition, the stimulus presentation had four temporal stages: (i) the prime (the target sentence without the last four syllables) was started in quiet immediately after the button press (e.g., if the English translation of a target sentence was "These sailing boats will symbolize that milk", the prime was "These sailing boats will symbolize"); (ii) the masker occurred immediately after the end of the prime presentation; (iii) about 1 s (0.8–1.2 s) after the masker onset, a target sentence started; (iv) the masker and target ended simultaneously. Thus, for a session with the priming condition, participants were informed that after they started a trial by pressing the button, the priming sentence, the target sentence without the last keyword, was presented in quiet, and the masker and then the complete target sentence occurred after the prime presentation. In other words, participants were told that the words they were hearing in quiet were then going to be presented in a background of the masker. The participant's task was to repeat aloud the whole target sentence (which started about 1 s after the masker onset) immediately after the stimuli was ended.

Eighteen target sentences were used for each testing condition. Six target sentences (out of 18 sentences) were recited by each of the three target voices. As described above, the voice reciting the prime was always different from that reciting the target sentence in a trial to avoid any voice-priming effects (Huang et al., 2010; Yang et al., 2007). Performance was scored as the number of correctly identified syllables for each keyword (each keyword contained two syllables). To ensure that all the participants fully understood and correctly followed the experimental instructions, there was one training session before formal testing (Yang et al., 2007).

2.2. Results and discussion

condition and SMR was significant ($F[3,69] = 47.102$, $p < 0.001$), the interaction between keyword position and SMR was significant ($F[6,138] = 14.957$, $p < 0.001$), and the three-way interaction was significant ($F[6,138] = 25.257$, $p < 0.001$). A separate two-way ANOVA shows that under the priming condition, the interaction between keyword position and SMR was significant ($F[6,138] = 42.774$, $p < 0.001$). Further one-way ANOVAs show that the SMR effect was significant for each of the three keyword positions (first keyword: $F[3,69] = 4.357$, $p < 0.01$; second keyword: $F[3,69] = 17.230$, $p < 0.001$; third keyword: $F[3,69] = 169.957$, $p < 0.001$). The results indicate that although younger participants heard the content prime in quiet, their recall of the keywords in the prime (the first and second keywords) following the target/masker presentation was significantly affected by the SMR.

For older participants, a two (priming condition) by three (keyword position) by four (SMR) within-subject ANOVA shows that the interaction between priming condition and keyword position was significant ($F[2,22] = 83.579$, $p < 0.001$), the interaction between priming condition and SMR was not significant ($F[3,33] = 1.024$, $p = 0.394$), the interaction between keyword position and SMR was significant ($F[6,66] = 4.277$, $p < 0.01$), and the three-way interaction was significant ($F[6,66] = 3.801$, $p < 0.01$). A separate two-way ANOVA shows that under the priming condition, the interaction between keyword position and SMR was significant ($F[6,66] = 47.375$, $p < 0.001$). Further one-way ANOVAs show that the SMR effect was significant for each of the three keyword positions (first keyword: $F[3,33] = 11.090$; $p < 0.001$; second keyword: $F[3,33] = 28.132$, $p < 0.001$; third keyword: $F[3,33] = 54.214$, $p < 0.001$). The results indicate that similar to younger participants' performance, older participants' recall of the keywords in the content prime following the target/masker presentation was also significantly affected by the SMR.

2.2.2. Content-priming effects on recognition of the last keyword

Fig. 3 plots group-mean percent-correct syllable identification for the third keyword as a function of the SMR for younger participants (solid curves) and older participants (broken curves) under the no-priming condition (open circles) and the priming condition (filled circles), along with the group-mean best-fitting psychometric functions.

There is evidence in Fig. 3 to suggest that for younger participants the presence of the content prime improved recognition of the third keyword. A two (priming condition) by four (SMR) ANOVA confirms that the main effect of priming condition was significant ($F[1,23] = 56.423$, $p < 0.001$), the main effect of SMR was significant ($F[3,69] = 345.149$, $p < 0.001$), but the interaction of the two factors was not significant ($F[3,69] < 1$).

However, there is no evidence in Fig. 3 to suggest that for older participants the presence of the content prime improved recognition of the third keyword. A two (priming condition) by four (SMR) ANOVA shows that the main effect of priming condition was not significant ($F[1,11] = 1.817$, $p = 0.205$), the main effect of SMR was significant ($F[3,33] = 83.482$, $p < 0.001$), and the interaction of the two factors was not significant ($F[3,33] < 1$).

