ORIGINAL ARTICLE
J Audiol Otol 2015;19(1):39-44

pISSN 2384-1621 / eISSN 2384-1710
http://dx.doi.org/10.7874/jao.2015.19.1.39

Speech Recognition in Real-Life Background Noise by Young and Middle-Aged Adults with Normal Hearing

Ji Young Lee[1,2,3], Jin Tae Lee[1], Hye Jeong Heo[1], Chul-Hee Choi[1,2,3], Seong Hee Choi[1,2,3], and Kyungjae Lee[1,2,3]

[1]Department of Audiology and Speech-Language Pathology, [2]Research Institute of Biomimetic Sensory Control, [3]Catholic Hearing Voice Speech Center, Catholic University of Daegu, Gyeongsan, Korea

Received October 31, 2014
Revised December 17, 2014
Accepted February 27, 2015

Address for correspondence
Ji Young Lee, PhD
Department of Audiology and Speech-Language Pathology, Catholic University of Daegu, 13-13 Hayang-ro, Hayang-eup, Gyeongsan 712-702, Korea
Tel +82-53-850-2544
Fax +82-53-359-6780
E-mail jyslwm@gmail.com

Background and Objectives: People usually converse in real-life background noise. They experience more difficulty understanding speech in noise than in a quiet environment. The present study investigated how speech recognition in real-life background noise is affected by the type of noise, signal-to-noise ratio (SNR), and age. **Subjects and Methods**: Eighteen young adults and fifteen middle-aged adults with normal hearing participated in the present study. Three types of noise [subway noise, vacuum noise, and multi-talker babble (MTB)] were presented via a loudspeaker at three SNRs of 5 dB, 0 dB, and -5 dB. Speech recognition was analyzed using the word recognition score. **Results**: 1) Speech recognition in subway noise was the greatest in comparison to vacuum noise and MTB, 2) at the SNR of -5 dB, speech recognition was greater in subway noise than vacuum noise and in vacuum noise than MTB while at the SNRs of 0 and 5 dB, it was greater in subway noise than both vacuum noise and MTB and there was no difference between vacuum noise and MTB, 3) speech recognition decreased as the SNR decreased, and 4) young adults showed better speech recognition performance in all types of noises at all SNRs than middle-aged adults. **Conclusions**: Speech recognition in real-life background noise was affected by the type of noise, SNR, and age. The results suggest that the frequency distribution, amplitude fluctuation, informational masking, and cognition may be important underlying factors determining speech recognition performance in noise.

J Audiol Otol 2015;19(1):39-44

KEY WORDS: Real-life background noise · Speech recognition · Speech recognition in noise · Signal-to-noise ratio · Age.

Introduction

Typically, conversation does not occur in completely quiet environment, but rather in interfering real-life background noise. It is more difficult to have a conversation in noise than in quiet. In particular, people with sensorineural hearing loss are more handicapped when they listen to speech sounds in background noise compared to people with normal hearing [1]. However, some studies showed that even people with normal hearing needed more effort to understand speech in noise than in quiet condition. Although numerous research have studied the effects of the type of noise, signal-to-noise ratio (SNR), and age on speech recognition in noise, there have been only a few studies which examined speech recognition in real-life background noise [1-10].

For the type of noise, white noise, narrow band noise, speech-shaped noise, and multi-talker babble (MTB) have been usually used to study speech perception in noise [1,2] because they are easily available in the clinical setting. However, we live in a variety of real-life background noise such as traffic noise, industrial noise, loud music, etc. Such background noises have more complex characteristics in spectral and temporal aspects compared to the steady noise used in speech audiometric procedure. People with normal hearing understood speech better in cocktail party noise and white

noise than in four-talker competitor and nine-talker competitor [3]. Also, all subjects including young and old people with normal hearing and young and old people with hearing loss showed better speech discrimination in traffic noise and continuous discourse than in speech noise and cocktail party noise [4]. Stationary noises such as crowd, car, and construction noises masked speech more than fluctuating or interrupted noises such as gun noise and bird noise [5].

