performance between younger and older adults is that older adults needed a 2.8 dB increase in signal-to-noise ratio to perform equivalently to younger adults.

This study illustrates how psychophysical techniques can be employed to ascertain the source (perceptual or cognitive) of age-related changes in the ability to recognize speech in noisy backgrounds. The use of the precedence effect to manipulate spatial position ensured that the degree of energetic or peripheral masking (the co-activation of different regions along the basilar membrane) was essentially the same for all perceived locations of the target and masker. Hence, any change in sensitivity with perceived spatial position had to reflect changes in more central auditory or cognitive processes. If, for example, older adults had experienced less of a release from informational masking due to spatial separation than did younger adults, this age difference would have to have been attributed to age-related declines in more central auditory or cognitive processes. The fact that older and younger adults exhibited an equivalent release from informational masking indicates that, at least with respect to nonsense sentences, older adults are equally adept at parsing the auditory scene as younger adults, and obtain equal benefit from this parsing. Finally, the fact that older adults required 2.8 dB higher signal-to-noise ratios than younger adults indicates that their auditory systems are more sensitive to the effects of energetic masking than are those of younger adults.

In the following talks you will see examples of how psychophysical techniques can be used to investigate the sources (both perceptual and cognitive) of speech comprehension difficulties when listening to speech in complex and noisy auditory environments.

References

Bregman, A. S. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, MA.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech, *Journal of the Acoustical Society of America*, *106*, 3578-3588.

Li, L., Daneman, M., Qi, J., & Schneider, B.A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older Adults? *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 1077-1091.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, *18*, 578-591.

TRANSIENT AUDITORY STORAGE OF ACOUSTIC DETAILS IS ASSOCIATED WITH RELEASE OF SPEECH FROM INFORMATIONAL MASKING IN REVERBERANT CONDITIONS

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

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

Perceptual integration of the sound directly emanating from the source with the reflections needs temporal storage and correlation computation of acoustic details. In Experiment 1, a break in correlation (BIC) between interaurally correlated wideband or narrowband noises was detectable even when an interaural interval (IAI) was introduced. The longest IAI, which varied markedly across participants, decreased as the center frequency was increased for narrowband noises. In Experiment 2, when the interval between target speech and its single-reflection simulation (inter-target interval, ITI) was reduced from 64 to 0 ms, intelligibility of target speech was markedly improved under speech-masking but not noise-masking conditions. The longest effective ITI correlated with the longest IAI for detecting the BIC only in the low-frequency (≤ 400 Hz) narrowband noise. Thus the ability to temporally store fine details contributes to perceptual integration of correlated sounds, which in turn, contributes to releasing speech from informational masking.

Both transient storage of acoustic features and temporal integration of relevant signals are important for detecting, recognizing, and localizing sounds in the every-day environment (Bregman, 1990). At the early stage of auditory perception, fine-structure details of sound waves must be faithfully maintained for a period of time, otherwise auditory processing at later stages would be impossible. The human's auditory system has the dramatic ability to process acoustic details and represents the consequence of the processing at the perceptual level (Blauert & Lindemann, 1986). For example, human listeners are able to detect a very transient break in correlation (BIC) between the two ears, showing the marked ability to temporally resolve fast changes in interaural configurations (Akeroyd & Summerfield, 1999; Boehnke et al., 2002). And this ability is frequency dependent (Akeroyd & Summerfield, 1999). However, previous studies did not investigate whether this ability can be maintained when an IAI is introduced. Thus measuring the longest IAI when the BIC is detectable can provide a way of investigating the temporal storage of acoustic details.

In noisy, reverberant environments, listeners receive not only direct waves from various sources but also filtered and time-delayed reflections from surfaces at various locations. The reflected waves should be perceptually integrated with their direct wave to weaken auditory echoes. And this perceptual integration can also 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 between target and masking speech to improve selective attention to target speech (Schneider et al., 2007). 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 purpose of the present study was to investigate the functional connection between the two types of abilities. One is the ability to temporally maintain

acoustic details in order to achieve the perceptual integration between an arbitrary noise and its delayed copy at the early auditory processing stage, which is measured by the longest IAI. The other is the ability to perceptually integrate target speech with its reflection simulation to achieve the perceptual segregation between target speech and maskers, and improve the recognition of target speech under adverse conditions, which is measured by the longest effective ITI.

