Effects of Emotional Prosody and Attention on Semantic Priming

Seung Kyung Kim (skim2@stanford.edu)
Department of Linguistics, Margaret Jacks Hall, Bldg. 460
Stanford, CA 94301-2150 USA

Meghan Sumner (sumner@stanford.edu)
Department of Linguistics, Margaret Jacks Hall, Bldg. 460
Stanford, CA 94301-2150 USA

Abstract

We use an auditory-visual semantic priming paradigm to investigate the effect of phonetically-cued emotional information (*emotional prosody*) on semantic activation of a lexical carrier. In two experiments, we show that words uttered in emotional prosody, although infrequent and atypical, do not necessarily hinder lexical access nor hamper subsequent semantic spreading, and that effects of emotional prosody on word processing crucially depend on the global context in which different types of prosody are presented. These results illustrate the complex nature of spoken word recognition and raise questions about how listeners incorporate multi-faceted information from spoken words.

Keywords: spoken word recognition; emotional prosody; semantic priming; attention

Introduction

Spoken language is full of variation. A single word is never uttered exactly the same way twice. The variation gets compounded as we consider speakers with different genders, ages, and accents. Over the last several decades, we have learned that listeners are highly sensitive to phonetic variation (e.g., Gow, 2002; LoCasto & Connine, 2011) and that listeners' memory for spoken words is also highly detailed (e.g., Bradlow, Nygaard, & Pisoni, 1999; Church & Schacter, 1994). Whenever a piece of speech is uttered, a great deal of social information, as well as linguistic information, is conveyed by phonetic variation, including the speaker's gender, age, geographic origin, emotional state, etc. However, our understanding of the role and function of social information embedded in speech on spoken word recognition is still quite limited. We know that social information cued from speech can influence speech perception and word recognition (e.g., Strand & Johnson, 1996; Sumner & Kataoka, 2013) but much is unknown about how phonetically-cued social information influences spoken language understanding. One important goal of the current study is therefore to provide insights into what mechanisms are involved in this process. To better situate this study, we need first to understand the complex relationship between frequency, typicality, and attention in the process of lexical access.

In episodic models of lexical access (e.g., Goldinger, 1996; Johnson, 2006; Pierrehumbert, 2002), lexical representations are conceptualized as clusters of acoustically detailed exemplars of spoken words. A spoken word is successfully recognized when the sum of activation of exemplars exceeds a certain threshold, and the strength of activation of each exemplar is determined by the (acoustic) distance between the exemplar and the incoming stimuli and the distance between the exemplar and the center of the category. A main prediction of this system is that frequent or typical forms are processed better (in accuracy, latency, etc.) than infrequent or atypical forms because more exemplars will be strongly activated by a frequent or a typical form. Although supported by many studies reporting frequency and typicality effects on spoken word processing, this system is limited in one important way. Since every instance of an exemplar contributes to the system the same way, quantity of experience is successfully accounted for, but the system does not accommodate difference in quality of linguistic experience.

Not accommodating qualitative difference in the model might not be a problem if quantitative difference alone is enough to explain the spoken word recognition process. However, frequency effects hold in only limited contexts and are often not observed, and there is evidence that qualitative difference in linguistic experience has a significant effect on spoken word recognition (Sumner, Kim, King, & McGowan, 2014). For example, Sumner and Kataoka (2013) showed that American English listeners recognized words spoken in Standard American English and Standard British English equally well. The same listeners recognized words spoken in New York dialect worse than words spoken in either standard accent. This result cannot be explained alone by quantity of experience in hearing various accents; the qualitative difference in hearing various accents must be considered as well.

