THE EFFECT OF AN EMOTIONAL CARRIER PHRASE ON WORD RECOGNITION

Dario Coletta[1], Kate Dupuis[1], and M. Kathleen Pichora-Fuller[1, 2]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd North, Mississauga, Ontario, Canada
[2]Toronto Rehabilitation Institute, 550 University Ave, Ontario, Canada

1. INTRODUCTION

The ability to understand emotion in speech is crucial for successful communication. Emotional information can be transmitted both through the words a talker uses and through affective prosody (i.e., the acoustic cues in speech used to express a talker's emotional state). Although numerous studies have examined how well listeners can recognize affective prosody, and have identified the acoustical parameters which are modified when different emotions are expressed, little is known about how emotion affects a listener's ability to understand speech in noise.

1.1 Intelligibility of emotional speech

Traditional speech intelligibility tasks have used stimuli which were recorded in neutral voices; however, these are not representative of real-world communication. This may explain, at least in part, the poor correspondence between performance on speech tests and listeners' subjective reports of their everyday communication difficulties [1]. Emotion has yet to be examined in the context of speech intelligibility. Emotion could potentially exert an influence on intelligibility through a form of cognitive processing, such as an attentional mechanism and/or by modifying acoustical cues produced when a target word is spoken.

1.2 Current experiments

These experiments were designed to test the cognitive hypothesis by examining if emotion could influence word recognition when the acoustical properties of the target word were unaltered. In order to determine whether emotion could potentially improve word recognition in noise, the presence of emotional prosody was manipulated such that it was only present in the carrier phrase, with a neutral target word either following (Experiment 1) or preceding (Experiment 2) the carrier. A final set of analyses compared results from these two experiments.

2. EXPERIMENTS

2.1 Experiment 1

Method

Twenty-eight participants were tested in each of the two experiments. All participants were University of Toronto undergraduates in good health who had clinically normal hearing thresholds in the speech range (see Table 1 for participant characteristics). The stimuli used were taken from the Toronto Emotional Speech Set (TESS, [2]). In this set, 200 items from the NU-6 test [3] were recorded by two female actors, one 26 years old and the other 64 years old, to portray seven different emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral), for a total of 2800 stimuli. Each NU-6 item consists of the carrier phrase "Say the word" followed by a target word, which is a unique monosyllabic noun (e.g., "Say the word *dog*"). In the current experiments, the stimuli spoken by the younger actor were modified such that the carrier phrase spoken by her with emotion preceded each neutral target word which was spoken by a male in a commercially available recording of the NU-6 test (Auditec, St. Louis). The stimuli were equated for RMS and presented in the multi-talker babble background from the Speech Perception in Noise Test [4]. In order to determine the presentation level, a pilot study was conducted using only stimuli spoken in a neutral voice (similar to the neutral stimuli used in typical intelligibility tasks). A 75% identification rate was achieved with an SNR of -5 dB. It is common for the NU-6 test items to be divided into 4 half-lists of 25 items. Seven half-lists were used in this experiment while the eighth was used for practice. Participants were instructed to repeat the target word.

Table 1. Demographic information (Means and SDs) for participants in each of the two experiments.

	Exp 1 (N=28)	Exp 2 (N=28)
Age (years)	18.50 (1.17)	20.80 (2.14)
Education (years)	13.18 (1.59)	15.00 (1.74)
Vocabulary (out of 20)	12.39 (2.93)	12.30 (2.53)
Health (1-4)	3.32 (0.55)	3.30 (.60)

Results

The measure of interest for all the analyses reported in this paper was the number of target words correctly repeated by each participant. Results from an ANOVA indicated that emotion had a significant influence on performance, $F(6,126) = 2.55$, $p = .02$. Specifically, sad (67%) and pleasant surprise (66%) were the emotions that yielded the best intelligibility while fear (61%) was the emotion that yielded the worst intelligibility.

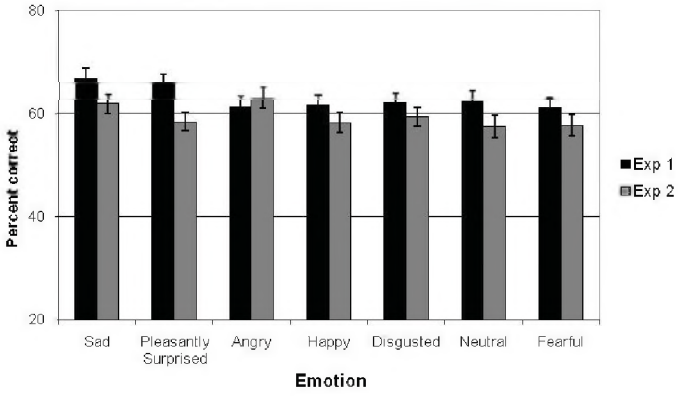

Figure 1. Mean percent accuracy for word recognition with each of the seven emotions in each of the two experiments.

