THE EFFECT OF INFORMATIONAL MASKING AND WORD POSITION ON RECALL

Payam Ezzatian[1], Liang Li[1,2], Kathy Pichora-Fuller[1], and Bruce Schneider[1]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd North, Mississauga, Ontario, Canada L5L1C6
[2]Dept. of Psychology, Peking University, Beijing, China, 100871

1. INTRODUCTION

In order to perceive speech in noisy backgrounds, listeners need to perceptually separate the target stream from competing streams. Although competing streams such as construction noise may reduce target audibility (energetic masking), they are easy to distinguish from target speech because they are highly dissimilar to speech. Speech maskers on the other hand, can reduce target audibility as well as interfere with the processing of the target due to linguistic and acoustic similarities to target speech (informational masking, [1]). Because of these similarities, it may initially be difficult to perceptually segregate a speech signal from competing speech. Due to this initial delay in segregation, we might expect speech perception to improve as sentences unfold over time in situations where informational masking is prevalent. Informational masking may especially trouble older adults due to age-related declines in ability to take advantage of the F_0 difference between speakers' voices [2], or a reduced ability to take advantage of modulations in background maskers, thereby making stream segregation more difficult for older adults.

The current study examined age-related differences in the time course of perceptual streaming in an informational masking paradigm. In Experiment 1, younger and dder adults' ability to repeat sentences was measured in the presence of a two-talker speech masker (informational masking) or a speech-spectrum noise masker (energetic masking). In Experiment 2, the speech masker was noise-vocoded using 3 bands to determine the extent to which performance in the presence of speech masker is related to the fluctuations in the amplitude envelope of such maskers.

2. METHOD

2.1 Participants

Two independent groups of 16 younger and 16 older adults participated in each experiment. All participants were native English speakers, and had clinically normal audiograms (thresholds $\leq$ 25 dB HL) from .25 to 3 kHz, with interaural differences not exceeding 15 dB.

2.2 Stimuli

Target sentences were 208 nonsense sentences spoken by a female talker, e.g. "A house should dash to the bowl" [3]. In Experiment 1, a speech-spectrum noise masker was used in one half of the conditions, and a two-talker nonsense speech masker was used in the other half. In Experiment 2, the same speech-spectrum noise masker from Experiment 1 was used, but the two-talker nonsense speech masker from Experiment 1 was noise vocoded using 3 bands. The speech masker was noise-vocoded as follows: First, it was divided into 3 frequency bands using the following boundaries: 300, 814, 1528, and 6000 Hz. Then, the amplitude envelope was extracted in each band and used to modulate bands of noise having the same widths and center frequencies, thus creating a "vocoded" signal [4]. This procedure preserves amplitude envelope cues while removing fine structure cues. By using only 3 bands, the signal is largely unintelligible.

All stimuli were digitized at 20 kHz using a 16-bit Tucker Davis Technologies (TDT, Gainesville, FL) System II. The stimuli were converted to analog using the TDT system. The stimuli were then low-pass filtered at 10 kHz and amplified by a Harmon Kardon amplifier (HK 3370) prior to transmission via a single 40 watts loudspeaker. The loudspeaker was situated in the left far corner of a 9.3 x 8.9 x 6.5 foot Industrial Acoustic Company (Bronx, NY, USA) double-walled sound-attenuated chamber, and participants were seated at the center of the chamber facing the loudspeaker at a distance of 1.03 meters. Target sentences were presented at 60 dBA. Masker levels were adjusted to produce 4 signal-to-noise ratios (SNRs) for the younger adults: -12, -8, -4, and 0 dB, and 4 SNRs for the older adults: -8, -4, 0, and 4 dB. Sentence lists, SNRs, and masker types were counterbalanced across participants such that each list was presented in each masker and SNR condition an equal number of times.

2.3 Procedure

Prior to the start of each experiment, participants were familiarized with the task by being presented with one of the practice sentences ("A house should dash to the bowl") at the easiest SNR. The experimenter initiated the presentation of each sentence with the press of a keyboard button, which was followed immediately by the onset of background noise. Exactly 1 second later, the female target speaker uttered a target sentence, at the end of which the masker terminated as well. Participants were asked to repeat each target sentence, and were scored for 3 keywords (e.g., in "A rose can paint a fish", keyword 1 = Rose, keyword 2 = Paint, keyword 3 = Fish).

2.4 Analysis

The dB SNR required for 50%-correct performance was computed for each individual at each of the three keywords. These thresholds were then entered into an analysis of variance (ANOVA) to examine the effect of Age, Masker, and Word Position on performance.