Emphasizing the role of qualitatively different linguistic experience in spoken word recognition, Sumner et al. (2014) argue that each exemplar is encoded with different strength depending on the amount of attention it receives. Given greater attention, a certain incoming stimulus is encoded more deeply, and it can be (representation-wise) as robust as many instances of weakly encoded stimuli. Among other things, social information cued from speech modulates attention. Words spoken in a British English accent attract attention, for example, because the social information of "standardness" and "properness" conveyed by a British accent attracts attention. This *social-weighting* (Sumner et al., 2014) approach to spoken word processing predicts that forms with varying frequency can lead to equally successful recognition provided that enough attention is given to those forms. It also predicts that the same form can result in varying recognition success depending on the attention it receives.

The idea that each exemplar should have different weight-

ing is not new. In Pierrehumbert (2002)'s model, recent exemplars are given greater weight than remote ones to account for a recency effect. Johnson (2006) did not vary weights but stated that differential weighting needs to be considered. Nosofsky (1986) had identified differential attention as an important factor in an episodic model. The contribution of the social-weighting approach therefore lies less in the idea of differential weighting and more in foregrounding the role of attention and qualitative linguistic experience in spoken word recognition, and in linking attention and phonetically-cued social information.

The current study investigates the predictions made by the social-weighting approach by focusing on the effect of phonetically-cued *emotional* information (what we call *emotional prosody*) on spoken word recognition. Emotional prosody is an ideal place to examine the social-weighting approach to spoken word recognition because processing of emotional stimuli and attention allocation are known to be deeply related (e.g., Adelman & Estes, 2013; Brosch, Pourtois, & Sander, 2010). Also, emotional prosody is one of the few types of social information in speech that can be varied within an individual, which makes the effects we find attributable solely to prosody, not to different talker characteristics. Several studies have shown that emotional prosody facilitates (or hinders) processing of emotional words when prosody and word meaning are congruent (or incongruent) (e.g., Nygaard & Lunders, 2002; Nygaard & Queen, 2008; Schirmer, Kotz, & Friederici, 2002). These studies provide initial evidence that emotional prosody influences the recognition of the lexical carrier, but this type of congruency effect between emotional prosody and emotional words cannot tease apart the role of frequency and typicality from the role of social weighting because congruent forms are always more typical and/or frequent than incongruent forms. Therefore, we use non-emotional lexical carriers for emotional prosody in the current study so that we can separate the effect of prosody from the lexical meaning of the carrier.

Using cross-modal semantic priming, we specifically ask whether words spoken in emotional prosody (e.g., *pineapple*$_{Angry}$ or *pineapple*$_{Happy}$) facilitate the recognition of semantically-related words (e.g., *fruit*) as much as words in neutral prosody (e.g., *pineapple*$_{Neutral}$). This paradigm was chosen because a successful semantic priming indicates understanding the meaning of the target word, not just a shallower form-based processing. Also, semantic priming is shown to be sensitive to the degree of attention such that semantic priming is significantly hindered when there is decreased attention like in a divided attention task (e.g., Otsuka & Kawaguchi, 2007; Smith, Bentin, & Spalek, 2001).

We ask two specific questions. First, in Experiment 1, we investigate whether words produced in emotional prosody hinder the understanding of the lexical carrier given their infrequent and atypical status. We test whether non-emotional words uttered in emotional prosody (e.g., *pineapple*$_{A/H}$) facilitate recognition of semantically-related target words (e.g.,

fruit) as much as words in neutral prosody (*pineapple*$_N$). Importantly, different types of prosody are presented to different group of listeners. Frequency-driven lexical access, without attention or a weighting mechanism, predicts that infrequent and atypical productions of words spoken in emotional prosody will hinder the recognition of the lexical carrier, compared to more frequent and typical productions of the same words spoken in neutral prosody. The social-weighting view predicts that words produced in emotional prosody will not necessarily hinder the recognition process.

Second, we investigate whether the presentation format of emotional prosody influences understanding of the lexical carrier. By presentation format, we mean whether listeners experience only one type of prosody (Experiment 1), or they experience different types of prosody in a mixed way (Experiment 2). Change in presentation format implies change in the way attention is allocated and how each stimulus will be attended to (e.g., Barreda, 2012). The frequency-driven view makes no different predictions between single vs. mixed presentation. However, if spoken word recognition is sensitive to social weighting modulated by attention, there will be meaningful differences between Experiment 1 and Experiment 2.