Method

Nineteen young university students (19 – 25 years old, 13 females) with normal and balanced pure-tone hearing participated in the longest IAI testing, and 18 of them also participated in the longest effective ITI testing. In the IAI testing, 2000-ms (including 30-ms rise/fall times) Gaussian wideband noises were first generated and then the center of the noises was replaced by a 200-ms BIC. The whole noises were passed through either a 10k-Hz low-pass filter to get wideband noise or a group of 1/3-octave band-pass filters to get narrowband noises with a center frequency of 200, 400, 800, 1600 or 3200 Hz. Stimuli were presented to listeners by two headphones at 58 dBA SPL. At the beginning of each testing session, the IAI was fixed at 0 ms, and then the IAI was changed according to the 3-up-1-down procedure. The longest IAI was measured using the 2AFC paradigm.

The effective ITI testing was conducted in an anechoic chamber (Beijing CA Acoustics). All acoustic stimuli were presented to participants through two loudspeakers (Dynaudio Acoustics, BM6 A), which were in the frontal azimuthal plane at the left and right 45° positions with respect to the median plane. Speech stimuli used in the testing were Chinese "nonsense" sentences, each of which has 3 key components: subject, predicate, and object, which are also the 3 key words, with two characters for each. Target speech was spoken by a young female talker (Talker A). In each testing trial, the two loudspeakers presented identical target speech with the left loudspeaker led or lagged behind the right loudspeaker by 0, 0.5, 1, 2, 4, 8, 16, 32, or 64 ms. One second before the presentation of the leading target speech, the two loudspeakers also presented either speech maskers or speech-spectrum noise maskers simultaneously. The speech maskers presented in the left loudspeaker were spoken by talker B and C, and that in the right loudspeaker were spoken by talker D and E. The speech-spectrum noise maskers presented from the two loudspeakers were uncorrelated with each other. For single loudspeaker presentation, the target speech was presented at a sound level of 56 dBA and the masker was presented at a sound level of 64 dBA. 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

The center-frequency effect on the longest IAI is shown in Figure 1. Clearly, participants were able to detect the BIC over longer IAIs when the center frequency was low (200, 400, or 800 Hz) than when it was high (1600 or 3200 Hz). A one-way within-participant ANOVA shows that the effect of noise type was significant ($F_{5,75} = 40.189$, $p < 0.001$). Bonferroni post hoc analyses indicate that the longest IAI for wideband noises was not significantly different from that for narrowband noises with the center frequency of 200, 400 or 800 Hz, but significantly longer than that for narrowband noises with the center frequency of 1600 or 3200 Hz; and the longest IAI for the center frequency of 1600 or 3200 Hz was significantly shorter than that for the center frequency of 200, 400 or 800 Hz.

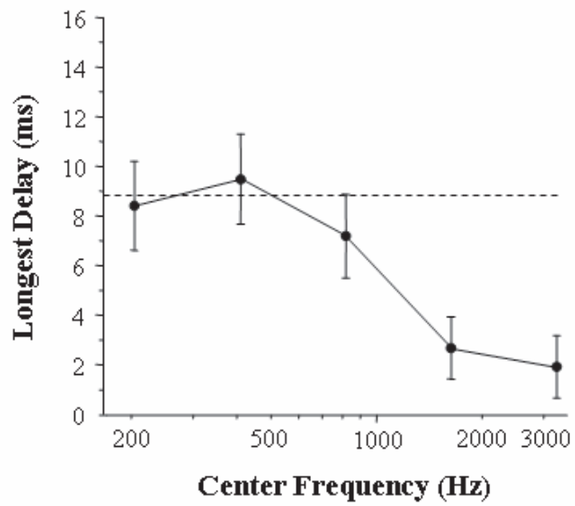

Figure 1. *The group mean of longest IAI at which the BIC in the narrowband noise could be detected as a function of the center frequency. The broken line represents the longest IAI when the noise was wideband. The error bars in this and following figures represent the standard errors of the mean.*