3. RESULTS AND DISCUSSION

3.1 Experiment 1: Two-talker speech masker vs. speech-spectrum noise masker

As shown in Figure 1, younger adults outperformed older adults, and both groups performed worse with the speech masker than the noise masker in the background. The mean thresholds for both age groups improved in a linear fashion as a function of word position, but for the speech masker only. These observations were confirmed by 2 Masker by 3 Word Position ANOVA, with Age as a between-subjects factor, which revealed significant main effects of Age; $F(1, 30) = 7.12$, ($p = 0.01$), and Masker; $F(1, 30) = 23.43$, ($p < 0.001$), and Word Position on thresholds; $F(2, 60) = 4.23$, ($p = 0.019$). On average, older

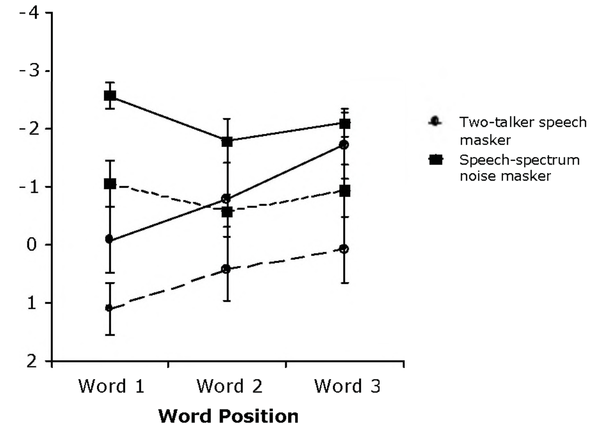

Fig. 1. Average thresholds from Experiment 1 for the younger adults (solid lines), and the older adults (dashed lines), are shown as a function of word position for the two-talker speech masker (squares), and speech-spectrum noise masker (circles) conditions. Error bars represent ± 1 SE.

adults required a 2.3 dB higher SNR to have the same performance as younger adults, and performance was an average 1.34 dB SNR better with the noise masker in the background (supporting the notion that speech maskers result in informational masking). Finally, overall performance was better for the third word, but the interaction of Word Position by Masker was significant; $F(2, 60) = 8.82$, $p < 0.001$, confirming the trend in the figure whereby performance improves over time with the speech masker, but not the noise masker. More importantly, the three-way interaction of Age by Masker by Word Position was not statistically significant, suggesting that the two age groups do not differ in the way in which they perceptually segregate target streams from background maskers.

3.2 Experiment 2: 3-band noise-vocoded speech masker vs. speech-spectrum noise masker

To confirm that the results from Experiment 1 are due to informational masking and not energetic masking, the speech masker from Experiment 1 was noise-vocoded using 3 bands (which removes content and fine structure cues) and tested in Experiment 2. The speech-spectrum noise masker was also tested in Experiment 2. Figure 2 plots the average performance for younger and older adults as a function of masker and word position. As can be seen from this figure, younger adults outperformed older adults, and both groups performed better with the noise-vocoded masker than with the speech-spectrum noise masker in the background. An ANOVA confirmed a main effect of Age; $F(1, 30) = 21.11$, ($p < 0.001$). On average, older adults required a 2.87 dB higher SNR to match the performance as younger adults. The interaction of Masker by Age was also significant; $F(1, 30) = 5.69$ ($p = 0.024$). Overall, younger adults improved by an average 1 dB more than older adults when going from the noise masker to the vocoded masker. Finally, there was a significant main effect of Masker; $F(1, 30) = 65.27$ ($p < 0.001$), and a significant interaction of Masker by Word Position; $F(2, 60) = 4.53$ ($p = 0.014$). Performance was an average 1.9 dB SNR better with the vocoded masker in the background, and the positive change in performance between Word 2 and 3 was greater for the speech masker.

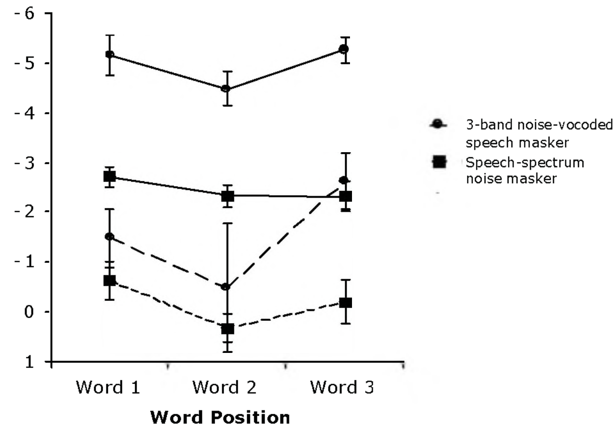

Fig. 2. Average thresholds from Experiment 2 for the younger adults (solid lines), and the older adults (dashed lines), are shown as a function of word position for the 3-band vocoded masker (squares), and speech-spectrum noise masker (circles) conditions. Error bars represent ± 1 SE.

4. DISCUSSION

Results from both experiments replicate previous findings of negative age-related differences when listening to speech in noisy background. Results from Experiment 1 show that for both younger and older adults, performance improves systematically with word position when the background consists of a speech masker, but not when it consists of a noise masker, indicating that stream segregation takes time to build up in informational masking situations. In Experiment 2, the pattern of improvement for the speech masker was no longer present once it was noise-vocoded. Older adults experienced less benefit from the noise-vocoded masker than did younger adults (possibly due to a decline in the ability to benefit from temporal fluctuations in a masker); however the pattern of performance as a function of Word Position was again equivalent for both groups, indicating that the time course of stream segregation is not affected by aging.

REFERENCES

[1] 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, 559-572.

[2] Vongpaisal, T., Pichora-Fuller, M.K. (2007). Effects of age on F0 difference limen and concurrent vowel identification. *Journal of Speech Language and Hearing Research*, 50, 1139-1156.

[3] Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech Language and Hearing Research*, 40(2), 432-443.

[4] Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wigonski, J. Boothroyd. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*, 107(5), 2704-2710.

ACKNOWLEDGMENTS

This project was funded by CIHR.