Experiment 1: Within Prosody

Experiment 1 tests whether non-emotional words uttered in emotional prosody (e.g., *pineapple*$_{Angry/Happy}$) facilitate recognition of semantically-related targets (e.g., *fruit*) as much as words in neutral prosody (e.g., *pineapple*$_{Neutral}$).

Experiment 1 has three conditions. 1A investigates the semantic priming effect using prime words spoken in neutral prosody. All the trials, both critical and filler, use neutral prosody primes. This neutral condition provides the baseline for the other two conditions. 1B investigates the semantic priming effect using prime words spoken in angry prosody. All the critical and filler trials use angry prosody primes. Finally, 1C uses words uttered in happy prosody for both critical and filler trials. These three conditions are identical with each other except for the prosody of the spoken words.

Methods

Participants 200 native speakers of English from the Stanford community participated in Experiment 1 for either pay or class credit (about 65 participants for each prosody condition).

Critical Auditory Primes 24 non-emotional words (e.g., *pineapple, transmission*) were recorded by a female speaker of American English in three types of prosody—angry, happy, and neutral.

Which acoustic features make up emotional prosody is an important question, but acoustic analysis of emotional prosody is beyond the scope of the current study. For our purposes, it suffices if our stimuli are heard as the intended emotion by listeners. To check this, we conducted a separate study. For each auditory stimulus, 25 naive listeners rated it on three 9-point scales: how emotional, how angry, and how

happy a given spoken word sounded. The results are shown in Table 1. Both angry and happy stimuli sounded much more emotional than neutral stimuli ($ts > 24$, $ps < 0.001$); Angry stimuli sounded much more angry than the other two ($ts > 17$, $ps < 0.001$); Happy stimuli sounded much happier than the other two ($ts > 21$, $ps < 0.001$). These results verify that the auditory stimuli were perceived as intended.

Table 1: Mean ratings (s.d.) of auditory stimuli

Auditory Primes	Sounds Emotional	Sounds Angry	Sounds Happy
Angry Prosody	7.0 (0.7)	5.0 (1.1)	3.6 (1.1)
Happy Prosody	6.7 (0.7)	1.5 (0.4)	7.1 (0.6)
Neutral Prosody	3.6 (0.4)	2.4 (0.5)	3.9 (0.5)

Note: The scale ranges from 1 (not at all) to 9 (definitely).

Critical Visual Targets The top semantic associate for each prime word (e.g., *fruit* for *pineapple*; *car* for *transmission*), obtained from a separate study, was chosen as a semantically-related target word for each prime word.

Design and Procedure The design is shown in Table 2. The critical auditory primes and the critical visual target words were pseudo-randomly paired and crossed in two lists for each prosody condition. Each list had 24 critical trials with half related and half unrelated pairs. In addition, each list had 72 filler trials. All filler trials had real word auditory primes, and they were paired with 24 real-word visual targets and 48 non-word visual targets.

Table 2: Design of Experiment 1

Exp	Prosody	Condition	Auditory Prime	Visual Target
1A	Neutral	Related	pineapple$_N$	fruit
		Unrelated	specialist$_N$	
1B	Angry	Related	pineapple$_A$	fruit
		Unrelated	specialist$_A$	
1C	Happy	Related	pineapple$_H$	fruit
		Unrelated	specialist$_H$	

The experiment consisted of 96 trials and took approximately 10 minutes to finish. Each trial started with a spoken prime word played through headphones. 100 ms after the offset of the prime, a written target word was presented on the screen until the participant made a lexical decision. Participants were instructed to respond as quickly and as accurately as possible. The accuracy and the latency of each lexical decision were recorded.