The top panels in Figure 2 show the release of speech from masking under speech masking and that under noise masking conditions. The release of speech at certain ITI condition was calculated 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. Under both masking conditions, the release increased with the decrease of the absolute value of ITI, but larger releases occurred under the speech-masking condition. A 2 (masker type) by 9 (absolute value of ITI) within-participant ANOVA shows that the main effect of masker type was significant ($F_{1, 17} = 109.349$, $p < 0.001$), and the main effect of ITI was significant ($F_{8, 136} = 111.520$, $p < 0.001$), the interaction between the two factors was significant ($F_{8, 136} = 37.205$, $p < 0.001$). The bottom panels of Figure 2 show the percent release as a function of ITI for individual participants under speech masking conditions or noise masking conditions. Clearly, there was a remarkable variability in the release across participants, particularly under speech masking conditions. The effective ITI for individuals is defined as the longest ITI (among those used in the experiment) at which the correct recognition of keyword syllables was significantly better than the correct recognition of keyword syllables at the ITI of 64 ms. In this experiment, effective ITIs could be obtained in each of the participants under speech-masking conditions. For individual participants under noise-masking conditions, however, significant differences in performance between the ITI of 64 ms and any other ITIs could not be obtained, except only for one participant whose performance at the ITI of 0.5 ms was significantly difference from that at the ITI of 64 ms. Thus we could not obtain reliable effective ITIs for individuals under noise masking conditions.

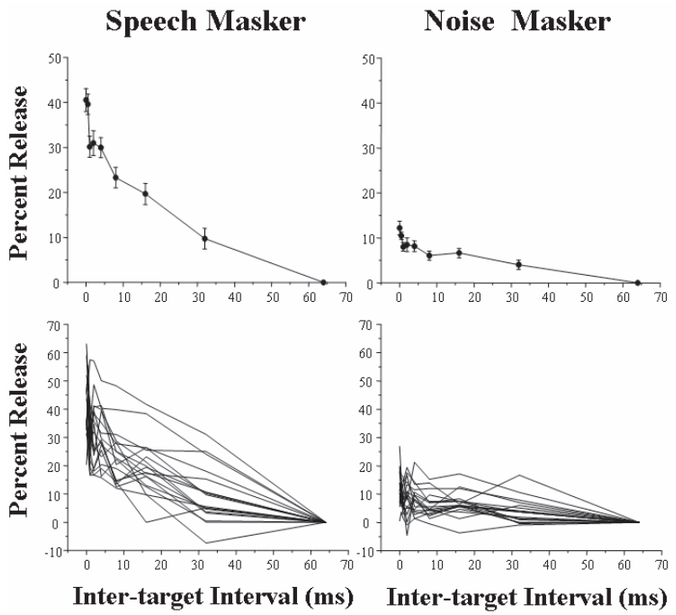

Figure 2. *Top panels: The group mean of percent release of target speech as a function of the absolute value of ITI under the speech masking condition (left panel) or speech-spectrum noise masking condition (right panel). Bottom panels: The percent release of target speech as a function of the absolute value of ITI under the speech masking condition (left panel) or under speech-spectrum noise masking condition (right panel) for individual participants.*

Correlation coefficients were calculated between the longest IAI (in the longest IAI testing) and the effective ITI under speech masking conditions (in the longest effective ITI testing) across the same eighteen participants. The correlation coefficients are shown in Figure 3. For the two narrowband noises with the low center frequencies (200 Hz, 400 Hz), significant correlations (200 Hz: $p = 0.041$; 400 Hz: $p = 0.005$) were obtained between the longest IAI and the effective ITI. When the center frequency of the narrowband noise was 800 Hz or higher, no significant correlations were found. Even for wideband noise, the correlation was only marginally significant ($p = 0.075$).

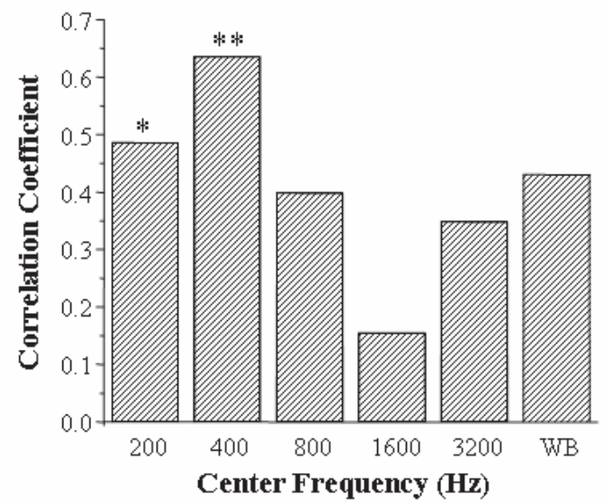

Figure 3. *Correlation coefficient between the longest IAI obtained in Experiment 1 for each noise type and the critical ITI obtained in Experiment 2 across the same eighteen participants. The critical ITI is defined as the longest ITI at which the correct recognition of keyword syllables was significantly better than that at the ITI of 64 ms. WB, wideband.*

