EFFECTS OF EMOTIONAL CONTENT AND EMOTIONAL VOICE ON SPEECH INTELLIGIBILITY IN YOUNGER AND OLDER ADULTS

Kate Dupuis[1] and Kathy 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

Everyday speech contains emotional information which is conveyed through emotional inflection. (e.g., a disgusted tone of voice) and through the lexical content of the message (e.g., "I hate your dress."). Recent advances have been made in determining the exact nature of emotional representations in speech (e.g., [1]), and in describing the numerous methodologies which have been used to examine understanding of emotion based on both content and tone (e.g., [2]). Standardized speech tests are used extensively by clinicians and researchers to measure speech perception and spoken language understanding in different listening conditions and in people of all ages. However, the emotion represented in the tone and content of most speech intelligibility test materials seems to be either minimized or poorly controlled. Speech stimuli are typically recorded in an artificially neutral way, devoid of any affective cues. Additionally, the lexical content of many of the stimuli have emotional connotations which are not controlled for or equalized across lists. For example, the sentence "The airplane went into a dive." in the Speech Perception in Noise test [3] and the sentence "The young people are dancing." in the Hearing in Noise Test [4] can have negative and positive connotations, respectively.

Literature in the visual and audio-visual domains suggests that printed words or spoken sentences with an emotional meaning are more efficiently recognized and recalled better than stimuli with neutral meaning [5, 6]. Indeed, databases of words, pictures, and sounds have been developed to provide normative ratings to allow researchers to equate stimuli on the emotional characteristics of arousal and valence [7, 8, 9]. Nevertheless, most commonly-used speech tests were developed before the advent of such databases. The effect of the emotional implications of an utterance's emotional content and/or affective tone on intelligibility has not been controlled for or investigated. In addition, the interpretation of emotional cues in speech differs depending on age [10, 11], and it may be a factor influencing older adults' reported difficulties understanding speech in noise (e.g., [12]). Thus, performance on speech intelligibility tests may be influenced by an interaction of age with emotional content and/or tone, but this possibility has not been examined.

The first purpose of the current study was to determine the emotional valence and emotional arousal properties of 200 words spoken in a neutral tone of voice which were taken from a standardized test (Northwestern University Auditory Test No. 6; NU6 [13]). Each target word is presented in the carrier phrase "Say the word...". The relationship between the emotional ratings and the ability of listeners to identify the target words under varying signal-to-noise conditions was then examined; pilot data from younger adults will be described in the present paper. The secondary purpose of the study was to create an emotionally-spoken version of the NU6 words so that the effect of emotional speech production on intelligibility could be tested. A younger and an older female actor each re-recorded the 200 NU6 words in seven emotional tones of voice (happy, sad, neutral, angry, fearful, pleasantly surprised and disgusted). Numerous tokens of each stimulus were recorded and three raters chose the token which they felt best represented each emotion. The ratings of the three judges were compared to analyze inter-rater reliability. A final set of 2800 stimuli (200 sentences x 7 tones of voice x 2 speakers) will be presented to both younger and older listeners in subsequent studies to determine the accuracy with which each portrayed emotion can be identified and how emotion influences speech intelligibility.

2. EXPERIMENTS

2.1 Experiment 1

2.1.1 Method

To date, twenty-eight younger adults (mean age = 19.9 years, *SD* = 2.5) with good health and clinically normal hearing thresholds in the speech range have been tested. The stimuli used were the 200 sentences from the NU6 lists. These stimuli were presented to one group (N=12) visually as text on paper, and to two groups auditorally, through two loudspeakers in a sound-attenuating booth. In the auditory conditions, one group (N=9) heard the sentences presented by a recorded female voice [14] and the other group (N=7) heard presentations by a recorded male voice [15]. In all cases participants rated emotional valence from 1-9 (1 was "extremely negative"; 9 was "extremely positive") and emotional arousal from 1-9 (1 was "least arousing"; 9 was "most arousing"), either by circling the appropriate number in the visual conditions or by pressing the appropriate box on a touch-screen in the auditory conditions. The participants were tested individually and provided ratings for each of the 200 words.

2.1.2 Results

Means were obtained for both valence and arousal ratings for all three groups (visual, auditory female voice and auditory male voice). These means were then correlated with ratings of frequency, familiarity, and neighbourhood density for each word, as well as the mean level (in dB SNR) at which participants could reliably identify each word 50% of the time. The SNR threshold data were collected on young adults with normal hearing by Richard Wilson and colleagues using the female voice. Analyses revealed a significant positive correlation between valence and arousal for participants in the visual condition, $r = .314$, $p < .001$, and auditory female voice condition, $r = .360$, $p < .001$. However, participants who listened to the stimuli spoken by the male voice exhibited a negative correlation between valence and arousal, $r = -.143$, $p = .044$.