Results

Statistical Analysis Procedures We used mixed-effects models for analyzing both accuracy (generalized linear models) and latency (linear models). All analyses were carried out using *R*'s lme4 and lmerTest packages. We report the results from the models with a (nearly) maximal random effect structure justified by the experiment design (e.g., Barr, Levy, Scheepers, & Tily, 2013). In some cases (especially

for accuracy analyses), the random structure had to be simplified to avoid convergence errors. The reported p-values are calculated by the lmerTest package. In all models, continuous variables were centered. Categorical variables were sum-coded unless a treatment-coding was necessary for making relevant inferences. For latency analyses, target word frequency, prime word frequency, and trial count were included as control factors.

Accuracy The mean accuracy rate was 96.7%, 96.0%, and 96.4% for 1A, 1B, and 1C, respectively. There was no significant difference in accuracy among the three conditions. Furthermore, within critical trials, there was no significant difference by semantic relatedness on accuracy. Therefore, accuracy rate is not considered in subsequent analyses.

Latency Reaction times (RTs) of correct responses to critical trials were included in the analyses. RTs that fell outside of 3 s.d. of the grand mean were discarded (2.5%). All analyses were carried out with log RTs. The mean RT and the priming sizes for the three conditions are reported in Table 3. For ease of interpretation, the log RTs have been transformed back to milliseconds.

Table 3: Mean RTs in ms [logRT (s.d.)] and priming size in Experiment 1

Exp	Prosody	Related	Unrelated	Priming Size
1A	Neutral	511 [6.24 (0.23)]	526 [6.27 (0.22)]	15
1B	Angry	511 [6.24 (0.22)]	526 [6.27 (0.23)]	15
1C	Happy	543 [6.30 (0.22)]	553 [6.32 (0.22)]	10

We found a main effect of semantic relatedness ($\beta = -0.012$, s.e $= 0.006$, t $= -2.1$, p $= 0.04$), suggesting an overall semantic priming effect. We also found a main effect of the prime prosody condition such that RTs in 1C (happy prosody) are significantly slower than the average ($\beta = 0.05$, s.e., $= 0.02$, t $= 2.3$, p $= 0.02$). There was no interaction between semantic relatedness and the prosody condition.

Testing further for simple effects of semantic relatedness for each prosody condition, we found that visual targets following semantically related primes were recognized more quickly than following semantically unrelated primes when primes were produced in neutral prosody or angry prosody, but not when produced in happy prosody (1A—Neutral: 15 ms, $\beta = -0.012$, s.e. $= 0.006$, t $= -2.1$, p $= 0.04$; 1B—Angry: 15 ms, $\beta = -0.015$, s.e. $= 0.006$, t $= -2.4$, p $= 0.03$; 1C—Happy: 10 ms, $\beta = -0.01$, s.e. $= 0.006$, t $= -1.6$, p $= 0.1$).

In sum, words spoken in angry prosody result in semantic facilitation on par with the typical forms in neutral prosody, which is contrary to what we might expect if an atypical production hinders the spoken word recognition process. Words in happy prosody, on the other hand, did slow down lexical access and seemed unable to produce a robust semantic priming effect. The results thus suggest that different types of atypical productions have different consequences in the recognition process.

Experiment 2: Mixed Prosody

Experiment 2 tests whether the effect of emotional prosody on semantic priming depends on the presentation format in which emotional prosody appears. Experiment 2 has two conditions. 2A includes words spoken in angry prosody and in neutral prosody in the critical trials; 2B includes words spoken in happy prosody and in neutral prosody in the critical trials. Both 2A and 2B include words produced in angry, happy, and neutral prosody in the filler trials.

Participants 162 native speakers of English from the Stanford community participated in Experiment 2 for either pay or class credit (about 80 participants for each condition).

Materials The materials are the same as Experiment 1.

Design and Procedure The design is shown in Table 4. The critical primes and targets were pseudo-randomly paired and crossed in four experimental lists for each experiment. Each list had 24 critical trials with half related and half unrelated pairs. In 2A, half of the critical primes were spoken in neutral prosody and the other half in angry prosody; in 2B, half were in neutral and half in happy prosody. In addition, each list had 72 filler trials. All filler primes were real words and were paired with 24 real-word targets and 48 with non-word targets. For both 2A and 2B, neutral, angry, and happy filler primes each took up a third of the filler trials. The experiment procedure was the same as in Experiment 1.