The SNR is one of the primary factors determining speech recognition performance in noise. Generally, the higher SNR, the better speech recognition in noise [6]. People with sensorineural hearing loss experience more difficulty in understanding speech at the decreased SNRs. When ten monosyllabic words were presented in quiet and MTB, people with sensorineural hearing loss required a 5.5 dB more favorable SNR than people with normal hearing to obtain the same level of speech recognition in MTB while their performance was similar to people with normal hearing in quiet [7].

Age is another factor widely studied to understand speech perception in noise [8,9]. People have to extract speech signal from competing background noise to understand speech in noise. Older adults, however, experience more difficulty understanding speech in noise than young adults because they have disadvantages in peripheral and central auditory function as well as cognition [10]. They are more distracted when they listen to speech in noise as well as they have increased hearing thresholds [10]. In particular, cognitively meaningful background noise such as babble noise degraded speech recognition more in old people than young people. Older adults were good at separating speech from a steady-state noise, but not from babble compared to younger adults [2]. Whereas old people's speech recognition in quiet or steady noise was well predicted by audibility, their speech recognition in MTB was poorer than predicted [11]. To examine the age-related effect on speech recognition in noise, most studies included people over the age of sixty as older subjects. However, not only elderly but also middle-aged adults report the complaints related to the background noise in communication. The routine audiometric examination, however, does not reveal their poor speech processing in adverse listening conditions because they usually show normal hearing sensitivity and normal speech recognition in quiet. In addition, there is some behavioral and electrophysiological evidence to suggest the pre-senescent decline in auditory processing. Middle-aged adults showed more decreased N1-P2 response to stimulus duration, poorer gap detection, and poor discrimination compared to younger adults [12,13]. The relationship between auditory temporal processing and speech perception has been widely studied. In particular, the gap detection showed significant correlation with speech recognition even when hearing threshold is factored out, and predicted speech recognition in noise [14,15]. Thus, the findings that middle-aged adults showed decreased auditory temporal processing bring up the question of whether their speech recognition in noise may also decrease.

The present study aimed to investigate the effects of noise type, SNR, and age on speech recognition in real-life background noise. To reflect real-life environment, various real-life background noises were presented via a loudspeaker in sound filed. Also, middle-aged adults with normal hearing were included as subjects to examine whether they show different speech recognition performance from young people as older people do.

Subjects and Methods

Subjects

Thirty three monolingual Korean speakers participated in the present study. Eighteen were young adults in twenties (seven males, eleven females, M=22.4 yrs) and fifteen were middle-aged adults in forties and fifties (eight males, seven females, M=48.3 yrs). All subjects had no history of audiological, neurological, or psychological disorders nor speech, language, hearing, or learning disorders. Their pure tone averages were below 25 dB HL and tympanograms were normal for both ears. This study was approved by the Catholic University of Daegu Institutional Review Board.

Procedure

Prior to the experiment, all subjects signed informed consent statements, filled out a case history form, and received pure tone audiometry and tympanometry as the screening test. The word recognition score (WRS) tests were administered in quiet and then in noise. Three types of noise (subway noise, vacuum noise, MTB) were presented at three SNRs (5 dB, 0 dB, and -5 dB). Thus, subjects received ten WRS tests in total (one in quiet, nine in noise: three types of noise×three SNRs). The noise was presented in random order to get rid of fatigue effect.

Subjects were seated 1 m from a loudspeaker and instructed to listen to 25 monosyllabic words carefully and write down the words they would listen to. Of fifty monosyllabic words in total on each word list, twenty five were used for each WRS test to avoid fatigue effect. The noise stimuli and the speech stimuli were presented simultaneously via the loudspeaker (Acoustical Analyzer AA1200, Starkey Hearing Technologies, Eden Prairie, MN, USA). Speech stimuli were presented at 55 dB HL, 60 dB HL, and 65 dB HL to make the SNR -5, 0, and 5 dB with the level of noise fixed at 60 dB HL.