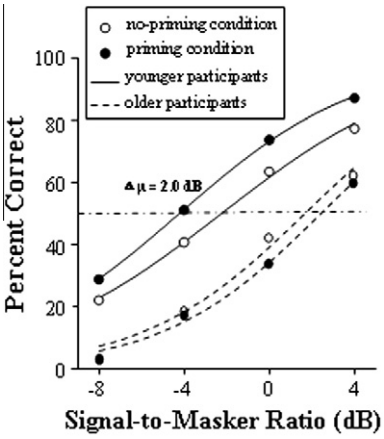

Fig. 3. Group-mean percent-correct syllable identification for the third keyword as a function of SMR in Experiment 1 for younger participants (solid curve) and older participants (broken curves) under the no-priming condition (open circles) and the priming condition (filled circles), along with the group-mean best-fitting psychometric functions.

3. Experiment 2: Without the onset delay cue

In Experiment 2, the masker/target onset delay was removed, and the content prime became the only semantic cue providing participants a means of attending to the target sentence. Thus, Experiment 2 was to investigate whether listeners are still able to determine the target-speech stream in a speech complex with three female voices when the onset-delay cue is absent, and whether an increase of the prime length from four syllables (including the first keyword) to eight syllables (including both the first and second keywords) improves recognition of the last (third) keyword in younger adults and older adults. Since the content prime was the only semantic cue for participants to attend to the target sentence when the onset-delay cue was removed, the no-priming condition, which had neither the semantic cue nor the onset-delay cue, was not used in Experiment 2.

3.1. Materials and methods

3.1.1. Participants

Eighteen younger adults (12 females and 6 males, mean age = 23.1 yr between 20 and 26 yr) recruited from Peking University and 12 older adults (6 females and 6 males, mean age = 67.8 yr between 57 and 74 yr) recruited from the local community participated in Experiment 2. They did not participate in Experiment 1. All the participants had pure-tone hearing thresholds and symmetrical hearing as required in the Experiment 1. The participants also gave their written informed consent to participate in the

experiment and were paid a modest stipend for their participation.

As shown in Fig. 4, the thresholds of older participants were generally higher than those of younger participants, and the age difference in threshold increased with frequency. Particularly for frequencies of 4, 6, and 8 kHz, the thresholds of older adults exceeded 30 dB HL. Thus, the two groups of participants in Experiment 2 were different not only in age but also in hearing sensitivity. Although these older adults were clinically normal in hearing for their age population, they were best characterized as being in the early stages of presbycusis.

3.1.2. Apparatus and stimuli

Both the apparatus and stimuli were the same as used in Experiment 1.

