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,

1990) and can greatly reduce the extent of informational masking (e.g., spatial separation, familiarity with the speaker's voice, etc.). However, stream segregation often takes time to develop. Hence it would seem reasonable to expect that under conditions where potential cues that might aid stream segregation are sparse, streaming will develop gradually over time. The study reported here was designed to test this hypothesis. In two experiments, we examined the influence of target and masker similarity on the ability of listeners to repeat a target sentence. If stream segregation develops slowly over time we would expect word identification to increase as a function of word position. If, however, stream segregation develops rapidly, there should be very little improvement in performance as the sentence unfolds.

In Experiment 1, we presented participants with short nonsense sentences (e.g. "A house should dash to the bowl") against either a background of speech-spectrum noise, or a speech masker consisting of nonsense sentences spoken by two other talkers. To facilitate informational masking, the two talkers in the speech masker were of the same gender, age, and background as the target talker, and both the target speech and the two-talker speech masker were presented from the same loudspeaker. In Experiment 2, the two-talker speech masker was noise vocoded using 3 bands. This procedure removes the fine structure cues and semantic content from the speech masker, while preserving amplitude envelope information, and thus should result in reduced informational masking. If it is true that stream segregation develops slowly when the masker and target are acoustically and semantically similar, then we would expect performance in Experiment 1 to improve as target sentences unfold in the presence of the two-talker speech masker. However, this should not be the case for the speech-spectrum noise masker in either Experiment 1 or Experiment 2 where stream segregation is likely to occur more rapidly because the masker is both acoustically and semantically quite different from the target. We might also expect less of a word position effect in the presence of the three-band noise-vocoded speech masker because this masker, although it retains the amplitude envelope information in the speech masker, is also acoustically distinct from the target speech, and has no or minimal semantic content.

Participants, Materials, and Procedures

Sixteen college-aged adults participated in Experiment 1, and a separate group of 16 college-aged adults participated in Experiment 2. All participants were native English speakers, and had clinically normal audiometric thresholds in the speech range. Target sentences consisted of 208 nonsense sentences spoken by a female talker, e.g. "A frog will arrest the pit" (developed by Helfer, 1997). Two maskers were used in each experiment. 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. Experiment 2 was identical to Experiment 1 with the exception that the two-talker nonsense speech masker from Experiment 1 was noise vocoded using 3 bands. Noise vocoding is a process in which the signal is partitioned into different frequency bands, the amplitude envelope is extracted in each of these bands, and a vocoded signal is created by using these amplitude envelopes to modulate bands of noise having the same widths and center frequencies used in the original partitioning of the spectrum (Eisenberg, Shannon, Martinez et al., 2000; Shannon, Zeng, Kamath, Wygonski, et al., 1995). This procedure preserves amplitude envelope cues while eliminating fine structure cues. Intelligibility of noise-vocoded speech increases as the number of available bands is increased (Shannon et al., 1995). In the current experiment, we vocoded the speech masker using the following frequency boundaries: 300, 814, 1528, and 6000 Hz. All stimuli were digitized at 20 kHz using a 16-bit Tucker Davis Technologies (TDT, Gainesville, FL) System II and custom software. The stimuli were converted to analog

using the TDT system. The stimuli were then low-pass filtered at 10 kHz, amplified by a Harmon Kardon amplifier (HK 3370), and transmitted 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 at a distance of 1.03 meters from the loudspeaker. Participants faced the loudspeaker at 0^0 azimuth.

Target sentences were divided into 16 lists containing 13 target sentences each. Four additional sentences were included as practice sentences and added to the beginning of each list in a random order. Target sentences were presented at 60 dBA. Masker levels were adjusted to produce 4 signal-to-noise ratios: -12, -8, -4, and 0 dB. Signal-to-noise ratios remained constant throughout the presentation of a single list, but sentence lists and signal-to-noise ratios were counterbalanced across participants such that each list was presented at each of the 4 different signal-to-noise ratios an equal number of times.

