A PROSODIC CONFIGURATION THAT CONVEYS POSITIVE ASSESSMENT IN AMERICAN ENGLISH

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-Long-Quiet” construction of American English, used for positive assessments. From subjects’ judgments of systematically 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 [20, 22, 19, 17], not least because of the diversity of types of prosodic forms which may bear meaning. Some of these are comparatively simple to model, including paralinguistically meaningful prosodic features which may directly reflect physiological states, and prosodic feature patterns that add specific meaning to specific linguistic units [3], such as that used with awww to show admiration for a baby picture. Recent research has also identified configurations of diverse prosodic features in specific temporal relations that can bear specific meanings. Following [18], we will refer to these as “prosodic constructions.” While many prosodic constructions have been discussed impressionistically, only a handful have been quantitatively described [4, 13, 14, 21]. All 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. Tradition-
High then Long then Quiet Construction

<table>
<thead>
<tr>
<th>Function: Express Positive Assessment</th>
</tr>
</thead>
</table>

**Form:**
- typical timespan: prosodic properties
- $-1600$ to $-800$ ms: raised pitch
- $-800$ to $-200$ ms: increased loudness
- $-200$ to $0$ ms: lengthened vowels
- $0$ to $800$ ms: silence or low intensity

**Figure 1:** The High-Long-Quiet Construction.

tokens of *yeah* conveying agreement with something the interlocutor had suggested were more likely to have a) longer vowel duration, b) pitch ranges that extend higher, c) lower mean intensity, and d) a earlier and steeper intensity drop [7], and, more generally that the stressed vowels in content words were longer in positive utterances [6]. Thus we know that prosodic features of many types are involved in conveying positive feeling, but the question of how these features are related or configured remains open.

More recently, a statistical study of the prosody of a corpus of American English conversations [1] 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. This appears in the corpus, for example, in *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*. 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 in fact 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 coincidental or the result of a hidden variable, we test this by direct comparisons, holding everything constant but the prosody.)

Hypothesis 2: This pattern conveys positive assessment in a gradient manner. (We expect to find that, across a wide range, stimuli where the component features are more strongly present are perceived as more positive.)

Hypothesis 3: This pattern functions as a configuration, not a temporally amorphous conjunction of independent features. (We expect to find that the component features 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, and chosen 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 ex-
Table 1: Selected properties of the source samples, for the first syllable (s1) and the second (s2). Pitch 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.

<table>
<thead>
<tr>
<th></th>
<th>good job</th>
<th>thank you</th>
<th>avg. ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>neu pos</td>
<td>neu pos</td>
<td></td>
</tr>
<tr>
<td>s1 peak</td>
<td>254 323</td>
<td>106 219</td>
<td></td>
</tr>
<tr>
<td>s1 avg.</td>
<td>239 313</td>
<td>106 195</td>
<td>1.57</td>
</tr>
<tr>
<td>s1 max.</td>
<td>75 69</td>
<td>59 65</td>
<td></td>
</tr>
<tr>
<td>s1 duration</td>
<td>.21 .16</td>
<td>.13 .19</td>
<td></td>
</tr>
<tr>
<td>s2 peak</td>
<td>263 345</td>
<td>102 117</td>
<td></td>
</tr>
<tr>
<td>s2 avg.</td>
<td>245 287</td>
<td>99 101</td>
<td></td>
</tr>
<tr>
<td>s2 max.</td>
<td>77 77</td>
<td>52 60</td>
<td></td>
</tr>
<tr>
<td>s2 duration</td>
<td>.39 .56</td>
<td>.14 .25</td>
<td>1.61</td>
</tr>
</tbody>
</table>

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 preferences 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, 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 [2]. The results we judged to be adequately natural: despite some slight impairments, especially in the stimuli whose creation involved pitch reductions, all sounded like human productions. The Praat script, the source samples, and all stimuli can be downloaded from [URL redacted for anonymous review].

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 = 68 stimulus pairs.

Subjects were a convenience sample, recruited from a graduate Computer Science class at a university in the United States, located [details omitted to allow anonymous review]. 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 all 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
Table 2: Number of more-positive judgments. (Hypothesis 1)

<table>
<thead>
<tr>
<th>prosody source</th>
<th>neutral</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>1</td>
<td>71</td>
</tr>
<tr>
<td>positive</td>
<td>7</td>
<td>65</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>syllable-</th>
<th>overall</th>
<th>localized</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>17</td>
<td>55</td>
</tr>
<tr>
<td>100%</td>
<td>11</td>
<td>61</td>
</tr>
</tbody>
</table>

Based on the screening question, and two for not providing judgments for all stimuli. The 18 remaining subjects included 6 females.

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 stepwise comparisons. All differences are significant, supporting Hypothesis 2. We also note that discriminating between adjacent members of the eight step series was harder than discriminating 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 than when they were present throughout the stimulus, and 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, motivated by a concern that the use of stimuli derived from samples by one of the authors 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%.

6. SUMMARY AND DISCUSSION

Thus we find that this High-Long-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 whether the same results hold for uses of this prosodic construction in ongoing dialog. One might also measure the relative contributions of the 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 [9]. 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, or whether the contributions are compositional.

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 XXX...
7. REFERENCES