Table 4: Design of Experiment 2

Exp	Critical Prosody	Condition	Auditory Prime	Visual Target
2A	Angry & Neutral	Related	pineapple$_{A/N}$	fruit
		Unrelated	specialist$_{A/N}$	
2B	Happy & Neutral	Related	pineapple$_{H/N}$	fruit
		Unrelated	specialist$_{H/N}$	

Note: Both 2A and 2B included filler trials with words in angry, happy, and neutral prosody.

Results

Accuracy The mean accuracy rate was 96.7% and 95.8% for 2A and 2B, respectively. There was no significant difference between the two conditions on accuracy rate. Further, within critical trials, there was no significant difference in accuracy by semantic relatedness nor by prosody type. Therefore, accuracy rate is not considered in subsequent analyses.

Latency RTs of correct responses to critical trials were included. RTs that fell outside of 3 s.d. of the mean were discarded (2.6%). The mean RTs for the two experiments are reported in Table 5.

We found a main effect of semantic relatedness ($\beta = -0.01$, s.e. $= 0.003$, t $= -3.1$, p $= 0.003$) suggesting an overall semantic priming. We found no main effect of experiment condition (2A vs. 2B) nor a main effect of prime prosody (Emotional vs. Neutral). There were no significant two-way interactions, but the three-way interaction among semantic relatedness, experiment condition and the prime prosody type was

Table 5: Mean RTs in ms [logRT (s.d.)] and priming size for Experiment 2

Exp	Prosody	Related	Unrelated	Priming Size
2A	Angry	516 [6.25 (0.22)]	529 [6.27 (0.23)]	13
	Neutral	516 [6.25 (0.23)]	521 [6.26 (0.21)]	5
2B	Happy	526 [6.27 (0.22)]	529 [6.27 (0.21)]	3
	Neutral	516 [6.25 (0.21)]	535 [6.28 (0.21)]	19

significant ($\beta = -0.006$, s.e. $= 0.003$, t $= -1.99$, p $= 0.05$). To unpack the interaction, we further fit separate models to 2A and 2B.

2A shows a main effect of semantic relatedness ($\beta = -0.008$, t $= -1.93$, p $= 0.05$) and no main effect of prime prosody. The interaction between semantic relatedness and prosody was not significant. Testing for simple effects reveals that visual targets following semantically related primes were recognized more quickly than following semantically unrelated primes when primes were produced in angry prosody but not when produced in neutral prosody (Angry: 13, $\beta = -0.013$; s.e. $= 0.006$, t $= -2.3$, p $= 0.02$; Happy: 5 ms, $\beta = -0.004$, s.e. $= 0.006$, t $= -0.6$, p $= 0.5$).

In 2B, the result shows a main effect of semantic relatedness ($\beta = -0.01$, s.e. $= 0.005$, t $= -2.5$, p $= 0.02$) and no main effect of prime prosody. The interaction between semantic relatedness and prosody was marginally significant ($\beta = 0.007$, s.e. $= 0.004$, t $= 1.7$, p $= 0.09$). Testing for simple effects show that visual targets following semantically related primes were recognized more quickly than following semantically unrelated primes when primes were produced in neutral prosody but not when produced in happy prosody (Neutral: 19 ms, $\beta = -0.018$, s.e. $= 0.006$, t $= -3.2$, p $= 0.003$; Happy: 3 ms, $\beta = -0.004$, s.e. $= 0.007$, t $= -0.6$, p $= 0.5$).

In sum, the results show that when different types of prosody are presented in a mixed way, semantic priming patterns greatly vary. These varying priming patterns, in particular the different patterns produced by neutral prosody, strongly indicate that lexical access and semantic spreading of spoken words are critically influenced by surrounding prosodic context.