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 signal-to-noise ratio. The experimenter who sat outside the sound-attenuated chamber initiated the presentation of each sentence with the press of a keyboard button. This button press was followed immediately by the onset of background noise. Exactly 1 second after the onset of the background masker, the female target uttered a target nonsense sentence, at the end of which the masker terminated as well. Participants were asked to repeat back the target nonsense sentences after each presentation, and were scored online by the experimenter for 3 keywords in each target sentence (e.g. in "A rose can paint a fish", keyword 1 = Rose, keyword 2 = Paint, keyword 3 = Fish).

Analysis and Results

Psychometric functions were computed using the percentage of correctly identified keywords (13 keywords at each of three word positions) at each signal-to-noise ratio using:

$$y = \frac{(1 - a)}{1 + e^{-\sigma(t - \mu)}} \qquad (1)$$

Where y is the probability of correctly identifying a keyword, (1 – a) determines the asymptote of the psychometric function, t is the signal-to-noise ratio, μ corresponds to the signal-to-noise ratio for 50%-correct identification (threshold), and σ corresponds to the slope of the psychometric function. Figure 1 plots 50%-correct thresholds as a function of word position for the speech-spectrum noise masker, and the two-talker speech masker in Experiment 1. As can be seen from this figure, overall performance is better for the speech-spectrum noise masker than the two-talker speech masker. But more importantly, when the background consists of competing speech, performance improves systematically with word position. However, this is not the case when the background consists of speech-spectrum noise. This pattern of results was confirmed by a 2 Masker by 3 Word position within-subjects Analysis of Variance (ANOVA). The ANOVA revealed a significant main effect of Masker on thresholds, F (1, 15) = 7.49, p = 0.015. Thresholds were an average 1.29 dB higher with the two-talker speech masker in the background than with the speech-spectrum noise masker in the background. The main effect of Word position was not significant (p = 0.084). However, there was a significant interaction between Masker and Word position: F (2, 30) = 5.81, p = 0.007. A Student Neuman Kuels test of multiple comparisons showed that thresholds were statistically equivalent at each word position when the masker was speech-spectrum noise. However, the average threshold for Word 3 with the two-talker masker in the background was significantly lower than the threshold for Word 1 ($p < 0.01$), and Word 2 ($p < 0.05$). Thresholds for Word 1 and Word 2 were statistically equivalent. An ANOVA examining the effect of Masker and Word position on the slopes of the psychometric function did not yield significant results.