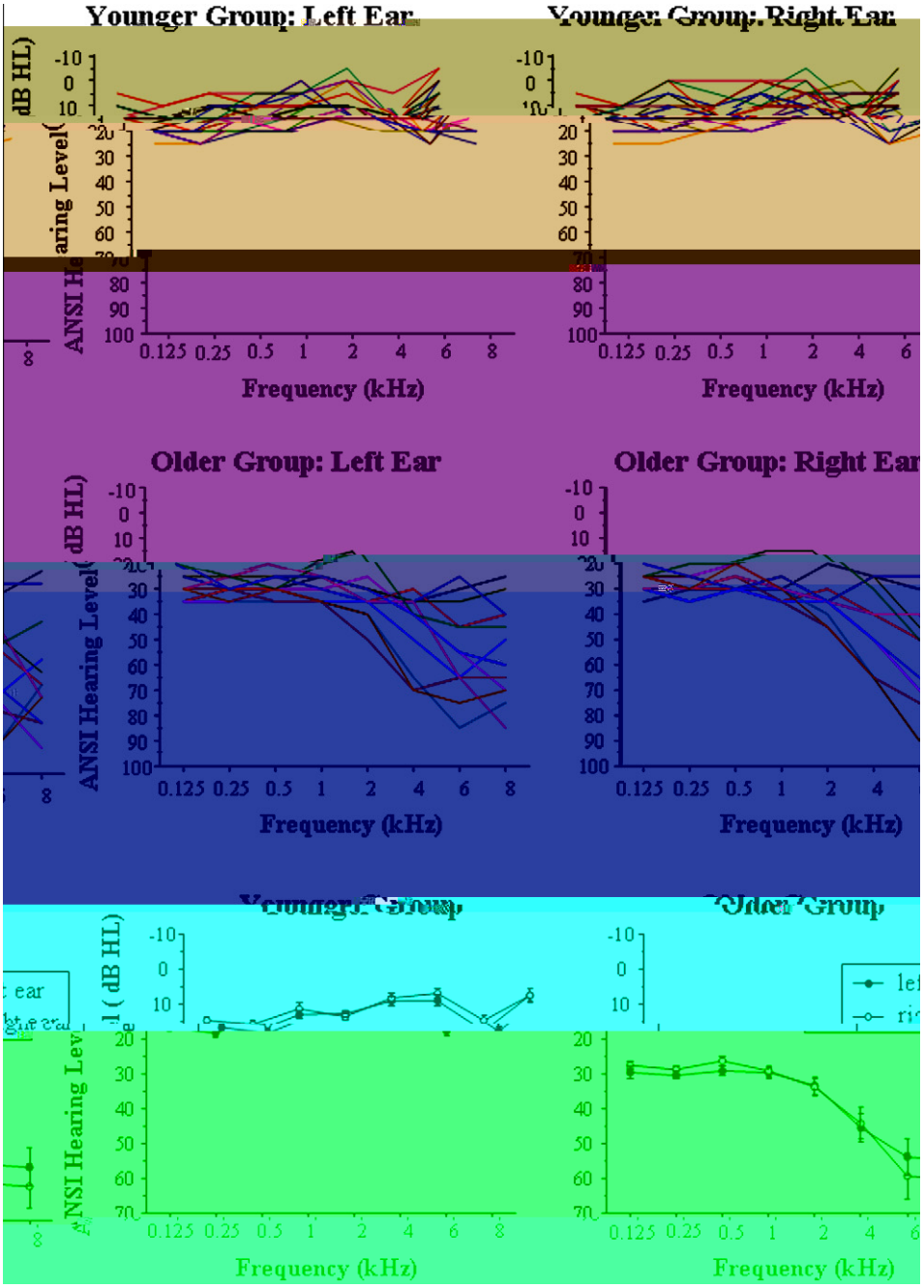

Fig. 4. Top panels: Hearing thresholds in the left ear (left panel) and the right ear (right panel) for individual younger participants who participated in Experiment 2. Middle panels: Hearing thresholds in the left ear (left panel) and the right ear (right panel) for individual older participants who participated in Experiment 2. Bottom panels: Average hearing thresholds in the left ear (filled circles) and the right ear (open circles) for the younger-participant group (left panel) and the older-participant group (right panel). ANSI: American National Standards Institute (S3.6-1989). The error bars represent the standard errors of the mean.

3.1.3. Procedures

In both the younger group and older group, there were 8 conditions (two priming types: 4-syllable (one-keyword) priming, 8-syllable (two-keyword) priming; four SMRs: −8, −4, 0, 4 dB) for each participant, and 18 target sentences were used for each of the conditions. The presentation order for the priming type was counterbalanced across participants in each age group, and the presentation order of the 4 SMRs for each priming type was arranged randomly.

Under a testing condition (with a particular priming-type/SMR combination) containing 18 trials, the participant was informed of the type of priming condition (4-syllable priming or 8-syllable priming). The participant pressed a button on a response box to start a trial. Under the four-syllable-priming condition, the content prime, which contained the first four syllables of the sentence (including the first keyword (with two syllables) and the article or pronoun (with two syllables) before the first keyword), was started in quiet immediately after the button press. Then both the masker and the complete target sentence occurred (without any masker/target onset delay) after the end of prime presentation. Under the eight-syllable-priming condition, the content prime, which was early part of the target sentence with the first eight syllables of the sentence (including both the first and second keywords), was started in quiet immediately after the button press, and both the masker and the complete target sentence occurred (without any masker/target onset delay) after the end of prime presentation. Immediately after all the stimuli ended, participants needed to repeat aloud the whole target sentence.