General Discussion

In this paper, we examined effects of emotional prosody on semantic activation of a carrier word. Experiment 1 showed that prime words spoken in neutral prosody (e.g., *pineapple*$_{Neutral}$) and in angry prosody (e.g., *pineapple*$_{Angry}$) both facilitate the recognition of semantically-related targets (e.g., *fruit*), whereas primes spoken in happy prosody did not show significant semantic priming. Importantly, this experiment showed robust facilitation of recognition to target words by both a frequent and typical word form and an infrequent and atypical word form. In Experiment 2, we found that variability of prosodic contexts within experiments greatly influences semantic priming patterns. In 2A, angry stimuli engendered semantic priming but neutral stimuli did not; in 2B, happy

stimuli did not but neutral stimuli did. From these results, we have three findings to understand: semantic priming equivalence; within vs. mixed-prosody differences; and the relationship between attention, emotion, and semantic priming.

Semantic Priming Equivalence

In Experiment 1, we found that infrequent and atypical forms (e.g., *pineapple*_{Angry}) are as successful as frequent and typical forms of the word (e.g., *pineapple*_{Neutral}) in accessing the lexical representation and inducing semantic spreading. We explain this equivalence via social weighting.

This equivalence is unexpected from a frequency-based episodic lexical access system because it predicts atypical and infrequent forms will slow down lexical access, precluding priming effects. If this were the case, we would have seen stronger priming for words in neutral prosody than words in angry prosody. Once we consider attention that is mediated by social context, the equivalence effect can be readily explained. Priming in 1A occurs because it is the typical case with words in neutral prosody. Priming in 1B occurs because emotional stimuli, especially negative ones, grab more attention than non-emotional stimuli (e.g., Adelman & Estes, 2013; Brosch et al., 2010). This increased attention leads to stronger encoding, resulting in robust activation. In a sense, the increased attention due to negative emotionality offsets the fact that these forms are infrequent.

We are not dismissing the effect of frequency in lexical access. Infrequent forms are often actually processed slower than frequent forms (e.g., Connine, 2004; Strand, 2000). In fact, in our own data, the overall RT in 1C (happy prosody) was significantly slower than the other two conditions and it did not engender a significant semantic priming effect. Unlike angry prosody which could overcome its atypical infrequent status, happy prosody seems to fail to do so, and this is perhaps because positive emotional stimuli attract less attention than negative emotional stimuli (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Taylor, 1991)

Effect of Presentation Format

The only difference between Experiments 1 and 2 was the presentation format of the stimuli, that is, whether each participant was presented with one type or multiple types of prosody mixed during the experiment, but we found distinct priming patterns between the two. Crucially, when presented alone, neutral prosody primes facilitated recognition of semantically related target words, but when presented intermixed with other prosodies, especially with angry prosody, neutral prosody primes failed to facilitate recognition of semantically related target words. We believe that different acoustic contexts lead to different attention allocation, which in turn leads to different priming patterns.

Behavioral difference between blocked vs. mixed presentation in linguistic tasks is not uncommon. Mixed presentation is reported to make the task harder than blocked presentation (e.g., Sommers, Nygaard, & Pisoni, 1994). In the current study, however, the difference between single and mixed presentation was not simply about difficulty. In fact, we see no meaningful accuracy or latency difference between Experiments 1 and 2. Difference was rather in the priming patterns, likely due to difference in how attention is allocated to different stimuli. Barreda (2012) has suggested that the behavioral difference between mixed vs. blocked presentation is due to the attention needed to detect differences among stimuli. Along this line, one possible interpretation of the current results is that a listening environment with multiple emotional prosodies uses attentional resources to detect differences in stimuli, and this in turn may make the prosodic difference more meaningful and relevant in the linguistic task at hand than in a listening environment with a single prosody type. This means we would find differential effects of emotional prosody more easily in mixed-presentation than in blocked presentation, just as we did between Experiments 1 and 2.

