A PROSODIC CONFIGURATION THAT CONVEYS POSITIVE ASSESSMENT IN AMERICAN ENGLISH

Nigel G. Ward and James A. Jodoin

University of Texas at El Paso
nigelward@acm.org, jajodoin@miners.utep.edu

ABSTRACT

Among the meaning-bearing forms of language are, recent work suggests, "prosodic constructions," that is, configurations of diverse prosodic features in specific temporal relations. However the detailed properties of such constructions have been little studied. As a case study, we here examine the "High-Lengthened-Quiet" construction of American English, used for positive assessments. From subjects' judgments of modified stimuli, we find: that this construction indeed conveys positive assessment, that the positive interpretation is stronger to the extent that the prosodic components are more strongly present, and that this is truly a configuration, in that the contributions of the component features depend on their temporal positions.

Keywords: prosodic constructions; perception; temporal patterns; multistream prosody

1. RELATING PROSODY AND MEANING

Modeling the ways in which prosody conveys meaning is a challenge in many ways [18, 20, 17, 15], not least because of the diversity of types of prosodic forms which may bear meaning, ranging from the lexically-bound to the purely paralinguistic. Between these two extremes, recent research has also identified configurations of diverse prosodic features in specific temporal relations that bear specific meanings. Following [16], we will refer to these as "prosodic constructions." While many prosodic constructions have been discussed, only a handful have been quantitatively described [2, 11, 12, 19]. Most such previous work has relied on corpus studies and production studies, but here we seek to elucidate the nature of a prosodic construction using perception studies instead.

A related question in prosody modeling relates to the extent to which different types of prosodic information work in concert or independently. Traditionally work in prosody focused on pitch (intonation) alone, but there is increasing interest in multistream feature configurations [15, 13]. For example, for pitch and intensity, much work has investigated how different alignments of peaks for these two streams can convey different meaning [14, 21]. Here we investigate another multistream configuration.

Another central question in prosody is that of when and whether the perception of meaning-bearing prosody is categorical or gradient in nature. For a construction conveying contrast, Kurumada showed that the strength of perception of contrast was roughly proportional to the strength of the components of the construction, in a warping experiment that manipulated pitch and duration [9]. Here we explore this issue for another construction.

2. PROSODIC EXPRESSIONS OF POSITIVE FEELING IN AMERICAN ENGLISH

To date most work on conveying positive feelings with prosody has examined emotions, most often in monologue or acted speech. Across languages, positive emotions associate with features including higher pitch with "upward inflections," higher intensity, and breathy voice [10]. Classifiers that use multiple prosodic features in concert achieve better performance than those that use just one type of feature [8], but it is not known whether this is due specifically to the importance of feature configurations.

There has also been some work on the prosody of positive feeling in interaction and dialog. For American English, our language of interest here, we know of four studies. Fernald observed that expressions of approval in infant-directed speech often have "exaggerated rise-fall F_0 contours" [3]. Freese and Maynard's study of the prosody of good news as it occurs in conversation, for example in a sentence announcing that a relative had found a job, found that it often includes 1) high pitch level, 2) increased pitch range, 3) abrupt step-ups and rises, 4) modal voice, 5) loudness on key words, 6) and fast and increasing rate [6]. Freeman and colleagues did two quantitative studies. Examining the expression of positive stance in problem-solving meetings, they found that tokens of *yeah* conveying agreement with something the interlocutor had suggested tended to have a) longer vowel duration, b) pitch ranges that extend

High then Lengthened then Quiet Construction	
Function: Express Positive Assessment	
Form:	
typical timespan	prosodic properties
–1600 to –800 ms	raised pitch
–800 to –200 ms	lengthened vowels increased loudness
–200 to 0 ms	sharp drop in loudness (clipped end)
0 to 800 ms	silence or low intensity

Figure 1: The High-Lengthened-Quiet Construction.

higher, c) lower mean intensity, and d) a earlier and steeper intensity drop [5], and, more generally that the stressed vowels in content words were longer in positive utterances [4]. Together these studies show that prosodic features of many types are involved in conveying positive feeling, but not specifically how these features are related or configured.

More recently, a statistical study of the prosody of a corpus of American English conversations [22] led to the discovery of a frequently co-occurring set of prosodic features in a specific temporal configuration which is often used to convey positive assessment, as summarized in Figure 1. Utterances with this prosody in the corpus include *I love teaching, I love helping kids*; *[in terms of being prepared for Monday's exam], I feel good*; *[I like magical fantasy movies but] I also really love the Boondock Saints*; *it's really cool coming up with a program, and then being able to see that program on someone else's phone*; and *I like her style*. The audio is available at [23]. This pattern is of course not rigid: there is significant variation in which features are present, how strongly they are present, their timespans, how they align with syllables, and what meanings beyond positive feeling are also present.