Among the 18 target sentences that were assigned into a particular combination of priming type and SMR, 6 were produced by each of the three target voices. The voice reciting the prime was always different from that reciting the target sentence in each trial. Performance was scored as the number of correctly identified syllables for each keyword. To ensure that all the participants fully understood and correctly followed the experimental instructions, there was one training session before formal testing (Yang et al., 2007).

3.2. Results and discussion

When no masker was presented, the accuracy for recognizing the keywords was near perfect in younger participants. For older participants, an accuracy of 87% was reached across the three keyword positions.

3.2.1. Recognition of the first keyword

The first keyword occurred in the prime under both the four-syllable-priming condition and the eight-syllable-priming condition. Participants always heard this keyword in quiet before the target/masker presentation. Thus, participants needed to use a memory resource to hold the content of the first keyword against speech masking, and then recall this keyword after the target/masker presentation.

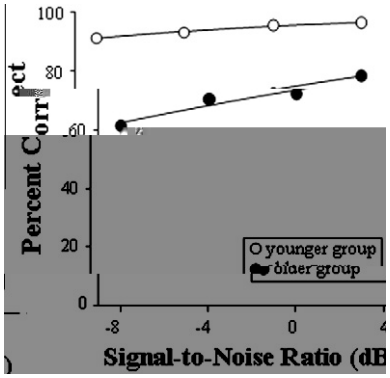

Fig. 5. Averaged group-mean percent-correct syllable identification for the first keyword in Experiment 2 over the two priming conditions as a function of SMR for younger participants (open circles) and older participants (filled circles), along with the group-mean best-fitting psychometric functions. Note that the first keyword was always presented in the prime.

Since there were no significant differences in recognizing the first keyword between the two priming conditions for both younger participants and older participants, percent-correct responses for individual participants were averaged over the two priming conditions. Fig. 5 plots averaged group-mean percent-correct syllable identification for the first keyword over the two priming conditions as a function of SMR for younger participants (open circles) and older participants (filled circles) along with the group-mean best-fitting psychometric functions.

As shown in Fig. 5, recognition of the first keyword was affected by the SMR, and it was markedly poorer in older participants than in younger participants. A two (age group) by four (SMR) mixed two-way ANOVA shows that the main effect of age group was significant ($F[1,28] = 33.254$, $p < 0.001$), the main effect of SMR was significant ($F[3,84] = 15.971$, $p < 0.001$), and the interaction of these two factors was significant ($F[3,84] = 4.408$, $p < 0.01$). Further one-way ANOVAs show that the effect of SMR was significant in both the younger group ($F[3,51] = 6.777$, $p < 0.01$) and older group ($F[3,33] = 7.800$, $p < 0.001$). Also, at each of the SMRs, the difference between the two groups was significant ($p < 0.001$). The results indicate that recognition of the first keyword was significantly worse in older participants than in younger participants, and more importantly, the effect of SMR was significant in both younger participants and older participants. Fig. 5 also shows that the steepness of the psychometric function curve appears to be affected by the age-group factor.

3.2.2. Recognition of the second keyword

The second keyword occurred in the prime under the eight-syllable-priming condition but not the four-syllable-priming condition. Thus, under the four-syllable-priming condition, participants needed to use the first four syllables as the cue to determine the target sentence, recognize the second keyword during the presentation of the target

sentence and masker, and then repeat the whole sentence. Under the eight-syllable-priming condition, after listening to the prime including the second keyword in quiet, participants needed to both use a memory resource to hold the content of the second keyword, recognize this keyword during the target/masker presentation, and then repeat the whole sentence. Fig. 6 plots group-mean percent-correct syllable identification for the second keyword as a function of SMR for younger participants (solid curves) and older participants (broken curves) under either the four-syllable-priming condition (open circles) or the eight-syllable-priming condition (filled circles), along with the group-mean best-fitting psychometric functions.

As shown by Fig. 6, recognition of the second keyword was affected by the SMR, and it was generally poorer in older participants than in younger participants. Obviously, for both younger participants and older participants, repeating the second keyword was much easier under the eight-syllable-priming condition (because this keyword was presented in quiet) than under the four-syllable-priming condition (when this keyword was not presented in the prime). For example, at the SMR of −8 dB, an increase of the prime length from four to eight syllables resulted in an improvement of the accuracy of recognizing the second keyword from 27.2% to 87.0% in the younger group and from 5.3% to 52.6% in the older group. More importantly, even under the eight-syllable-priming condition, recall of the second keyword was also affected by the SMR in both younger participants and older participants.