More broadly, we might predict that a listening environment with multiple social categories, be they multiple prosodies, multiple talkers, etc., makes the differences among them more meaningful and makes listeners evaluate and process the current item by comparing it to previous items. We would therefore expect that effects of social information on listener behavior will be greater and more readily observable when listeners experience multiple types of social categories rather than a single type.

Attention, Emotion, and Semantic Priming

In Experiment 2, when the prosodies are presented in a mixed way, we found, roughly speaking, that words spoken in angry prosody are the strongest in their ability to induce semantic priming, then words in neutral prosody, and words in happy prosody the weakest. We suspect this pattern is related to the relationship between attention, emotion, and semantic priming for two reasons.

First, emotional stimuli tend to grab more attention than non-emotional stimuli (e.g., Adelman & Estes, 2013; Brosch et al., 2010), but negative stimuli tend to grab still more attention than positive stimuli (e.g., Baumeister et al., 2001; Taylor, 1991). Second, we know that semantic priming is sensitive to attention (e.g., Otsuka & Kawaguchi, 2007; Smith et al., 2001). Semantic priming increases as attention to the prime increases; it decreases as attention to the prime decreases. Therefore, it may be that, in 2A (comparing neutral and angry), words spoken in neutral prosody did not receive enough attention to induce semantic priming because words in angry prosody grabbed more attention. In 2B (comparing neutral and happy), it appears that words spoken in neutral prosody received more attention than words spoken in happy prosody. Perhaps, in the case of happy prosody, even though it receives extra attention due to its emotionality, this is still not enough to overcome its atypical status, whereas angry prosody receives the most attention, being a negative emotion.

One open question is, given that both 2A and 2B included filler trials where the spoken words were produced in angry,

happy, and neutral prosodies, how to explain the fact that filler words in angry prosody did not deprive attention from neutral critical trials in 2B. This might be related to the different ratio of angry stimuli present in 2A and 2B. This is one of many questions that future investigation needs to look at to better understand dynamics of attention allocation depending on different combinations of the stimulus set.

In sum, we showed that atypical productions of words like *pineapple* in emotional prosody do not always cause difficulty in spoken word recognition and that the effect of emotional prosody is context-dependent. These results suggest that the spoken word recognition process is crucially mediated by the qualitative experience of the spoken word and the amount of attention it attracts. We believe that the current study illustrates the complex nature of spoken word recognition and raises questions about how listeners incorporate multi-faceted information from speech signals.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Barreda, S. (2012). Vowel normalization and the perception of speaker changes: an exploration of the contextual tuning hypothesis. *The Journal of the Acoustical Society of America*, *132*, 3453–3464.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219.

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion*, *24*(3), 377–400.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521–533.

Connine, C. M. (2004). It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, *11*(6), 1084–1089.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 163–179.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

LoCasto, P. C., & Connine, C. M. (2011). Processing of no-release variants in connected speech. *Language and Speech*, *54*(2), 181–197.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Pscyhology: General*, *115*(1), 39–57.

Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, *30*(4), 583–593.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017–1030.

Otsuka, S., & Kawaguchi, J. (2007). Divided attention modulates semantic activation: evidence from a nonletter-level prime task. *Memory & Cognition*, *35*(8), 2001–2011.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology, vol. vii* (pp. 101–139). Berlin: Mouton de Gruyter.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233.

Smith, M. C., Bentin, S., & Spalek, T. M. (2001). Attention constraints of semantic activation during visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1289–1298.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition: Effects of variability in speaking rate and verall amplitude. *The Journal of the Acoustical Soceity of America*, *96*(3), 1314–1324.

Strand, E. A. (2000). *Gender Stereotype Effects in Speech Processing*. Ph.d. dissertation, Ohio State University.

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In *Natural language processing and speech technology: Results of the 3rd konvens conference* (pp. 14–26).

Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America*, *134*(6), EL485.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, *4*, 1–13.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological bulletin*, *110*(l), 67–85.