Both speech and noise stimuli were presented after calibration on the volume unit meter to make sure that the stimuli were presented at the levels they should be. There was no difference in speech recognition performance between when the level of speech was fixed and when the level of noise was fixed to make the SNR vary [16].

Stimuli

For noise stimuli, three types of real-life background noise were used: subway noise, vacuum noise, and MTB. These types of noise were selected as stimuli in that they represent the real-life background noise from transportation (subway noise), home (vacuum noise), and public places (MTB) we are commonly exposed to in our daily life. Subway noise was recorded at the center in a train of on Daegu subway line 2 where few people are seated to minimize the additional noise from people's transit and conversation. The announcement and the short excessive noise from the departure and arrival of the train were also digitally deleted to select the typical noise inside subway train (Adobe Audition version 7.0, Adobe Systems Inc., Mountain View, CA, USA). Vacuum noise was recorded 1 m from the suction portion of a vacuum (VC4905F-HA, LG electronics Inc., Seoul, Korea) by keeping a vacuum working strong in the house while MTB was recorded during the break time when forty undergraduate students were conversing simultaneously in a lecture room before the lecture would start. After noises were recorded at a sampling rate of 44.1 kHz and a bit depth of 16 bits, using a mobile phone (SHV-E250S, Samsung Electronics Co., Suwon, Korea) for ten to twenty minutes, the 5-minute noise stimuli were constructed for each noise and normalized to have the same total root mean square amplitude (Adobe Audition version 7.0, Adobe Systems Inc., Mountain View, CA, USA). The segments of 5-minute noise stimuli were chosen based on two of researchers' auditory perceptual evaluation, using a 5-point scale with 1 point for least natural and 5 points for most natural. They agreed that the selected noise stimuli were all 5 points and representative of the typical real-life environment. Figs. 1 and 2 show frequency spectra and waveforms of the noise stimuli.

For speech stimuli, the monosyllabic words on the Korean Standard-Monosyllabic Word Lists [17] were used. These monosyllabic words were presented via a compact disc recording. Twenty five monosyllabic words on each word list were used for each WRS test.

Statistical analysis

The speech recognition was analyzed on the WRS, using an independent t-test for the quiet condition and using a three-way mixed analysis of variance (ANOVA) for the noise condition. On a three-way mixed ANOVA, the noise type (3 levels: subway noise, vacuum noise, MTB) and SNR (3 levels: -5 dB, 0 dB, 5 dB) were the within-subject factors, and the age group (2 levels: young adult, middle-aged adult) was the between-subject factor. Bonferroni corrections were used for post-hoc tests. Statistical analysis was performed with the IBM SPSS Statistics for Windows (version 19, IBM corp., Armonk, NY, USA). The criterion used for statistical signifi-

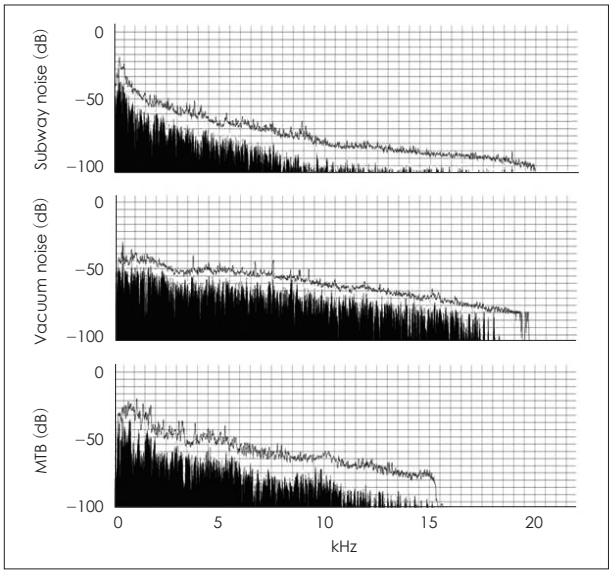

Fig. 1. Frequency spectra (colored) and peak amplitude (black line) at 20 seconds after onset of the noise stimuli. MTB: multi-talker babble.

