older adults is shallower than that for younger adults, however, this difference is not statistically significant (p > .2, two-tailed). Finally, in Figure 3, the Laplace projection values of older adults in the current identification experiment and those in the experiment by Murphy et al. (2006) are plotted as a function of the Laplace projection values obtained from the younger adults in the same experiments. The straight line in this plot (slope = 1, intercept = 0) is what we would expect if the identification performance of older adults was exactly identical to that of younger adults. As can be seen from this figure, despite the fact that pairwise discriminability was reduced in the current experiment due to smaller intensity separation between adjacent tones, older adults appear to perform as well as younger adults in both experiments. These results indicate that younger and older adults perform equivalently in identification tasks, independent of the range of stimuli employed.

Discussion

The current experiment replicates Murphy et al's (2006) findings that auditory channel capacity is preserved in aging. Despite using a much smaller range of stimuli than Murphy et al. (2006), we found no age differences in the performance of younger and older adults in absolute identification task with tones varying in intensity only. Our findings, together with those obtained by Murphy et al. (2006), indicate that auditory channel capacity does not diminish with healthy aging.

References

Martin, J. S., & Jerger, J. F. (2005). Some effects of aging on central auditory processing. *Journal of Rehabilitation Research and Development*, *42*(4), 25-43.

Murphy, D. R., Schneider, B. A., Speranza, F., & Moraglia, G. (2006). Comparison of higher order auditory processes in younger and older adults. *Psychology and Aging, 21(4)*, 763-773.

Parker S., Murphy D. R., and Schneider B. A. (2002). Top-down gain control in the auditory system: Evidence from identification and discrimination experiments. *Perception and Psychophysics, 64* (4), 598-615.

Schneider, B. A. (2007). The shape of the underlying distributions in absolute identification experiments. In S. Mori, T. Miyaoka, & W. Wong (Eds.) *Fechner Day 2007: Proceedings of the 23rd Annual Meeting of the International Society for Psychophysics.* Tokyo, Japan: International Society for Psychophysics.

Schneider, B. A., and Pichora-Fuller, M. K. (2000). Implications of sensory deficits for cognitive aging. In F. I. M. Craik & T. Salthouse (Eds), *The handbook of aging and cognition* (2nd ed., pp. 155-219). Mahwah, NJ: Erlbaum.

PERCEPTUAL INTEGRATION BETWEEN TARGET SPEECH AND TARGET-SPEECH REFLECTION REDUCES MASKING FOR TARGET-SPEECH RECOGNITION IN YOUNGER AND OLDER ADULTS

Ying Huang, Qiang Huang, Xun Chen, Tianshu Qu, Xihong Wu, Liang Li[a]

Department of Psychology, Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing, China, 100871
[a]E-mail: liangli@pku.edu.cn

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 adults and older adults with clinically-normal hearing, reducing the time interval between target speech and target-reflection simulation (inter-target interval, ITI) from 64 to 0 ms progressively released target speech from either speech masking or noise masking. 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.

In noisy, reverberant environments, it is more difficult for older adults than for younger adults to recognize speech (e.g., Nabelek and Robinson, 1982). Under these conditions, listeners receive not only sound waves that directly emanate from various sources but also filtered and time-delayed reflections from surfaces at various locations. Fortunately, the reflected waves can be perceptually integrated with their direct wave by the auditory system to form the "precedence effect" (Litovsky et al, 1999). The precedence effect weakens auditory echoes. Also, the perceptual integration of direct and reflected waves can increase speech recognition under multiple-talker conditions (e.g., Freyman et al., 1999; Li et al., 2004; Wu et al., 2005) by enhancing perceptual differences (i.e. differences in perceived spatial location and in auditory image such as compactness/diffusiveness, timbre, and/or loudness) between target speech and masking speech, leading to improved selective attention to target speech (Schneider et al., 2007). When the time interval between the direct and the reflected waves is short (such as 3 or 4 ms), there is no difference of the advantage of perceptual difference for target speech recognition between older and younger adults (Li et al., 2004; Helfer and Freyman, 2008). However, some studies on younger adults have shown that the

advantage of perceptual integration can occur over a large range of time intervals (Brungart et al., 2005; Rakerd et al, 2006), about 32 ms under speech masking conditions. The present study investigated whether the age-related difficulties in speech recognition under simulated adverse conditions are related to an assumed age-related decline in the advantage of perceived spatial separation when the time interval between the direct and the reflected waves is changed over a large range.