A two (age group) by two (priming condition) by four (SMR) mixed three-way ANOVA shows that the three main effects were all significant (age group: $F[1,28] = 1232.461$, $p < 0.001$; priming condition: $F[1,28] = 345.424$, $p < 0.001$; SMR: $F[3,84] = 180.261$, $p < 0.001$), the three-way interaction was significant ($F[3,84] = 7.516$, $p < 0.001$), the two-way interaction between SMR and priming condition was significant ($F[3,84] = 63.893$, $p < 0.001$), but both the two-way interaction between age group and SMR and that between age group and priming condition were not significant ($p > 0.070$).

Separate two (age group) by four (SMR) mixed two-way ANOVAs show that under the four-syllable-priming condition, the main effect of SMR was significant ($F[3,84] = 298.748$, $p < 0.001$), the main effect of age group was significant ($F[1,28] = 135.455$, $p < 0.001$), and the interaction of the two factors was also significant ($F[3,84] = 8.957$, $p < 0.001$). Further one-way ANOVAs show that the difference between the two age groups was significant at each of the SMRs (for all, $p < 0.001$). Moreover, under the eight-syllable-priming condition, the main effect of SMR was significant ($F[3,84] = 11.931$, $p < 0.001$), the main effect of age group was significant ($F[1,28] = 48.498$, $p < 0.001$), but the interaction of the two factors was not significant ($F[3,84] = 2.480$, $p = 0.067$). The results indicate that recognition of the keyword was significantly worse in older participants than in younger participants, and more importantly, under the eight-syllable-priming condition, the effect of SMR was significant in both younger participants and older participants.

3.2.3. Content priming effects on recognition of the last (third) keyword

The last (third) keyword did not occur in the prime under either of the two priming conditions. Fig. 7 illustrates group-mean percent-correct syllable identification for the third keyword as a function of SMR for younger participants (solid curves) and older participants (broken curve) under the four-syllable-priming condition (open circles) and the eight-syllable-priming condition (filled

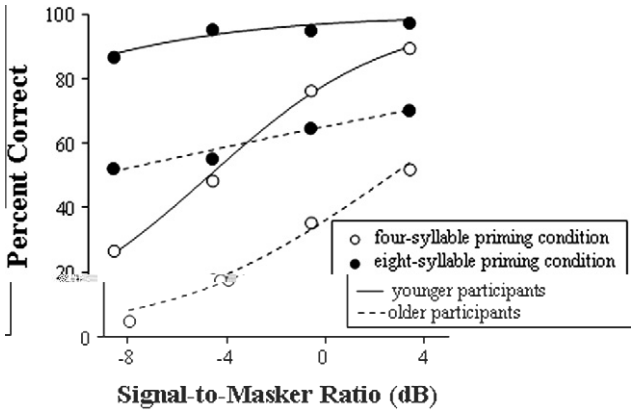

Fig. 6. Group-mean percent-correct syllable identification of the second keyword as a function of the SMR in Experiment 2 in younger participants (solid curves) and older participants (broken curves) under either the four-syllable-priming condition (open circles) or the eight-syllable-priming condition (filled circles). The group-mean best-fitting psychometric functions for each of the priming conditions are shown in each of the panels. Note that under the eight-syllable-priming condition, the second keyword was in the content prime.

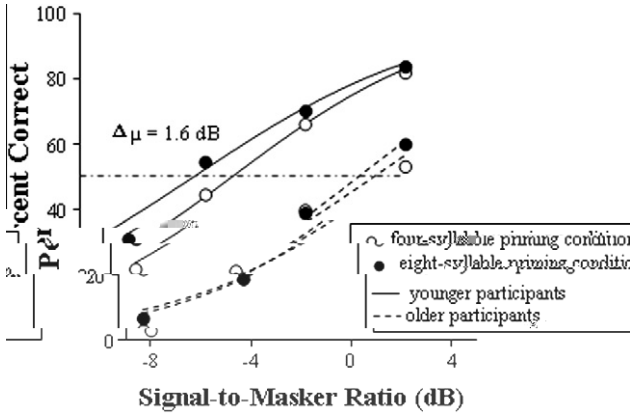

Fig. 7. Group-mean percent-correct syllable identification of the third keyword as a function of the SMR in Experiment 2 in younger participants (solid curves) and older participants (broken curves) under either the four-syllable-priming condition (open circles) or the eight-syllable-priming condition (filled circles). The group-mean best-fitting psychometric functions for each of the priming conditions are shown in each of the panels. Note that only in younger participants, an increase of the prime length from four syllables to eight syllables shifted the threshold (μ) by 1.6 dB.

circles), along with the group-mean best-fitting psychometric functions.

