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.

Methods

Five experiments were conducted in this study. Sixty young university students (19 – 27 years old) with normal audiograms (< 25 dB HL at test frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz) and with less than a 15 dB difference in threshold between the two ears at all testing frequencies participated in this research. Twelve participants were used for each experiment. Their first language was Mandarin Chinese. The participant was seated at the center of an anechoic chamber (Beijing CA Acoustics). Acoustic signals were presented to participants through a loudspeaker (Dynaudio Acoustics, BM6A), which was in the frontal azimuthally plane at 0° position (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). The sentence frame does not provide any contextual support for recognition of the key word. Target speech was spoken by one of the three young female talkers A, B and C in a trial. Seventy-two lists (24 list/talker and 18 sentences/ list) of nonsense sentences were used as target and priming sentences (Yang et al, 2007). Both target and priming were presented at the same level (52 dB SPL). Speech masking and speech-spectrum-noise masking were used. Two serials of Chinese nonsense sentences spoken by two other female talkers D and E were combined together as the speech masker whose content were different from target stimuli. A stream of steady-state Chinese speech babble noise with the duration of 10 s was obtained by mixing 113 female speech voices using Mat lab programming.

The study firstly investigated the role played by voice priming in releasing energetic and informational masking in the noise and speech masking conditions when target speech started 1 second after masking speech. In Experiment 1, three within-subject variables were used: (1) two masking conditions: noise masking and speech masking; (2) three priming conditions: no priming, same voice priming, repeated voice priming; (3) four different SNRs: -12, -8, -4 and 0 dB. Totally there were 24 (2×3×4) conditions for each listener and each condition included 18 trials. For the same voice priming condition, the prime and the target sentence were randomly selected from the sentences spoken by talker A, B or C, ensuring they were spoken by the same talker but without any content connection. In the repeated voice priming condition, one prime sentence was presented twice by the target voice, whose contents were different from that of the target sentence. According to the masker and prime type, the experiment was separated into six blocks which was completely balanced across 12 listeners, and the 4 SNRs were randomly arranged in each block.

In Experiment 2, the noise masking condition was excluded and a learning procedure on target voice was used. All other aspects (including the target and masker stimuli, the SNRs and the onset asynchrony) were identical to Experiment 1.

In the following studies, the onset asynchrony was excluded (masker and target started at the same time and ended at the same time).

In Experiment 3, the target and masker were presented simultaneously. Two within-subject variables were used: (1) Four priming conditions: same voice priming, no priming, different voice priming, and maker priming; (2) four SNRs: -12, -9, -6, and -3 dB.

In Experiment 4, the target and masker were presented simultaneously, three within-subject variables were used: (1) two masking conditions: noise masking and speech masking;(2) three priming conditions: no priming, same voice priming, repeated same voice priming; (3) four different SNRs: -12, -9, -6, and -3 dB.

In Experiment 5, the noise masking condition was excluded and a learning procedure on target voice was used. All other conditions (including the target and masker stimuli, the SNRs and the onset asynchrony) were identical to Experiment 4.

At the beginning of the experiment, participants were provided the details about the experiment tasks. Hearing "now the experiment is ready", the participant pressed the button of a response box to start the sound. They were instructed to try their best to repeat the target sentences loudly immediately after sounds were completed. The experimenter scored the 3 key words of every target sentence syllable by syllable outside the anechoic chamber. The number of correctly identified words was tallied later. A training procedure was conducted before the formal experiments to ensure participants to understand the experiment tasks well and be familiar with the experimental stimuli. Stimuli used in training were different from those used in formal experiments.