METHODS

Thirty-six young university students (18 – 33 years old) and thirty-six older adults (60 – 75 years old) participated in this study. Their first language was Mandarin Chinese. All young students all had normal and symmetrical 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. Both the 36 younger participants and the 36 older adults were randomly divided into three groups with twelve for each group. Different groups were assigned with different SNRs in speech-recognition testing: -4 dB for Younger Group 1, -6 dB for Younger Group 2, and -8 dB for Younger Group 3; -2 dB for Older Group 1, -4 dB for Older Group 2, and -6 dB for Older Group 3.

The participant was seated at the center of an anechoic chamber (Beijing CA Acoustics). Acoustic signals were presented to participants through two loudspeakers (Dynaudio Acoustics, BM6 A) in the frontal azimuthal plane at the left and right 45° positions with respect to the median plane. Speech stimuli were Chinese "nonsense" sentences, which are syntactically correct but not semantically meaningful. 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).

Target speech was spoken by a young female talker A. The two loudspeakers presented the identical target sentences, but the left loudspeaker either led (in positive ITI values) or lagged behind (in negative ITI values) the right loudspeaker with the absolute ITI value of 0, 0.5, 1, 2, 4, 8, 16, 32, or 64 ms for younger participants. For older-participants, the 0.5-ms ITI was not used. The two loudspeakers also presented either (young female) two-talker-speech maskers (both talkers and contents were different between the two loudspeakers) or steady-state speech-spectrum-noise maskers that were not correlated between the two loudspeakers. The masking-speech sentences presented from the left loudspeaker were spoken by talkers B and C, and by talkers D and E for the right loudspeaker. The masker signals in the two loudspeakers were presented simultaneously but 1 s before the presentation of the leading target speech. The masker and the target were gated off at the same time.

During the testing, target-speech sounds were presented at a constant level of 56 dBA in each loudspeaker. The participant was instructed to loudly repeat the whole target sentence immediately after all the sounds were completed. Performance for each participant was scored on the number of correctly identified syllables in keywords.

RESULTS

Figure 1 shows the release of speech from masking for the older and younger listeners under different conditions. The release of speech at certain ITI condition was calculated
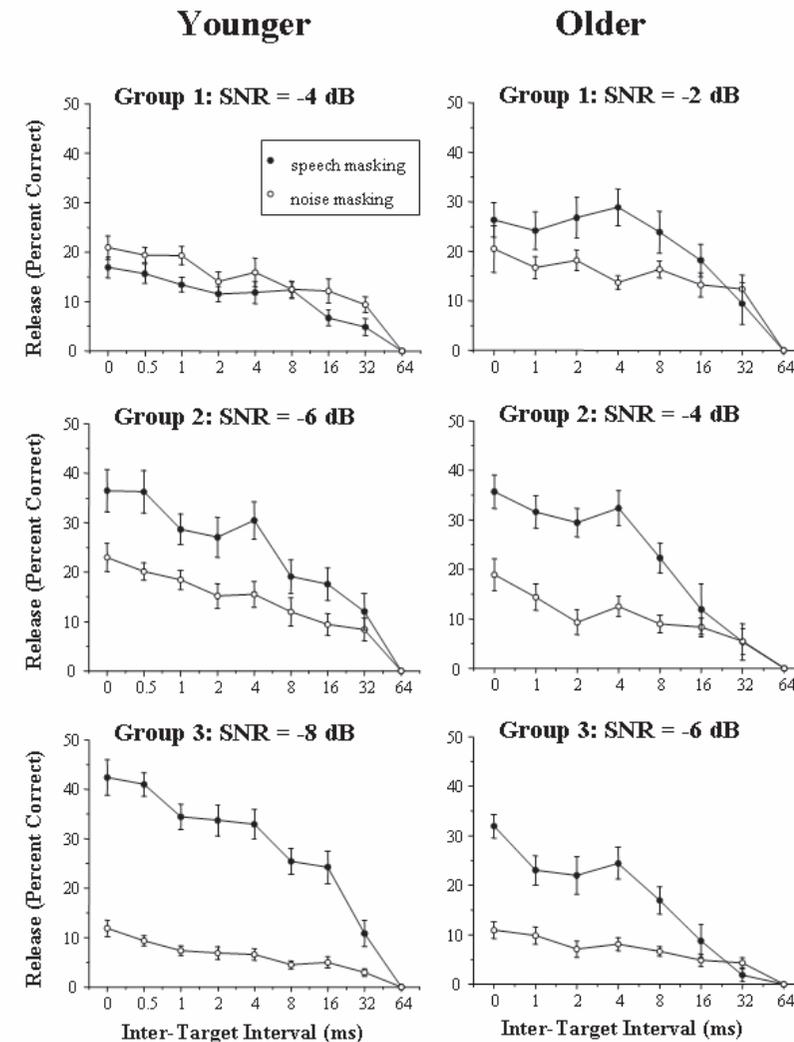