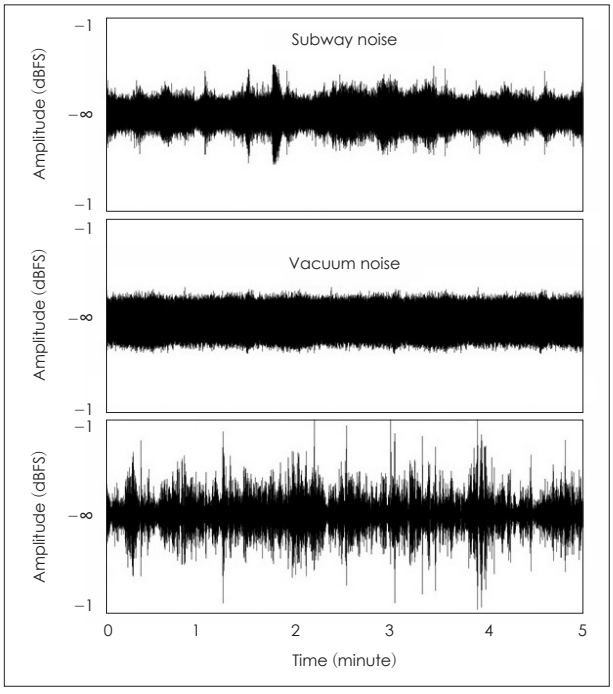

Fig. 2. Waveforms of the noise stimuli.

cance was $p<0.05$.

Results

For the WRS in quiet, there was no significant difference between young and middle-aged adults (young adults: mean=97.56, SD=2.79, middle-aged adults: mean=96, SD=4.06).

For the WRS in noise, the effects of nose type ($F_{2, 62}$=271.775, $p<0.001$) and SNR ($F_{2, 62}$=246.447, $p<0.001$) were significant, respectively and the interaction effect of the noise type and SNR was also significant ($F_{4, 124}$=6.436, $p<0.001$). The results of post-hoc tests of interaction effect between the noise type and SNR are as follows: for the noise type, when the SNR was 5 dB, the WRS was greater in subway noise than MTB and in MTB than vacuum noise ($F_{2, 64}$=67.805, $p<0.001$). When the SNR was 0 and -5 dB, however, the WRS was greater in subway noise than both vacuum noise and MTB and there was no difference between the WRSs in vacuum noise and MTB (SNR=0: $F_{2, 64}$=56.049, $p<0.001$, SNR=-5: $F_{2, 64}$=154.309, $p<0.001$). Fig. 3 shows the WRS by the noise type at the SNRs of 5 dB, 0 dB, and -5 dB. On the other hand, for the SNR, the WRS decreased as the SNR decreased regardless of the noise type. The WRS was greater at the SNR of 5 dB than 0 dB and at the SNR of 0 dB than -5 dB (subway noise: $F_{2, 64}$=72.165, $p<0.001$, vacuum: $F_{2, 64}$=103.081, $p<0.001$, MTB: $F_{2, 64}$=101.911, $p<0.001$). In addition, young adults showed better WRS than middle-aged adults ($F_{1, 31}$=3074.722,

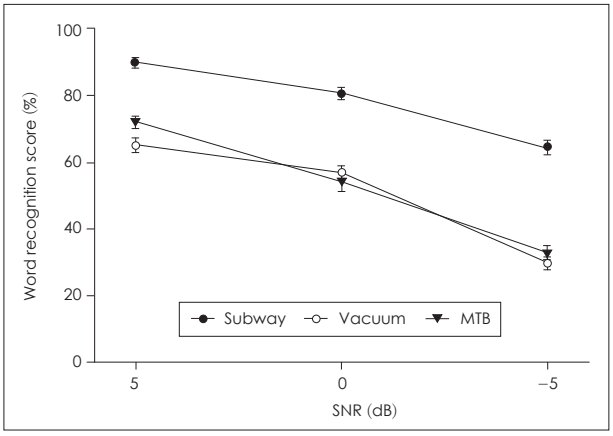

Fig. 3. Word recognition score by the noise type at the signal-to-noise ratios of 5 dB, 0 dB, and -5 dB. Error bars indicate 1 standard error from the mean. SNR: signal-to-noise ratio, MTB: multi-talker babble.

