EMOTION CO-EXISTS WITH LEXICAL EFFECTS: A CASE STUDY

Tatiana V. Kryuchkova[1], Benjamin V. Tucker[2]
[1]Dept. of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Alberta, Canada, T6G 2E7, kryuchko@ualberta.ca
[2]Dept. of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Alberta, Canada, T6G 2E7, benjamin.tucker@ualberta.caa

1. INTRODUCTION

In speech production many lexical factors influence the way in which words are produced. Munson and Solomon (2004) find that in laboratory speech shorter vowel duration is associated with larger neighborhood density and lower lexical frequency. While lexical factors have been shown to affect the way in which speech is produced, emotion has also been shown to affect acoustic features of speech, such as vowel and word duration, fundamental frequency, and intensity (e.g. Zupan et al., 2009).

With emotion (joy, fear, anger, happiness, etc.) being an important part of daily communication, emotional and lexical acoustic variation may come into conflict such that during emotionally produced speech lexical effects found in laboratory speech may not be present. The present work develops the results of previous work (Kryuchkova & Tucker, 2011), which found a robust effect of neighborhood density on word duration in production of acted emotional speech. This effect interacted with lexical Frequency and Emotion. In the present study we further examine the possible interaction of emotion and the lexicon by investigating effects of neighborhood density and lexical frequency for words and also at the vowel level.

2. EXPERIMENT

2.1 Materials

The stimulus set comprises 260 real words of English taken from Wurm & Seaman (2008), for which counts of lexical frequency, neighborhood density, and morphological family size were available. Two professional male actors (T and D) recorded the stimuli in five emotional modalities: neutral, anger, joy, content, and fear. Participants read items from a list: fillers were added at the beginning and end to account for list intonation. Each actor recorded the stimuli over two sessions. The recordings of three emotional modalities: anger, joy and neutral - were annotated and acoustic measures extracted using PRAAT (Boersma & Weenink, 2010). Mean vowel fundamental frequency (F0) and vowel duration were extracted for each word. Measurements for the mean word F0 and word duration were taken from the previous study and mean intensity for words and vowels was extracted.

2.3 Methods

The statistical analysis was performed using linear mixed effects regression modeling (*lme4*, Bates, Maechler, Bolker, 2011) with words as random effects and duration, mean F0, and mean intensity as dependent variables. Markov chain Monte-Carlo sampling (10,000 simulations) was used to estimate reported p-values (*languageR*, Baayen 2010). First, the dependent variables were modeled individually as a function of Emotion and Speaker to check whether the acoustic characteristics of the recorded speech were indeed modified by the emotional modality. Second, the dependent variables were modeled as a function of the predictors lexical Frequency (logged) and Neighborhood Density; the model also accounted for possible interactions of the two lexical predictors with Emotion and Speaker. The count for Number of Syllables and inherent vowel length (tense vs. lax) were also included in the model as control variables.

3. RESULTS

3.1 Emotion related acoustic variation

All six dependent variables are affected by the emotional modalities portrayed by the actors. The pattern of the effects for mean Intensity and mean F0 are similar at the word and vowel level (Figure 1, top panels)

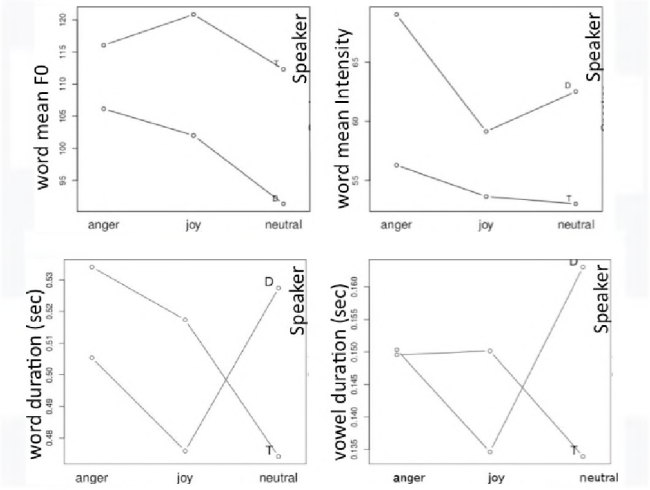

Figure. 1. Effects of Emotion and Speaker on F0, Intensity and duration.

The effect pattern for Duration is different for T at the word and vowel levels. For T anger has the longest word durations of all three emotional modalities, but at the vowel level T's anger and joy have similar vowel durations (Figure 1, bottom panels).

The results of this stage of analysis confirm that in the current dataset the target acoustic features (mean F0, mean Intensity and duration) are modified by the acted emotional modality both at the word and vowel levels. The patterns of variation are similar but not identical.