Furthermore, negative correlations were found between arousal rating and mean SNR threshold level only for the visual presentation group, r = -.153, p = .031. Neither valence nor arousal ratings correlated with ratings of word frequency, familiarity or neighbourhood density in any of the three conditions.

2.1.3 Discussion

This experiment was the first to our knowledge to gather emotional valence and arousal ratings for the target words of a test commonly used in speech audiometry. The results indicate that the emotional arousal of listeners to a particular word can affect intelligibility, depending on the modality of presentation. More arousing words can be reliably identified at lower SNR than less arousing words. However, this finding of a significant correlation between emotional rating and the SNR threshold for words only holds true for the emotional rating of words which were visually presented to participants. This modality-specific finding likely reflects the success of the auditory test developers in achieving recordings (in both male and female voices) which minimized the emotional response of listeners who rated the words. In contrast, those who read the words were free to react with an unrestricted emotional response during rating. The neutrality of tone of voice could, nevertheless, influence peoples' everyday understanding of the content of the word, as very rarely is natural speech devoid of emotional inflection. These data demonstrate the strong interaction between emotional tone and content in spoken language, but highlight that this aspect of natural communication has not been tapped by traditional speech tests which have been more focused on de-contextualized phonemic and lexical aspects of speech perception.

2.2 Experiment 2

2.2.1 Method

A younger (aged 26 years) and an older (aged 64 years) female actor were recruited from the community and consented to create voice recordings on separate occasions. Both actors spoke English as a first language and had clinically normal hearing thresholds in the speech range. Each actor recorded numerous tokens of 1400 sentences (200 stimuli x 7 emotional tones of voice -- happy, sad, neutral, angry, fearful, pleasantly surprised and disgusted) and instructions were given in an attempt to equate production styles across actors. The actors were given the 50 target stimuli from each of the four NU-6 lists on paper and were asked to speak each word in the carrier phrase "Say the word..." and to produce each item multiple times. The experimenter was present and directed the recording session. Tokens of each word spoken in each emotion were produced until the actors and experimenter were satisfied with the production of each word in each emotion.

2.2.2. Results

Using the Praat [16] acoustical analysis software, each token was saved as a separate sound file. Three English-speaking undergraduates listened to these stimuli and chose what they considered to be the best representation of the target emotion for each sentence from the set of tokens. Agreement was over 90% for at least two raters across all sentences.

2.2.3 Discussion

With knowledge that the emotional tone of presentation can have an effect on spoken language understanding and that this effect interacts with a listener's age, a novel battery of stimuli has been developed. In this battery, the 200 stimuli of the NU-6 test have been re-recorded by both a younger and an older female actor in seven different tones of voice. Agreement on the specific tokens which are the best representations of a particular emotion is high. The creation of these materials is the first step towards creating an emotional-tone speech intelligibility task which will enable us to determine whether the way in which a sentence is presented (e.g., in a sad or fearful tone of voice) will affect how well a listener can perceive and understand it.

REFERENCES

[1] Cowie, R. & Cornelius, R.R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40,* 5-32.

[2] Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 30,* 227-256.

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

[4] Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoustical Society of America, 95,* 1085-1099.

[5] Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5,* 394-400.

[6] Cahill, L. & McGaugh, J.L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition, 4,* 410-421.

[7] Bradley, M.M., & Lang, P.J. (1999a). *Affective norms for English words (ANEW): Instruction manual and affective ratings.* Technical Report C-1,

[8] Ito, T.A., Cacioppo, J.T., & Lang, P.J. (1998). Eliciting affect using the International Affective Picture System: Bivariate evaluation and ambivalence. *Personality and Social Psychology Bulletin, 24,* 856–879.

[9] Bradley, M.M., & Lang, P.J. (1999b). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Technical Report B-2).* Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

[10] Kiss, I., & Ennis, T. (2001). Age-related decline in perception of prosodic affect. *Applied Neuropsychology, 8,* 251-254.

[11] Orbelo, D. M., Testa, J. A., & Ross, E. D. (2003). Age-related impairments in comprehending affective prosody with comparison to brain-damaged subjects. *J Geriatric Psychiatry and Neurology, 16,* 44-52.

[12] Committee on Hearing, Bioacoustics and Biomechanics. (1988). Speech understanding and aging. *J Acoustical Society of America, 83,* 859-895.

[13] 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.

[14] Department of Veteran Affairs (1991). *Speech recognition and identification materials, Disc 1.1 (CD).* Auditory Research Laboratory V. A. Medical Center, Long Beach, CA.

[15] Wilson, R. H., Coley, K. E., Haenel, J. L., & Browning, K. M. (1976). Northwestern University Auditory Test No. 6: Normative and comparative intelligibility functions. *J American Audiology Society, 1,* 221-228.

[16] P. Boersma and D. Weenik, "Praat: Doing phonetics by computer (version 4.5.08)" [Computer program], 2006 Apr 4 [cited 2008 Jun 29], Available HTTP: http://www.praat.org/