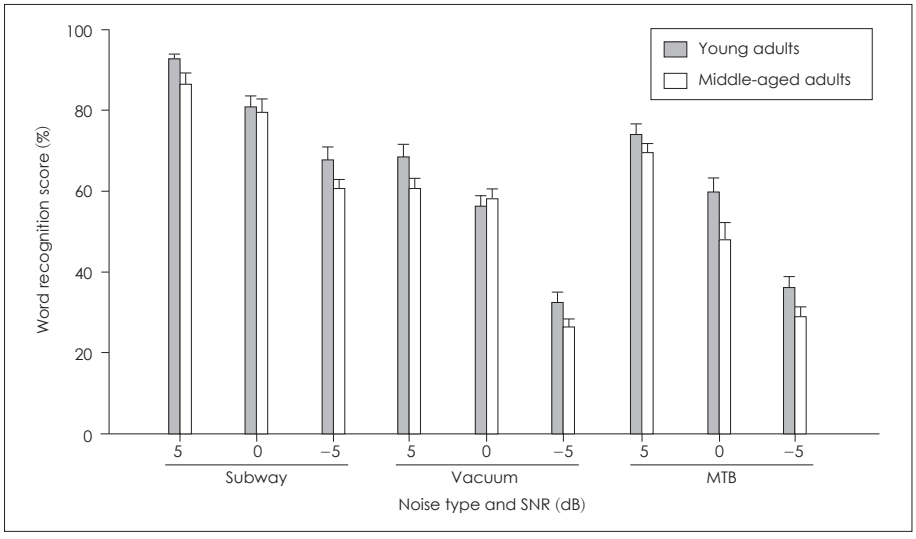

Fig. 4. Word recognition score by the noise type and SNR in young and middle-aged adults. Error bars indicate 1 standard error from the mean. SNR: signal-to-noise ratio, MTB: multi-talker babble.

Table 1. Word recognition score by noise type, SNR, and age group

	Subway noise			Vacuum noise			MTB		
	5 dB	0 dB	−5 dB	5 dB	0 dB	−5 dB	5 dB	0 dB	−5 dB
Young									
Mean	92.89	81.33	67.78	68.44	56.00	32.44	73.78	59.78	36.00
SD	4.24	9.41	13.68	13.57	12.12	11.14	12.06	14.22	11.88
Middle-aged									
Mean	86.40	79.73	60.53	60.80	58.13	26.40	69.60	47.80	28.80
SD	10.67	11.46	8.93	9.10	9.43	7.06	8.39	16.82	10.28

SD: standard deviation, SNR: signal-to-noise ratio, MTB: multi-talker babble

$p<0.001$) regardless of the noise type and SNR. For the noise type, young adults showed the WRSs of 80.67%, 52.30%, 56.52% whereas old adults showed the WRSs of 75.56%, 48.44%, 48.74% for subway noise, vacuum noise, and MTB, respectively. Also, for the SNR, young adults showed the WRSs of 78.37%, 65.70%, 45.41% while old adults showed the WRSs of 72.27%, 61.89%, 38.58% at the SNRs of 5 dB, 0 dB, -5 dB, respectively. The WRS by the noise type and SNR in young and middle-aged adults are shown in Fig. 4. Table 1 shows the summary of WRS by the noise type, SNR, and age group.

Discussion

For the WRS in quiet, there was no significant effect of age between young and middle-aged adults. It implies that middle-aged adults with normal hearing have no difficulty understanding speech as young adults, if there is no interfering background noise.