Obviously, in both younger participants and older participants, recognition of the third keyword was markedly affected by the SMR, and reached a level over 50% when the SMR was sufficiently high, indicating that when the onset-delay cue was absent the two age groups were able to use the content prime as the cue to determine and follow the target speech among the target/masker complex. Also, as expected, recognition of the third keyword was generally

the prime-induced release from speech masking must depend on higher-order processes. Note that in this study since the voice speaking the prime and the voice speaking the target were different, the voice-priming effect (see Yang et al., 2007; Huang et al., 2010) was minimized. However, it is possible that the prime-induced attention to the content of the target sentence, which was co-presented with the masker, enabled participants to attend to the target voice, which was the one throughout the target sentence. Thus, it is possible that the content-to-voice shift in cueing the target speech also played a role in target tracking. However, further investigation is needed to address this interesting issue.

The content priming effect is based on working memory holding the prime content during the course of the target/masker presentation. The increased masker-induced disruptions of the working memory in older participants suggest an age-related impairment of the content-priming effect. Indeed, the results of this study show that no significant content-priming effect on recognizing the last keyword was observed in older participants. According to the inhibition–reduction theory (Hasher and Zacks, 1988), the age-related decline in working memory is a result of a decrease in the ability to inhibit disrupting information in working memory. The age-related decrease in inhibitory mechanisms impairs the prevention of disrupting information from entering working memory and occupying storage capacity/processing resources. In addition, the increased vulnerability to irrelevant speech in aged people may be related to source-monitoring deficit that leads to confusion between distracters and targets (Bell et al., 2008). Since the content prime helps participants focus attention more quickly on the target, thereby facilitating recognition of the last keyword against informational masking (Freyman et al., 2004), the loss of the content-prime information makes older listeners more susceptible to informational masking. The present results are not in agreement with those of the Ezzatian et al. study (2011) that shows that English-speaking older adults are equivalent to their age controls (younger adults) in the amount of benefit they gain from content priming. It would be of interest to know why the results of the present study are not in agreement with those of the Ezzatian et al. (2011) study. First, in younger adults, the prime-induced improvement in speech recognition against speech masking was apparently larger in both the Ezzatian et al. (2011) study and the Freyman et al. (2004) study (using English-speaking listeners) than that in both the Yang et al. (2007) study and the present study (using Chinese-speaking listeners). Thus, the content prime may be more effective in reducing informational masking for English-speaking listeners than for Chinese-speaking listeners. In addition, in the Ezzatian et al. (2011) study, the prime and the target speech were identical, except that the last keyword in the prime was replaced by a broadband noise burst. Thus, older participants in Ezzatian et al. (2011) study might integrate several cues (e.g., content, voice, and prosody) to facilitate their recognition of the last keyword during the presentation of the target speech and masking speech. In the present study, since the prime and target were recited by different voices, cues other than prime content were minimized. Finally, in the present study, recall of the prime in older participants was markedly vulnerable to speech masking. The masking-induced decline in working memory of the content prime might be accounted for the reduction of the priming effect in older participants. However, whether recall of the prime is also vulnerable to speech masking in English-speaking older adults has not been reported before (Ezzatian et al., 2011; Freyman et al., 2004).Further investigation is needed to clarify whether English-speaking older adults are also equivalent to English-speaking younger adults in the accuracy of recalling keywords in the content prime.