Overall, the findings of these studies are largely compatible, despite the diversity of methods and of genres studied. The prosodic construction model however differs in ascribing to the prosodic features specific locations in a temporal configuration. We here seek to determine whether this actually matters.

3. HYPOTHESES

Based on these considerations we formulated three hypotheses.

Hypothesis 1: The high-long-quiet pattern indeed conveys positive assessment. (Since the connection found in the corpus study might have been due to a hidden variable, we test this by direct comparisons, holding everything constant but the prosody.)

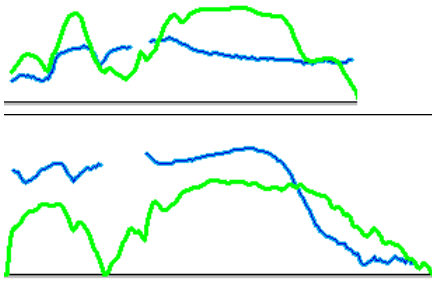

Figure 2: *good job* with neutral prosody (above) and with positive prosody (below). Pitch is in log Hz, blue, and intensity in decibels, green.

Hypothesis 2: This pattern conveys positive assessment in a gradient manner. (We expected stimuli where the component features are more strongly present to be perceived as more positive.)

Hypothesis 3: This pattern functions as a configuration, not a temporally amorphous conjunction of independent features. (We expected the component features to contribute more strongly to perceptions of positivity when they appear properly located relative to the other features.)

4. EXPERIMENT DESIGN

We tested these hypotheses with a perception study, in which subjects were presented prosodically-manipulated stimuli in pairs and indicated "which sounds more positive." Although previous research implicates a multitude of features, here we varied only those that seemed most consistently and saliently involved, namely pitch height and syllable duration.

4.1. Stimulus Sources

We used phrases observed in uses of this construction in everyday life, chosen also for their ability to sound natural both with strongly positive feeling and with neutral feeling. Samples of each phrase were recorded in these two conditions by native speakers of English asked to imagine contexts where they would have these feelings. The *thank you* samples were produced by the second author, and the *good job* samples by a teenage girl naive to our hypotheses. Pitch and intensity contours for the latter are seen in Figure 2, and some measurements in Table 1. We note that both positive productions had the expected relatively high pitch (region 1) and then the relatively longer syllable. In other respects the prosody was fairly diverse, apart from a weak ten-

Table 1: Selected properties of the source samples, for the first syllable (s1) and the second (s2). Pitch is in 100 Hz, intensity in decibels, duration in seconds. The last column is the average ratio of the positive to neutral measurements, shown only for the features that we chose to manipulate.

	good job		*thank you*		avg.
	neu	pos	neu	pos	ratio.
s1 peak pitch	254	323	106	219	
s1 avg. pitch	239	313	106	195	1.57
s1 max. intensity	75	69	59	65	
s1 duration	.21	.16	.13	.19	
s2 peak pitch	263	345	102	117	
s2 avg. pitch	245	287	99	101	
s2 max. intensity	77	77	52	60	
s2 duration	.39	.56	.14	.25	1.61

dency for the pitch of the second syllable also to be higher for the positive samples.

4.2. Stimuli Preparation and Statistical Tests

For Hypothesis 1: For each phrase we created two stimulus pairs: the neutral samples in original form and modified to exhibit the prosody of the positive form; and the positive samples in original form and modified to exhibit the prosody of the neutral form. Specifically, we used Praat to first adjust the syllable durations and then transplant the pitch contours [24]. We predicted that for both pairs the stimulus with the positive-sourced prosody would be judged to sound more positive, tested against the null hypothesis of no preference using the binomial distribution.

For Hypothesis 2: For each phrase we prepared stimuli with varying degrees of match to the Construction. The source in each case was the neutral production, the "0%" stimulus. To create, for example, the 20% stimulus, the pitch across the utterance was increased 20% of the way to that in the positive sample, in log Hertz, and the duration of region 2 was similarly increased by 20% of the way to the positive sample value. We predicted that for every pair of stimuli, the one with more intense manipulations would be judged more positive. To also investigate sensitivity, we created a set of stimuli pairs that spanned the 0-100% range in five steps and a set that spanned it in eight steps. Again we used the 0.5 binomial distribution to judge significance.

For Hypothesis 3: We prepared stimuli using the neutral productions as sources, modified either per-region as described above, or by globally increasing the pitch and duration across the sample. In each pair the average pitch and total duration were the same. We did this for manipulations to 50% and to 100% of the prosody of the positive sample. We predicted that the stimuli with the temporally-appropriate prosodic modifications would be selected as positive more often, by the same test.