For the WRS in noise, however, there were significant effects of the noise type, SNR, and age. First, there was a significant interaction effect between the noise type and SNR. In terms of noise type, the WRSs in vacuum noise and MTB showed different tendency of changing as the SNR changed whereas the WRS was consistently greatest in subway noise regardless of the SNR. The greatest WRS in subway noise could be attributed to the acoustic characteristics of noise with relatively lower energy at high frequency regions. Because the frequency of speech sound ranges from 100 Hz up to 8000 Hz, subway noise may not have masked the speech signal effectively at high frequency regions which include the information of most consonants such as stops, affricates, and fricatives. It is consistent with the finding of a previous study that speech discrimination was greater in traffic noise and continuous discourse where the background noise has the lower energy at high frequencies compared to the used other noises [4]. Thus, it is reasonable to speculate that energy allocation in the frequency spectrum of noise can affect the perception of speech predominatingly. Whereas the WRS was greater in MTB than vacuum noise at a favorable SNR (SNR=5 dB), there was no difference between the WRSs in vacuum noise and MTB at unfavorable SNRs (SNR=0 dB, SNR=-5 dB). It may be due to some combined effects of vacuum noise with relatively less amplitude fluctuation and MTB with relatively more informational masking. Vacuum noise is a steady noise whereas MTB is a noise with instantaneous amplitude fluctuation. Thus, when the SNR was 5 dB, more stationary vacuum noise with less fluctuation in amplitude may have affected speech recognition more negatively while the MTB with more fluctuation in amplitude may have allowed the listeners to recognize speech better during the presentation of less loud portions of noise. The listeners can 'glimpse' the speech in the presence of nose when the masking energy reduces momentarily [18,19]. As the SNR decreased from 5 dB to 0 dB and -5 dB, however, verbal information in MTB may have masked speech signal more effectively, resulting in no difference in the WRSs in vacuum noise and MTB. These explanations are in line with the conclusion of a previous study that less fluctuating noise in amplitude and more competing speech information affected speech understanding more negatively [20]. A previous study showed that subjects were good at separating speech from a steady-state noise, but not from babble noise, suggesting that informational masking affected the speech recognition more dominantly than amplitude fluctuation [10]. The greater the acoustic similarity between target and masker, the more difficult it is to separate target from masker [21]. The comparatively weaker effect of informational masking in the present study may be because the forty-talker babble includes less verbal informational than the fewer-talker babble used in most studies. Such interpretation the natural forty-talker babble can have the informational masking as well as energetic masking would need to be verified by future studies. On the other hand, in terms of the SNR, the WRS decreased as the SNR decreased regardless of the noise type. Speech recognition decreases as the SNR decreases. The relative intensity of the signal in background noise is important to accurately perceive speech in noise [6]. Given that people with sensorineural hearing loss require a more favorable SNR than normal people do to reach the same level of speech understanding [7], the SNR should be emphasized as a critical factor in understanding their speech recognition in noise.

In addition, young adults showed better WRS in all types of noises at all SNRs than in middle-aged adults. Middle-aged adults with normal hearing showed difficulty perceiving speech in noise though they had no difficulty in quiet. It is consistent with the findings of the previous study that that old people recalled fewer speech items after listening to speech in noise than young people [22]. It may be because speech recognition involves with cognition such as attending to the auditory signal, performing acoustic analysis, mapping the signal to phonemic representation, temporarily storing acoustic information in memory, and finally mapping phonemes to meaning in addition to peripheral auditory processing [23]. Moreover, it needs to segregate signal from competing background to recognize speech in noise. Thus, older people's reduced attention, memory, and cognitive-linguistic function may result in the decreased speech recognition in noise. It is noticeable that the effect of age was more pronounced for people

with hearing loss. The old people with normal hearing had slightly reduced speech discrimination in noise than young people with normal hearing but old people with hearing loss had more reduced speech discrimination in noise than old people with hearing loss [4]. The age effect in the present study is also consistent with the findings that middle aged adults showed more decreased auditory temporal processing than younger adults based on the significant correlation between auditory temporal processing and speech recognition in noise [12-15]. Thus, the middle-aged adults may have deficit not only in auditory processing but also in speech processing in difficult listening conditions compared to younger adults. Because older or elderly people who participated in the previous studies were mostly over the age of sixty, the findings of the present study provides additional information about the effect of age, indicating that middle-aged adults also showed decreased speech recognition performance in noise compared to young adults.