Summary

The temporal storage of fine-structure information is frequency dependent. The storage of low-frequency details lasts longer than that of high-frequency details. And under the reverberation-simulation condition with the presentation of the speech masker, the reduction of the ITI from 64 to 0 ms progressively improves the recognition of target speech. Under noise masking conditions, however, the improvement is minor. Thus the reduction of the ITI enhances the temporal integration between target speech with its reflections and predominantly releases target speech from informational masking. This ability of temporal integrating speech with its reflections is functionally associated with the ability of temporal storing of low-frequency acoustic details. Thus the primitive auditory "memory", which occurs at the early stage of the transient auditory memory system, is critical for later high-level segregating target speech from informational masking in noisy, reverberant environments.

Reference

Akeroyd, M. A., & Summerfield A. Q. (1999). A binaural analog of gap detection. *Journal of the Acoustical Society of America, 105,* 2807 - 2820.

Blauert, J., & Lindemann, W. (1986). Spatial-mapping of intracranial auditory events for various degrees of interaural coherence. *Journal of the Acoustical Society of America, 79,* 806 - 813.

Boehnke, S. E., Hall, S. E., & Marquardt, T. (2002). Detection of static and dynamic changes in interaural correlation. *Journal of the Acoustical Society of America, 112,* 1617 - 1626.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press.

Brungart, D. S., Simpson, B. D., & Freyman, R. L. (2005). Precedence-based speech segregation in a virtual auditory environment. *Journal of the Acoustical Society of America, 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. *Journal of the Acoustical Society of America, 106,* 3578 - 3588.

Li, L., Daneman, M., Qi, G. Q., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? *Journal of Experiment Psychology: Human Perception and Performance, 30,* 1077 - 1091.

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.

Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *Journal of the Acoustical Society of America, 119,* 1597 - 1605.

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. *Hearing Research, 199,* 1 - 10.

MODULATION OF THE VOICE-CUING EFFECT ON RELEASING SPEECH FROM INFORMATIONAL MASKING

Lijuan Xu, 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, China,100871
E-mail: liangli@pku.edu.cn

Abstract

In cocktail-party environments, familiarity or knowledge of target talker's voice is useful for reducing speech-on-speech masking (Yang et al., 2007). This study investigated whether the voice-cuing effect can be modulated by either the degree of familiarity/knowledge of target talker's voice or the onset asynchrony between target speech and masking speech. When target speech started 1 second after masking speech, pre-presenting a priming sentence voiced by the target talker significantly improved the recognition of the target speech which was co-presented with masking speech. However, reinforcing the familiarity/knowledge of the target-talker's voice did not further improve the recognition. When target speech and masking speech started at the same time, a single presentation of voice-priming speech did not change participants' speech recognition against masking speech unless the familiarity/knowledge of target-talker's voice was reinforced by either a learning procedure or repeated presentation of the target-talker's voice before testing.

Cherry (1953) propose the idea that a few of factors, including the voice features of the attended talker can give the mental facility in recognizing what this talker is saying when others are speaking at the same time. Since any perceptual cues as long as they facilitate listeners' attention to target talkers, the identification of the target can be improved (Freyman et al, 1999, 2001; Li et al, 2004; Wu et al, 2005; Kidd et al, 2002; Freyman, 2004, 2006), interactions among different cues is definitely an important topic in search for the solution the the "cocktail party" problem. For example, Noble and Perrett reported that spatial cues are less important when other salient cues are used to segregate a signal from a masker (2002). One of our recent studies (Yang et al., 2007) has confirmed that in a simulated cocktail-party environment, familiarity or knowledge of target talker's voice is useful for reducing speech-on-speech masking (Yang et al., 2007). This study investigated whether the voice-cuing effect can be modulated by other cues such as the degree of familiarity/knowledge of target talker's voice and the speech onset asynchrony which is considered to be an effective cue for unmasking target speech.