Results

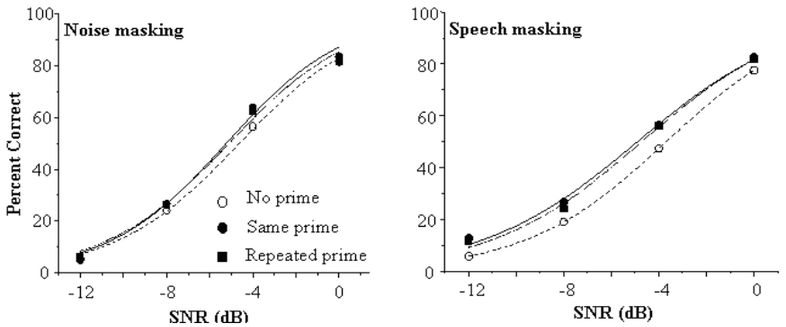

Figure 1. The group mean of recognition correct percent could be detected as a function of SNR under the speech-spectrum noise masking condition (left panel) or speech masking condition (right panel). Squares, repeated presentation of the same-voice prime; circles, single presentation of the same-voice prime.

Figure 1 shows the results for Experiment 1. The three curves were overlapped in the noise masking condition (left panel), indicating that there were little differences between the three prime types. Under speech masking conditions (right panel), when the onset asynchrony was 1 s, the curves of identification of the target following the same voice priming and repeated voice priming conditions were overlapped and they were both better than the identification in the no-priming condition.

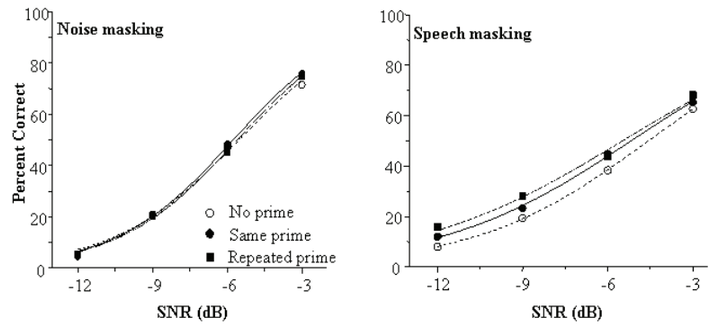

Figure 2. When the onset asynchrony was excluded, the group mean of recognition correct percent was detected as a function of SNR under the speech-spectrum noise masking condition (left panel) or speech masking condition (right panel). Squares, repeated presentation of the same-voice prime; circles, single presentation of the same-voice prime.

In Experiment 2, the voice-learning procedure was used. The results show that although participants became vary familiar with the target voice through the learning procedure, the single presentation of the same-voice prime and the repeated presentation of the voice prime produced similar improvement in speech recognition (single presentation: 0.911 dB; double presentations: 1.077 dB).

The results of Experiment 3 show that when the onset-asynchrony cue was excluded, introducing either the different-voice prime or the same-voice prime did not produce a significant effect.

Figure 2 shows results of Experiment 4. When the onset asynchrony was excluded, there was no significant difference among three different prime type conditions in noise masking. Under conditions of speech masking, introducing a repeated-voice prime produced a significant improvement in speech recognition, introducing the single presentation of the same-voice prime did not produce a significant effect.

The results of Experiment 5 show that when the learning procedure was used, either introducing the single or repeated presentations of the same-voice prime significantly improved speech recognition. Hence, when the learning procedure was used, there was a significant release from speech masking when the same voice or repeated-same-voice prime was used.

Summary

Voice priming cues can be used to reduce informational masking but not energetic masking when the cues are salient enough or when the other cues are co-presented. The voice-cuing effect can be modulated by the degree of familiarity/ knowledge of target talker's voice and the speech onset asynchrony. When target speech starts 1 second after masking speech, pre-presenting a priming sentence voiced by the target talker significantly improve the recognition of the target speech which is co-presented with masking speech. However, reinforcing the familiarity/knowledge of the target-talker's voice do not further improve the recognition. When target speech and masking speech start simultaneously, a single presentation of voice-priming speech do not change participants' speech recognition against speech maskers unless the familiarity/knowledge of target-talker's voice is reinforced by either a learning procedure or the repeated presentation of the target-talker's voice before testing. These results suggest that the voice-cuing effect on releasing speech from informational masking is graded, depending on both the degree of familiarity/knowledge of the target-talker's voice and the modulation by other cues such as speech-onset asynchrony. There must be various dynamic processes for integrating different cues to release informational masking. Different cues will be used at different central levels or under different conditions.