Figure 1. The release of target speech as a function of the absolute value of ITI for each of the three younger participant groups and each of the three older participant under the speech-masking condition (closed circles) and under the noise-masking condition (open circles). Note that data for the two 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.

through the following two steps: first, the percent-correct recognition of target speech under the ITI was averaged between the two leading conditions. Then the averaged percent correct at this ITI subtracted that at the ITI of 64 ms. For all the 6 groups, under either the speech-masking or the noise-masking condition, the release 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. When the SNR was at the level of -4 dB for younger participants and -2 dB for older participants, reducing the ITI from 64 to 0 ms led to an equivalent improvement of target-speech recognition for both masking conditions, 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. Two-way within-subject ANOVAs for each of these four groups show that the interaction between masker type and ITI was significant (Younger group 2: $F_{8,88}$ = 6.738, $p < 0.001$; Younger group 3: $F_{8,88}$ = 29.891, $p < 0.001$; Older group 2: $F_{7,77}$ = 9.193, $p < 0.001$; Older group 3: $F_{7,77}$ = 21.428, $p < 0.001$). 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 ($p < 0.050$) for Young group 2; from 0 to 32 ms ($p < 0.010$) for Younger group 3; from 0 to 8 ms ($p < 0.010$) for older group 2 and 3.

Further follow-up paired-sample *t*-tests at the level of 0.00625 (with a Bonferroni adjustment) were used in each of the six participant groups to find the longest effective ITI at which the release of speech from masking was significantly more than that at the ITI of 64 ms. The results are showed in Figure 2. For older participants, when the SNR was low (-4 or -6 dB), the release was significant only when the ITI was 8 ms or shorter under the speech-masking condition. However, for younger participants, when the SNR was -6 dB the release was significant when the ITI was 16 ms or shorter, and when the SNR was -8 dB the release was significant when the ITI was 32 ms or shorter.

These 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 listeners; -4 or -6 dB for older listeners), 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 to 16 or 32 ms) than in older participants (0 to 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.

CONCLUSIONS

For both younger and older listeners, recognition of target speech improves progressively with the reduction of the absolute ITI value from 64 to 0 ms especially under speech masking conditions because of the advantage of perceptual integration of target signals. Although older listeners' ability to use the perceptual cues provided by the reduction of ITI for improving speech recognition is well retained, the ITI range for a significant release under speech masking conditions 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.
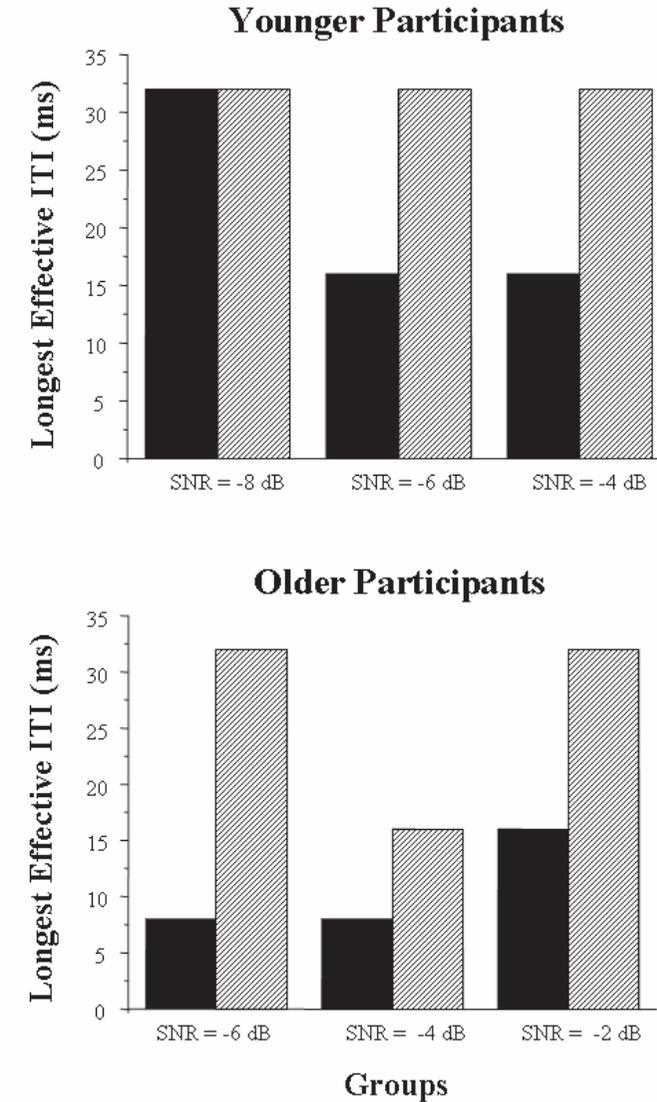