All of these manipulations were done using Praat [1]. The results we judged to be adequately natural: despite some slight impairments, especially in the two stimuli whose creation involved pitch reductions, all sounded like human productions. The Praat script, source samples, and stimuli are available [23].

For all tests we used a 5% confidence level.

4.3. Procedure and Subjects

To ease the subjects' task, we chose to elicit forced-choice judgments of stimulus pairs. The two samples in each pair were separated by 1 second of silence. We asked subjects to "ignore any distortions, and just focus on the feeling of what the speaker is saying." The presentation alternated *good job* stimulus pairs and *thank you* pairs, for three reasons: to reduce fatigue, to weaken possible anchoring effects between pairs, and to make it harder for subjects to listen analytically or infer the manipulations, promoting instead the quick judgments we wanted. For the same reasons we tested all 3 hypotheses at once, interleaving stimulus pairs related to the different hypotheses in a randomized order.

For each phrase, there were 2 pairs for Hypothesis 1, 8 + 5 pairs for Hypothesis 2, and 2 pairs for Hypothesis 3. Each pair was presented twice, once expected-positive first and once expected-neutral first. With the two phrases, this gave a total of 17 * 2 * 2 = 68 stimulus pairs.

Subjects were a convenience sample, recruited from a graduate Computer Science class at a university in the Southwestern United States, on the Mexican border. In this class about half the students were non-native speakers of English, but thinking that anyone with at least modest experience with conversational English would be familiar with this pattern, our only screening question was for "years lived in an English-speaking country," with choices "one or more" and "less than one." We offered students $10 to stay after class for 15 minutes to participate; 21 accepted. The stimuli were played over the wall-mounted speakers and the subjects marked their judgments on paper. Data from 3 subjects was excluded from analysis: one based on the screening question and two for not providing judgments for all stimuli. The 18 remaining subjects included 6 females.

Table 2: Number of more-positive judgments. (Hypothesis 1)

	prosody	
source	neutral	positive
neutral	1	71
positive	7	65

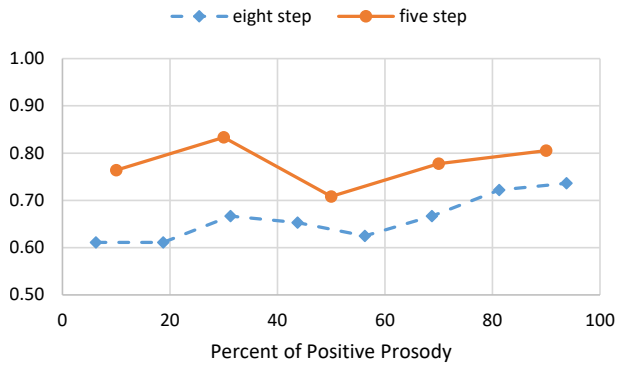

Figure 3: Fraction of judgments favoring the stimulus with higher first-syllable pitch and longer second-syllable duration, as a function of difference between stimuli and positions in the continuum. The dot at 10% represents judgments of the 0%-20% stimulus pairs, and so on. (Hypothesis 2)

5. RESULTS

Table 2 shows that, for both sources the stimulus with the positive prosody was judged to sound more positive. For both sources the differences were significant, supporting Hypothesis 1.

Figure 3 shows results for the stepwise comparisons. All differences are significant, supporting Hypothesis 2. The results also indicate that discriminating between adjacent members of the eight step series was harder than for the five step series.

Table 3 shows that stimuli were perceived as more positive when the associated features occurred where specified by the construction. The differences were significant for both the 50% and 100% manipulations, supporting Hypothesis 3.

Thus every prediction was borne out, all at the 5% confidence level. Moreover, when, concerned that the use of stimuli derived from author-produced samples may have biased the results, we repeated the analysis using only judgments of stimuli derived from the naive speaker's productions, all comparisons were still significant.

Finally, we observed substantial variation across subjects. The average subject's judgments matched our expectations 74% of the time, but this varied from 43% to 95%.

Table 3: Number of more-positive judgments as a function of the scope of the prosodic manipulation. (Hypothesis 3)

	overall	syllable-localized
50%	17	55
100%	11	61

6. SUMMARY AND DISCUSSION

Thus we find that this High-Lengthened-Quiet configuration of prosodic features indeed conveys meaning, and that it does so more strongly as the strength of the component features increases. Further, for the first time, we show that the temporal location of the component features of a prosodic construction is perceptually important. This confirms the utility of prosodic constructions in describing meaning-bearing prosody.