4.3. Using the content prime for determining the target stream

This study focused on the top-down, knowledge-driven process for segregating the target speech from masking speech. In Experiment 2, since the masker/target onset delay was removed and participants were instructed to use the prime for orienting to the target sentence, participants needed to hold the prime content in their mind, expect the prime occurrence in the target/masker complex, and find and follow the target stream. The results of Experiment 2 show that depending on the SMR, both younger participants and older participants, to certain degrees, were able to use the prime as the cue for finding and following the target stream among the target/masker complex. These results are consistent with previous reports that the call-sign type of semantic information is useful for segregating three female-voiced speech streams (Helfer and Freyman, 2009). However, although to a degree older participants maintained the ability to use the semantic information for reducing the target/masker confusion to determine the target stream under the four-syllable (two-keyword) priming condition, they did not benefit from the increase of the length of the content prime for further improving recognition of the last keyword. Also, even at the favorite SMR (4 dB), older participants' performance was still remarkably poorer than younger participants' performance, suggesting that the ability to use the semantic information was also lower in older participants than in younger listeners. In this study, since in a trial the voice speaking the prime was always different from that speaking the target, the potential voice-priming effect (Yang et al., 2007; Huang et al., 2010) was minimized. Note that although participants were instructed to use the prime content to determine the target sentence, the possibility that they might use other cues, such as the loudness difference between target and masking sentences, should not be excluded.

5. Conclusions

The results of this study indicate that the content prime plays a role in unmasking speech from informational mask-

ing. The content prime helps listeners determine the target-speech stream among a multi-people-talking complex. Thus, even when the powerful onset-delay cue is absent, listeners are still able to use the semantic cue to determine and follow the target speech against speech masking. Moreover, recall of the content prime is also affected by speech masking, and the masking effect is SMR dependent. In older adults, not only is recall of the content prime worse and more vulnerable to speech masking than that in younger adults, but also the content-priming effect declines.

Acknowledgments

This work was supported by the "973" National Basic Research Program of China (2009CB320901; 2010DFA31520; 2011CB707805), the National Natural Science Foundation of China (31170985; 30711120563, 90920302, 60811140086), the Chinese Ministry of Education (20090001110050), and "985" grants from Peking University.

References

Agus, T.R., Akeroyd, M.A., Gatehouse, S., Warden, D., 2009. Informational

recognition in younger and older adults? J. Exp. Psychol. Hum. Percept. Perform. 30 (6), 1077–1091.
Newman, R.S., Evers, S., 2007. The effect of talker familiarity on stream segregation. J. Phon. 35, 85–103.
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. Amer. 119 (3), 1597–1605.
Rosenblum, L.D., Johnson, J.A., Saldana, H.M., 1996. Point-light facial displays enhance comprehension of speech in noise. J. Speech Hear. Res. 39, 1159–1170.
Rossi-Katz, J., Arehart, K.H., 2009. Message and talker identification in older adults: effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. J. Speech Lang. Hear. Res. 52, 435–453.
Rudmann, D.S., McCarley, J.S., Kramer, A.F., 2003. Bimodal displays improve speech comprehension in environments with multiple speakers. Hum. Factors 45 (2), 329–336.
Salthouse, T.A., 1991. Mediation of adult age differences in cognition by reductions in working memory and speed of processing. Psychol. Sci. 2 (3), 179–183.
Schneider, B.A., 1997. Psychoacoustics and aging: implications for everyday listening. J. Speech-Lang. Pathol. Audiol. 21, 111–124.
Schneider, B.A., Daneman, M., Murphy, D.R., Kwong See, S., 2000. Listening to discourse in distracting settings: the effects of aging. Psychol. Aging 15 (1), 110–125.
Schneider, B.A., Li, L., Daneman, M., 2007. How competing speech interferes with speech comprehension in everyday listening situations? J. Amer. Acad. Audiol. 18, 578–591.
Shinoda, K., Watanabe, T., 1997. Acoustic modeling based on the MDL criterion for speech recognition. In: Proc. Eurospeech, 1997, pp. 99–102.
Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Amer. 26 (2), 212–215.
Summerfield, A.Q., 1979. Use of visual information for phonetic processing. Phonetica 36, 314–331.
Tun, P.A., O'Kane, G., Wingfield, A., 2002. Distraction by competing speech in young and older adult listeners. Psychol. Aging 17 (3), 453–467.
Verhaeghen, P., Marcoen, A., Goossens, L., 1993. Facts and fiction about memory aging: a quantitative integration of research findings. J. Gerontol. 48 (4), 157–171.
Wolfram, S., 1991. Mathematica: A System for Doing Mathematics by Computer. Addison-Wesley, New York.
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.-Y., Li, L., 2007. Effect of number of masking talkers on speech-on-speech 7(masking)-352.74.51.24