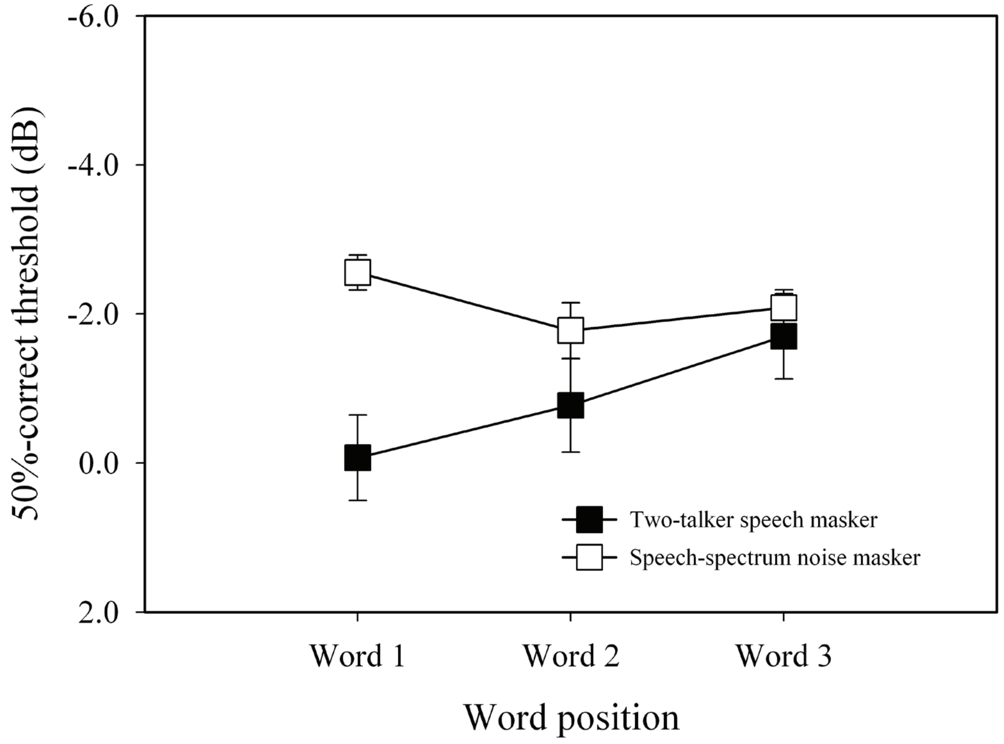

Figure 1. Average thresholds from Experiment 1 are shown as a function of word position for the two-talker speech masker (filled squares) and speech-spectrum noise masker (unfilled squares) conditions. Threshold values are plotted in reverse on the ordinate. Error bars represent ± 1 standard error.

Results from Experiment 2 were analyzed using the same procedures used for Experiment 1, and are plotted in Figure 2. As can be seen from this figure, overall performance was much better with the 3-band noise vocoded speech masker in the background than the speech-spectrum noise masker. What's more, the pattern of change in performance as a result of word position no longer resembles that of the intact two-talker speech masker. To examine these results, the data were analyzed using a 2 Masker x 3 Word position within-subjects ANOVA. As expected, there was a main effect of Masker on thresholds (F (1, 15) = 150.97, $p < 0.001$). Thresholds were an average 2.52 dB lower in the 3-band noise vocoded condition than the speech-spectrum masker condition. More importantly, the interaction of Masker by Word position was not significant. Finally, the effect of Word position on thresholds was significant, F (2, 30) = 3.74, $p = 0.036$. On average, thresholds for word 1 were 0.53 dB lower than those for word 2 ($p = 0.031$); however differences between word 1 and word 2, and word 2 and word 3 were not statistically significant. An analysis of the effect of Masker and word position on psychometric slopes showed a significant main effect of Masker [F (1,15) = 92.27, $p < 0.001$], and a significant main effect of Word position [F (2, 30)= 4.05, $p = 0.028$]. On average, slopes were higher for the 3-band noise vocoded speech masker, and higher for Word 2 than for Word 3.

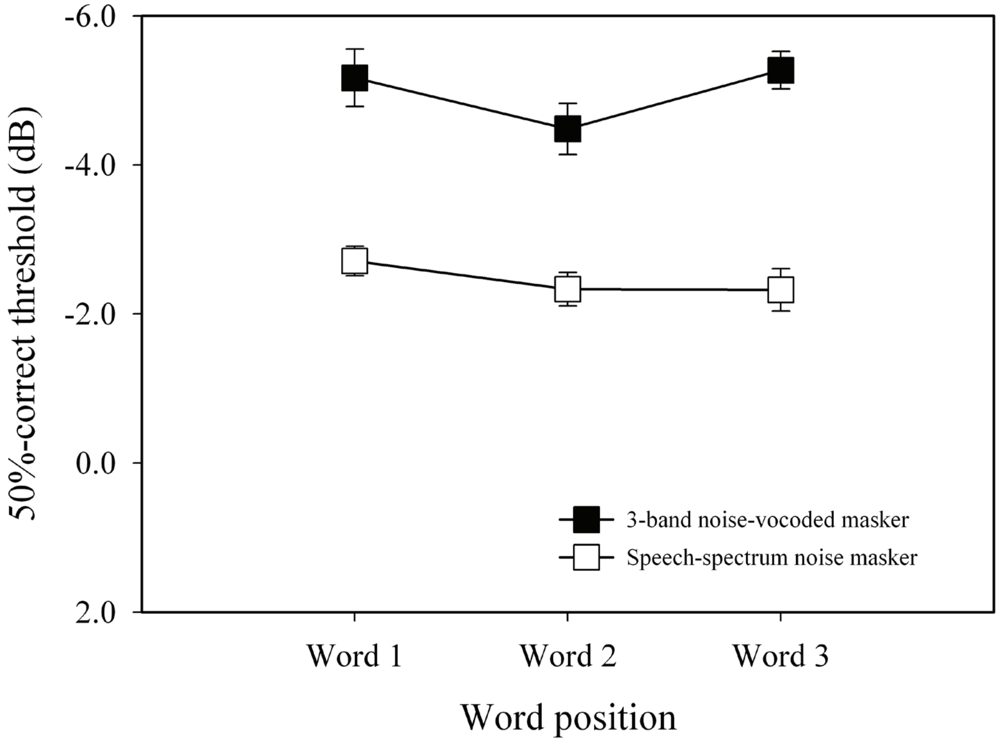

Figure 2. Average thresholds from Experiment 2 are shown as a function of word position for the 3-band noise vocoded speech masker (filled squares) and speech-spectrum noise masker (unfilled square) conditions. Threshold values are plotted in reverse on the ordinate. Error bars represent ± 1 standard error.