Figure 2. Summary of the longest effective ITI, which caused significant release of target speech from making, at each of the stimulus conditions. Black columns, speech masking; hatched columns, noise masking.

REFERENCE

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.

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.

Helfer, K.S., Freyman, R.L., 2008. Aging and speech-on-speech masking. Ear Hear. 29, 87–98.

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. Psych.: Hum. Per. Perf. 30, 1077 – 1091.

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.

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.

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.

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.

INTRINSIC RELATIONSHIP BETWEEN FACIAL MOTION AND ACOUSTICS IN INDIVIDUALS WITH PARKINSON'S DISEASE

Luyao Ma[1,2], Pascal van Lieshout[1,2,3,5], Huawei Colin Li[1,3,4], and Akiko Amano-Kusumoto[1,6]
[1]Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto; [2]Department of Psychology, University of Toronto; [3]Institute of Biomaterials and Biomedical Engineering, University of Toronto; [4]Department of Electrical and Computer Engineering, University of Toronto; [5]Toronto Rehabilitation Institute; [6]OGI School of Science & Engineering, Oregon Health & Science University, Portland, USA
luyao.ma@utoronto.ca; p.vanlieshout@utoronto.ca; huawei.li@utoronto.ca; akusumoto@bme.ogi.edu

Abstract

Parkinson's disease (PD) is closely associated with the death of dopaminergic neurons in the basal ganglia, which results in a reduction of facial dynamics during speech production. In young adult speakers, the relationship between facial motion and acoustics is robust. It can be hypothesized that this relationship between facial motion and speech acoustics is reduced in PD given the limitations in facial expression; however, virtually no study had addressed this relationship yet. The current project was designed to address this issue using a 3D video system in combination with Blacklight illumination to record facial motion with time aligned acoustic data. Findings show that in comparison to age-matched control speakers, PD subjects have significantly lower correlations for speech gestures, except for upper lip movement. The findings of this study have implications for the future development of facial motion based speech recognition software and rehabilitation tools.

In speech production abstract gestural goals are mapped onto time varying inputs of muscular activation levels that control the movement and positioning of biomechanical structures (Goldstein, Byrd, & Saltzman, 2006). The resulting shape of the vocal tract acts as an acoustic filter to produce resonance in specific frequencies, which is largely responsible for the various sounds that we hear (Stevens, 1989). Gestural patterns shape vocal tract dynamics in both visual and auditory ways, so its perception is not bound to a single (auditory) modality. This is captured in the notion of "audio-visual (AV) speech perception". From the perceivers' perspective, the coherence of observed visual information and acoustic signals is very important for comprehension. For example, a recent study indicated that subjects who had lip-read a speaker for one hour subsequently recovered speech in noise better when the acoustic signal was from the same talker as opposed to being from a different talker (Rosenblum, Miller, & Sanchez, 2007). Given the multi-modal nature of speech perception, facial motility is an important factor in how clearly a person produces speech and how others perceive the intended message. The study described here focuses on one particular population where facial motility is an issue, namely individuals with Parkinson's disease.

Parkinson's disease (PD) is a common neurological condition, characterized by muscle rigidity, tremor and slowness in physical movement and is particularly prevalent in people above 50 years of age (Pinto et al., 2004). One of the most severe consequences of PD is the loss of expressiveness in the body and especially in the face. As a result, the facial motility in patients affected by PD gradually diminishes and results in a mask-like status. This affects the natural facial dynamics, which accompany the production of speech (Smith, Ellgring, & Smith, 1996). In addition, PD motor impairment also often results in adverse