Discussion

Results indicate that emotion did have a significant influence on how well listeners were able to understand the target word in a sentence. However, it is not known whether simply having part of the sentence portrayed in an emotional voice would improve performance, or whether the emotion must precede the neutral target. This question was examined in the following experiment.

2.2 Experiment 2

Methods

The stimuli used were identical to those used in the previous experiment; however, Praat software [5] was used to move the neutral target word from the end of the sentence (e.g., "Say the word dog") to the beginning of the sentence (e.g., "Dog, say the word"). Testing procedures were identical to those used in the preceding experiment.

Results

Results from an ANOVA indicated that emotion did not have an effect on participants' ability to correctly identify the target word, $F(6, 126) = 1.58$, $p = .16$.

Discussion

In this experiment, having the emotional carrier phrase follow the neutral target word eliminated the effect of emotion. This suggests that the presence of emotion at any point in the sentence is not sufficient to improve intelligibility; the emotion must precede the target word.

2.3 Comparison between Experiments 1 and 2

Methods

In order to compare results from the two experiments, an ANOVA was run with emotion as the within-subjects factor and experiment as the between-subjects factor.

Results

Overall, the emotional carrier phrase presented prior to the neutral target word (Experiment 1) resulted in more accurate recognition of the target word. This description was confirmed by the ANOVA which revealed that the mean accuracy rates for participants in Experiment 1 were significantly higher ($M = 63.02\%$, $SD = 1.26$) than for participants in Experiment 2 ($M = 59.37\%$, $SD = 1.26$), $F(1,42) = 4.20$, $p = .047$. Although there was a main effect of emotion, $F(6, 252) = 2.44$, $p < .01$, the analysis failed to reveal an emotion by experiment interaction, $F(6, 252) = 1.54$, $p < .01$. This suggests that participants responded to emotions in the two experiments in a similar manner.

Discussion

The comparison of results from the two experiments indicates that, overall, participants identified a larger number of target words when the emotion preceded the target word. The influence of emotion on performance was similar across experiments.

3. GENERAL DISCUSSION

In these experiments, the effect of emotion on performance was only present when the participants heard the carrier phrase spoken with emotion before hearing the target word. Thus, the emotion in the carrier phrase had a prospective effect on subsequent word recognition. In contrast, when participants were presented with the neutral target word initially, the subsequent emotion depicted in the carrier phrase had no retrospective effect on performance. Participants may have simply focused on remembering the target word throughout the presentation of the emotional carrier phrase, rendering the emotion moot. Accuracy was highest to the portrayals of sadness, which may be the most acoustically similar to neutral, causing the sad carrier to serve as a prime for the neutral target word. The high accuracy for portrayals of pleasant surprise may be due to the positive and rewarding nature of this emotion, which may have caused listeners to orient to this emotion. In contrast, the low accuracy for portrayals of fear may reflect the fact that participants chose to orient away from this negative emotion

It is important to note that there was a discontinuity between the male-spoken neutral word and the female-spoken carrier phrase which may have influenced the participant's ability to correctly identify the target word. Future studies should use target words spoken in a neutral voice by the TESS talkers to control for this discontinuity.

The results from these experiments relate to the ways in which individuals communicate in the real world; it is very rare to hear speech devoid of some sort of affective prosody. It may be that individuals are much more responsive to the identity of a target utterance based on the emotion that was present prior to and/or during pronunciation.

REFERENCES

[1] Taylor, B. (2007). Predicting real world hearing aid benefit with speech audiometry: An evidence-based review. *Audiology Online, 1802.* Retrieved from http://www.audiologyonline.com/articles/article_detail.asp?article_id=1802

[2] Dupuis, K., & Pichora-Fuller, M. K. (2010). Toronto Emotional Speech Set (TESS). Available from https://tspace.library.utoronto.ca/handle/1807/24487

[3] Tillman, T. W., & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6. Technical report no. SAM-TR-66-135. San Antonio, TX: USAF School of Aerospace Medicine, Brooks Air Force Base.

[4] Bilger, R. C., Nuetzel, M. J., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research,* 27, 32-48.

[5] Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.5.08) [Computer program]. Retrieved from http://www.praat.org/