3.2 Lexical effects

No significant effects were found for intensity measures.

For vowel duration, Neighborhood Density significantly interacts with Speaker ($p<0.005$): higher Neighborhood Density is associated with longer duration; the slope of the effect is steeper for D. The effect of Neighborhood Density on vowel duration does not interact with lexical Frequency and lexical Frequency is not predictive of vowel duration. For word duration, the effect of Neighborhood Density interacts with Emotion: joy and neutral show a slightly steeper slope than anger, but for all three emotions larger Neighborhood Density is associated with shorter durations ($p<0.01$). For words, the effect of Neighborhood Density also significantly interacts with lexical Frequency ($p<0.005$), so that for words with lower lexical frequency larger Neighborhood Density is associated with shorter durations, but for words with high lexical frequency the effect disappears.

Neighborhood Density has a significant positive effect on the mean vowel F0 ($p<0.05$), with larger Neighborhood Density associated with higher F0. A similar effect of Neighborhood Density is observed at the word level ($p<0.054$). In both cases Neighborhood Density interacts with Speaker: T has overall higher mean F0 and a steeper slope for the effect. Lexical Frequency is also predictive of the mean word and vowel F0. At the word level, lexical Frequency interacts with Neighborhood Density ($p=0.0002$): for low frequency words the effect of Neighborhood Density is positive: larger Neighborhood Density is associated with higher mean word F0. For high frequency words the effect of Neighborhood Density on word duration reverses: larger Neighborhood Density in high frequency words is associated with lower F0. Neighborhood Density interacts with Vowel Type ($p=0.03$) for mean vowel F0: thus for lax vowels higher mean F0 is associated with larger Neighborhood Density, for tense vowels higher mean F0 is associated with smaller Neighborhood Density.

4. DISCUSSION

Our dataset was comprised of words recorded by two professional actors simulating anger, joy, and neutral emotion. To convey emotion, in this case study, both actors employed slightly different manipulations of duration, mean F0, and mean Intensity. Differences in vocal training and idiosyncratic differences in the way the actors convey emotion are likely the source of this variation.

Neighborhood Density is predictive of Duration and F0 at both word and vowel levels of analysis (Table 1); both longer duration and higher F0 are associated with higher Neighborhood Density. Like the first experiment in Munson and Solomon (2004) we find longer vowel duration for larger Neighborhood Density. We do not, however, find an effect of lexical Frequency on vowel duration. This may be due to differences in the word sets elicited (30% of our words were bisyllabic while Munson and Solomon (2004) only used monosyllabic words). Unlike Munson and Solomon (2004) we also investigated F0 and word duration. We find that lexical predictors, such as Neighborhood Density are also predictive of other acoustic characteristics at the word level. Most importantly, lexical effects, Neighborhood Density, are not removed by the acted emotional modality. The strength of the effect of Neighborhood Density is sometimes modified by the emotional aspect of the speech but is never the less present.

Table 1. Summary of statistically significant ($p<0.05$) lexical effects (ND=Neighborhood Density; Freq=lexical Frequency)

Dependent variable	Lexical predictors		
	ND (direction of the effect)	ND x Freq	ND x Vowel Type
Word Duration	Yes (x Emotion) (duration decreases)	yes	
Vowel Duration	Yes (x Emotion) (duration increases)		
Mean word F0	Yes (x Speaker) (F0 increases)	yes	
Mean vowel F0	Yes (x Speaker) (F0 increases)		yes

REFERENCES

Baayen, R.H. (2010). LanguageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics". R package 1.0 .http://CRAN.R-project.org/package=languageR.

Bates, D., Maechler, M., and B. Bolker (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4

Boersma, P. and D. Weenik. (2010). Praat: doing phonetics by computer (version 5.1.29) [computer program]. Available from hhttp://www.praat.org/

Kryuchkova, T. and B. V. Tucker. (2011). Lexical effects on the production of emotional speech. Presented at the 162nd Meeting of the Acoustical Society of America (San Diego, CA).

Munson, B. and N. Solomon. (2004). The effects of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, vol47, 1048-1058.

Wurm, L. and S. Seaman. (2008). Semantic effects of naming and perceptual identification but not in delayed naming: implications for models and tasks. *Journal of Experimental Psychology*, vol34, 381-398.

Zupan, B., Neumann, D., Babbage, D. R., and B. Willer. (2009). The importance of vocal affect to bimodal processing of emotion: Implications for individuals with traumatic brain injury. *Journal of Communication Disorders*, vol42, 1-17.