The present study has some limitations on the selection of noise type, SNR, age, and speech material, which are not sufficient to comprehensively represent the real world. Further studies would need to explore the speech recognition in real-life background noise more in depth in order to provide additional information to the routine audiometric examination.

Conclusion

The results of the present study showed that speech recognition in real-life background noise was significantly different for the noise type, SNR, and age. The results suggest that the frequency distribution, amplitude fluctuation, informational masking, and cognition may be important underlying factors determining speech recognition performance in noise.

Real-life background noise can reflect the realistic situation better than the speech-shaped noise used in the speech audiometric test. Because background noise in real life has more dynamic and complex characteristics, it may not be reasonable to predict our speech understanding in real-life environment by administering speech audiometric test. The present study could contribute to understanding speech recognition in real-life background noise and provide the basic information necessary to assess it for future clinical application.

Acknowledgments

This study was supported by research grants from the Catholic University of Daegu.

REFERENCES

1) Pittman AL, Wiley TL. Recognition of speech produced in noise. J Speech Lang Hear Res 2001;44:487-96.
2) Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. J Acoust Soc Am 1984;76:87-96.
3) Danhauer JL, Leppler JG. Effects of four noise competitors on the California Consonant Test. J Speech Hear Disord 1979;44:354-62.
4) Prosser S, Turrini M, Arslan E. Effects of different noises on speech discrimination by the elderly. Acta Otolaryngol Suppl 1990;476:136-42.
5) Rhebergen KS, Versfeld NJ, Dreschler WA. Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing. Ear Hear 2008;29:169-75.
6) Crandell CC, Smaldino JJ. Classroom acoustics for children with normal hearing and with hearing impairment. Lang Speech Hear Serv Sch 2000;31:362-70.
7) Wilson RH, Abrams HB, Pillion AL. A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. J Rehabil Res Dev 2003;40:321-7.
8) Helfer KS, Freyman RL. Aging and speech-on-speech masking. Ear Hear 2008;29:87-98.
9) Tun PA, O'Kane G, Wingfield A. Distraction by competing speech in young and older adult listeners. Psychol Aging 2002;17:453-67.
10) Ben-David BM, Tse VY, Schneider BA. Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? Hear Res 2012;290:55-63.
11) Sherbecoe RL, Studebaker GA. Audibility-index predictions of normal-hearing and hearing-impaired listeners' performance on the connected speech test. Ear Hear 2003;24:71-88.
12) Ostroff JM, McDonald KL, Schneider BA, Alain C. Aging and the processing of sound duration in human auditory cortex. Hear Res 2003;181:1-7.
13) Snell KB, Frisina DR. Relationships among age-related differences in gap detection and word recognition. J Acoust Soc Am 2000;107:1615-26.
14) Dreschler WA, Plomp R. Relations between psychophysical data and speech perception for hearing-impaired subjects. II. J Acoust Soc Am 1985;78:1261-70.
15) Snell KB, Mapes FM, Hickman ED, Frisina DR. Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. J Acoust Soc Am 2002;112:720-7.
16) Wilson RH, McArdle R. Speech-in-noise measures: variable versus fixed speech and noise levels. Int J Audiol 2012;51:708-12.
17) Lee JH, Jo SJ, Kim JS, Jang HS, Lim DW, Lee KW. Korean Speech Audiometry. Seoul: Hakjisa;2010.
18) Li N, Loizou PC. Factors influencing glimpsing of speech in noise. J Acoust Soc Am 2007;122:1165-72.
19) Cooke M. A glimpsing model of speech perception in noise. J Acoust Soc Am 2006;119:1562-73.
20) Wong LL, Ng EH, Soli SD. Characterization of speech understanding in various types of noise. J Acoust Soc Am 2012;132:2642-51.
21) Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am 2001;110(5 Pt 1):2527-38.
22) Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. J Acoust Soc Am 1995;97:593-608.
23) Wong PC, Jin JX, Gunasekera GM, Abel R, Lee ER, Dhar S. Aging and cortical mechanisms of speech perception in noise. Neuropsychologia 2009;47:693-703.