Reference

Cherry E. C., (1953). "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am. 25, 975–979.

Freyman, R. L., Balakrishnan, U., and Helfer K. S. (2001). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Amer. 109, 2112-2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Amer. 115, 2246-2256.

Freyman R. L., Balakrishnan U., and Helfer K. S. (2006). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Amer. 121, 1040-1046.

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," J. Acoust. Soc. Amer. 106, 3578-3588.

Kidd G. Jr., Mason C. R., and Arbogast T. L. (2002). "Similarity, uncertainty and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Amer. 111, 1367–1376.

Li, L., Daneman, M., Qi, J. Q., and 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.

Noble and Perrett. (2002). "Hearing speech against spatially separate competing speech versus competing noise," Perception & Psychophysics. 64 (8), 1325-1336.

Wu, X-H, Wang, C., Chen, J., Qu, H-W, Li, W-R, Wu, Y-H, Schneider, B. A, and Li, L. (2005). "The effect of perceived spatial separation on informational masking of Chinese speech," Hear. Res. 199, 1-10.

Yang Z. G., Chen J., Huang Q., Wu X. H., Wu Y. H., Schneider B. A., and Li L. (2007). "The effect of voice cuing on releasing Chinese speech from informational masking," Speech Communication, 49, 292–904.

THE EFFECT OF MASKER TYPE AND WORD POSITION ON IMMEDIATE SENTENCE RECALL

Payam Ezzatian[1], Liang Li[1, 2], M. Kathleen Pichora-Fuller1, Bruce A. Schneider[1]
[1]*Centre for Research on Biological Communication Systems, University of Toronto Mississauga*
[2]*Peking University*
payam.ezzatian@utoronto.ca, liang.li@utoronto.ca, k.pichora.fuller@utoronto.ca, bruce.schneider@utoronto.ca

Abstract

Noise maskers primarily result in energetic masking, whereas speech maskers create additional interference due to linguistic and acoustic similarities to the target (informational masking). Factors that facilitate stream segregation can greatly reduce the extent of informational masking. However, stream segregation often takes time to develop. In Experiment 1, nonsense sentences with 3 keywords were presented against a background of speech-spectrum noise or two-talker nonsense speech. With the speech masker, accuracy increased with word position. With the noise masker, accuracy did not vary systematically with word position. In Experiment 2, we noise-vocoded the speech masker using three bands to preserve envelope information while disrupting fine structure cues and minimizing semantic content. Here, performance was similar to that found with the noise masker. The results suggest that the ability to track a target sentence in conditions of informational masking improves as the target utterance unfolds over time.

When a masker and target overlap spectrally, the energy contained in the masker and target will activate similar regions along the basilar membrane, and the energy corresponding to the target stream can be completely or partially overwhelmed by the energy contained in the masker. This is referred to as energetic masking. When the competing sources are also speech, spectral and/or temporal fluctuations between the target and masker(s) provide brief glimpses of the target stream and can result in instances of reduced energetic masking. However, due to acoustic and semantic similarities to the target stream, speech maskers can produce interference beyond energetic masking, called informational masking (e.g. Freyman, Balakrishnan, Helfer, 2004; Li, Daneman Qi and Schneider, 2004; Schneider, Li, Daneman, 2007). This additional interference arises when the target and masking streams are confused with one another, or when the obligatory linguistic and semantic activation elicited by speech maskers interferes with the processing of the target stream (Schneider, Li, Daneman, 2007). Informational masking thus exerts its influence at a more cognitive level, making it difficult to segregate competing streams and attend to the target, whereas energetic masking mainly operates at a peripheral level by reducing target audibility. For example, it is quite common to experience difficulty understanding speech in the presence of background noise, such as ventilation or construction noise. However, it is quite unlikely that a listener will confuse such noises with the speech of a target talker. This is not the case for informational maskers. When competing streams are speech, similarities between the vocal characteristics of competing talkers can make it difficult to identify the target stream among competitors. Even when the target is perceptually isolated, the contents of the competitors may intrude into working memory making it more difficult to process the target stream. Generally, any acoustic features that distinguish competing streams facilitate stream segregation (Bregman,