Future work might examine many things. As our stimuli here were two-syllable isolated phrases, one might examine uses of this prosodic construction on longer phrases and in ongoing dialog. One might also quantify the relative contributions of the various prosodic features to the perception of positive feeling, including the contributions of different aspects of the intensity contour. One might also look for alternative, non-constructional ways to model these observations, and investigate aspects which may be difficult to account for in a constructional model [7]. One might investigate whether the meaning of this construction, is best characterized as positive assessment or as something more specific or more general. One might also test whether the meaning contribution of the prosodic components of this construction as a whole is greater than the sum of the contributions of the parts.

In any case, these results suggest that studies of the prosodic expression of semantic and pragmatic functions should use methods that allow the discovery of temporal configurations. Similarly applied work — in language description, speech synthesis, computational paralinguistics and so on — would be well-advised to use models that enable the representation of such configurations.

Acknowledgements: We thank Natalia Mazzaro for comments. This work was supported in part by DARPA's LORELEI program, but no official endorsement should be inferred.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2005. Praat: doing phonetics by computer [computer program]. Retrieved from http://www.praat.org/.

[2] Day-O'Connell, J. 2013. Speech, song, and the minor third: An acoustic study of the stylized interjection. *Music Perception* 30, 441–462.

[3] Fernald, A. 1989. Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development* 6, 1497–1510.

[4] Freeman, V. 2015. *The phonetics of stance-taking*. PhD thesis University of Washington.

[5] Freeman, V., Levow, G.-A., Wright, R., Ostendorf, M. 2015. Investigating the role of 'yeah' in stance-dense conversation. *Interspeech* 3076–3080.

[6] Freese, J., Maynard, D. W. 1998. Prosodic features of bad news and good news in conversation. *Language in Society* 27, 195–219.

[7] Huttenlauch, C., Feldhausen, I., Braun, B. 2018. The purpose shapes the vocative: Prosodic realisation of Columbian Spanish vocatives. *Journal of the International Phonetic Association* 48, 33–56.

[8] Koolagudi, S. G., Rao, K. S. 2012. Emotion recognition from speech: a review. *International Journal of Speech Technology* 15, 99–117.

[9] Kurumada, C., Brown, M., Tannenhaus, M. K. 2012. Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. *Cognitive Science Conference*.

[10] Murray, I. R., Arnott, J. L. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustic Society of America* 93, 1097–1108.

[11] Niebuhr, O. 2014. Resistance is futile: The intonation between continuation rise and calling contour in German. *Interspeech* 132–136.

[12] Niebuhr, O. 2015. Stepped intonation contours: A new field of complexity. In: Skarnitzl, R., Niebuhr, O., (eds), *Tackling the Complexity of Speech*. Charles University Press 39–74.

[13] Niebuhr, O., D'Imperio, M., Fivela, B. G., Cangemi, F. 2011. Are there shapers and aligners? Individual differences in signalling pitch accent category. *International Congress of Phonetic Sciences* 120–123.

[14] Niebuhr, O., Neitsch, J. 2019. Questions as prosodic configurations: How prosody and context shape the mutltiparametric acoustic nature of rhetorical questions in German. *International Congress of the Phonetic Sciences*.

[15] Niebuhr, O., Ward, N. G. 2018. Challenges in the modeling of pragmatics-related prosody: Introduction to the special issue. *Journal of the International Phonetics Association* 48, 1–8.

[16] Ogden, R. 2010. Prosodic constructions in making complaints. In: Barth-Weingarten, D., Reber, E., Selting, M., (eds), *Prosody in Interaction*. Benjamins 81–103.

[17] Prieto, P. 2015. Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science* 6, 371–381.

[18] Scherer, K. R., Banziger, T. 2004. Emotional expression in prosody: a review and an agenda for future research. *Speech Prosody*.

[19] Torreira, F., Grice, M. 2018. Melodic constructions in Spanish: Metrical structure determines the association properties of intonational tones. *Journal of the International Phonetics Association* 48, 9–32.

[20] Vaissiere, J. 2008. Perception of intonation. In: Pisoni, D., Remez, R., (eds), *The Handbook of Speech Perception*. John Wiley & Sons 236–263.

[21] Ward, N. G. 2018. A corpus-based exploration of the functions of disaligned pitch peaks in American English dialog. *Speech Prosody* 349–353.

[22] Ward, N. G. 2019. *Prosodic Pattterns in English Conversation*. Cambridge University Press.

[23] Ward, N. G. 2019. The prosody of positive assessment. http://www.cs.utep.edu/nigel/positivity.

[24] Yoon, K. 2007. Imposing native speakers' prosody on non-native speakers' utterances. *Modern British and American Language and Literature* 25, 197–215.