Discussion

The purpose of this study was to examine the time course of perceptual streaming in an informational masking situation. In such a situation the listener has to segregate the target sentence from the two-talker masker. If the degree of perceptual segregation in an informational masking situation improves over time, we would expect the identification of target words to improve as the nonsense sentence unfolds. The results of Experiment 1 confirmed that when the background consisted of two talkers, word identification improved as a function of word position. This effect was absent however, when the masker was speech-spectrum noise. In Experiment 2, we noise-vocoded the two-talker speech masker (using 3 frequency bands) to preserve envelope information while disrupting fine structure cues and minimizing semantic content. Vocoding the speech masker in this way eliminated the word position effect while significantly improving overall performance. These results indicate that the ability to track a target sentence in conditions of informational masking improves as the target utterance unfolds over time.

References

Bregman, A.S., (1990). *Auditory Scene Analysis: The Perceptual Organization of Sounds.* The MIT Press, London, England.

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.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America, 115*(5), 2246-2256.

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.

Li, L., Daneman, M., Qi, J. G., & 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*(6), 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*, 559-572.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

PLATEAU´S EXPERIMENTS REVISITED

Francisco J. Sanchez-Marin
Centro de Investigaciones en Optica
sanchez@cio.mx

Abstract

Joseph Plateau's pioneering experiments on the perception of contrast were replicated. The experiments were not done exactly as Plateau did. The main difference in our design was that, we made use of a computer to have the participants to generate grey scales using the bisection method. Our program interacted with each participant in such a way that he/she could generate a grey scale using the grey levels that were generated in the previous step. The experiments were done first with a black background and then with a white background, and with and without linearizing the computer monitor. Subjects with professional training for handling color as well as non trained subjects were invited to participate. The generated grey scales were analyzed with the help of a computer. Our results show that training is an important factor in contrast perception and that linearizing the computer display does not help the observers in the generation of a grey scale.

Plateau, in his classical experiment (Laming and Laming, 1996; Plateau, 1872a; Plateau, 1872b), apparently done more than ten years before the publication of Fechner's Elements, tried to determine if we humans have the ability of precisely estimating the intensity of visual stimuli. The method that he used is now known as the bisection method (Heller, 2001), or method of equal appearing intervals (Murray, 1993) given that he asked to professional painters to paint a gray whose intensity (i.e. brightness) should be midway between white and a given black. According to Plateau, the participant painters arrived to almost the same tone of gray even though they all used different levels of illumination. That is why he assumed that the perception of contrast differences was not affected by illumination. However, after Plateau reviewed the work of J. Delboeuf (Plateau, 1872b), who was a contemporary Belgian researcher that followed his steps, he arrived to the conclusion that illumination, in fact, played a role in the perception of contrast.

On the other hand, there are evidences that suggest that perceptions of brightness are generated empirically by experience with luminance relationships. According with this, it makes sense Plateau's decision of asking professional painters to participate in his experiments. However, Plateau did not investigate how experience (i.e. training) influenced his results.

In this work are presented some results that were obtained by somehow replicating those pioneering experiments, but with the help of a computer and with subjects with professional experience in handling colour and non-experienced subjects on this matter. Our results show that training influences contrast perception, and that the use of different backgrounds clearly affects brightness perception in both trained and untrained subjects.

In visual perception experiments is a common practice to linearize the devices used to display the visual information (i.e. the computer monitors) so that the non-linear relationship between the applied voltage in the monitor circuitry and the displayed intensity on the monitor screen is corrected. However, an interesting point is that the results obtained in our experiments with a linearized computer monitor were not better that those obtained without linearizing the monitor.