Some articulatory and acoustic changes associated with emphasis in spoken English

D. Erickson¹, K. Maekawa², M. Hashi³, J. Dang⁴

¹Gifu City Women’s College, Gifu, Japan; ²National Language Research Institute, Tokyo, Japan; ³Human Information Processing Laboratories, ATR, Kyoto, Japan

ABSTRACT

In order to understand better the relationship between stress and tonal patterning of prosodic changes in spoken English, this paper examines articulatory and acoustic data of a large number of American English speakers producing prosodic minimal pairs on high and low-front vowels. Articulatory and acoustic measurements (from the x-ray microbeam database, University of Wisconsin) of 45 American English speakers’ emphasized and unemphasized vowels /i/ and /æ/ were analyzed. The results confirm previous findings that articulation and acoustics (F1, F2, and F0) change as a function of emphasis. For both emphasized /i/ and /æ/, the jaw opens more, accompanied by more extreme tongue dorsum articulation as well as more extreme formant frequencies. Results of an F0 correlation with the articulatory measures suggest there may be at least two strategies for producing emphasis: one is to rely more on F0 whereas the other is to rely more on jaw opening. These findings are relevant to developments of phonological models of articulation, and especially to the C/D Model with its view of prosody as the basis for phonetic implementation of spoken English.

1. INTRODUCTION

Prosody of a language involves a combination of stress patterning and tonal patterning e.g., [1]. Changes in stress patterning, including assignment of emphasis, largely involve rhythm changes, which, according to the C/D Model [2], are associated with changes in supraglottal articulation, specifically, jaw opening [3]. A source of experimental support for this viewpoint comes from studies that report emphasized vowels, as compared with their unemphasized counterparts, have lower jaw position, accompanied by more extreme tongue dorsum articulation, as well as more extreme formant frequencies [e.g., 4,5,6].

With regard to changes in tonal patterning, these are largely related to F0 changes and generally are associated with changes in laryngeal configuration. Increased F0 is also a correlate of increased prominence or stress [e.g., 7]. It is not clear, however, exactly what the connection is between stress patterning and tonal patterning. More specifically, it is not clear what the relation is between increased F0, enhanced formant frequencies, increased jaw opening, and more extreme tongue dorsum articulation, other than that these are all characteristics of emphasis in spoken English. The purpose of this paper is to examine articulatory and acoustic data of a large number of American English speakers producing prosodic minimal pairs on high and low-front vowels in order to better understand the relationship between stress patterning and tonal patterning of prosodic changes in spoken English.

2. METHODS

Articulatory and acoustic data were obtained from the x-ray microbeam database (XRMB-SPD) [8] at the University of Wisconsin from 52 American English speakers for emphasized and unemphasized vowels. This paper reports on acoustic and articulatory measurements of emphasized and unemphasized vowels /i/ and /æ/ from sentences read by 45 (26 female, 19 male) AE speakers with 4 repetitions each.

Table 1. Utterance Types (emphasized words in capital letters; vowels measured in words underlined)

We look at the x-y movement tracings recorded from the movement of the spherical gold pellets (2.5-3 mm in diameter) attached to (1) the lower incisor (jaw) and (2) the second and third of four pellets (T2 and T3) attached along the longitudinal sulcus of each speaker’s tongue. The x-axis running backwards in negative numbers with increasing magnitude corresponds to the intersection of the midsagittal plane and the maxillary occlusal plane, and the origin of the coordinate system is the lowermost edge of the maxillary incisor. The y-axis is normal to the maxillary occlusal plane, intersecting the plane at the origin. Using a MATLAB program we compared measurements of F1, F2, jaw x-y and two positions on the tongue dorsum (T2 and T3) x-y at the time of lowest jaw position for the emphasized and unemphasized vowel during the target words. In addition, peak F0 measurements within the vowel were made using ESPS WAVES+.

3. RESULTS

In order to show sample results, plots of mean acoustic and articulatory measures for an American English speaker (female) are shown in Fig. 1. F2 vs. F1 are shown in the top left panel of the figure. The values of the x-y axes run from high to low in order to display the F1-F2 coordinates in the acoustic vowel space. Emphasized /i/ has a lower F1 and higher F2, making the vowel higher and further forward in the vowel space than its unemphasized counterpart. Emphasized /æ/ has a higher F1 and F2, making the emphasized vowel lower and further forward in the vowel space than its unemphasized counterpart.

The tongue dorsum (T2 and T3) x-y positions are shown in the top right and bottom left panels, respectively. The values of the x-y axes are displayed as if the speaker were facing to the
left of the page. The left-most bottom corner of the figure represents forward and downward position of the tongue dorsum and jaw. This orientation allows us to display the x-y articulator coordinates in an “articulatory” vowel space which can be compared with the acoustic vowel space. For the emphasized high front vowel, this speaker showed a more upward and forward position of the tongue, whereas for the emphasized low front vowel, the tongue position is more downward and forward. Tongue movement for emphasized vowels is in the direction of the phonological vowel specification: more raised/forwarded for the emphasized high front vowel /i/ and more lowered/forwarded for the low front vowel /ae/.

For jaw position for both high and low vowels, the speaker showed more downward movement.

To summarize, this speaker shows an acoustic and articulatory pattern for emphasized vowels similar to that reported in other recent studies: increased jaw opening, extreme tongue dorsum articulation in the direction of the phonological specification of the vowel, and extreme formant frequency coordinates.

In order to look at results for all the speakers, “difference-values” between emphasized and non-emphasized vowels were calculated for each acoustic and articulatory measure for each speaker. The averaged difference values between emphasized and unemphasized /i/ and /ae/ for male and female speakers separately are shown in Table 2 below. A one-sample t-test testing against the null hypothesis of mean difference =0 with alpha = 0.05 was done. A single asterisk indicates p<.05; and double asterisks, p<.01.

<table>
<thead>
<tr>
<th>i</th>
<th>N</th>
<th>V</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>T2x</th>
<th>T2y</th>
<th>T3x</th>
<th>T3y</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>26</td>
<td>[i]</td>
<td>61.7**</td>
<td>2.9</td>
<td>209.3*</td>
<td>0.6</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>m</td>
<td>18</td>
<td>[i]</td>
<td>42.3**</td>
<td>19.9</td>
<td>220.0</td>
<td>-0.3</td>
<td>0.2</td>
<td>0.9</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>f</td>
<td>26</td>
<td>[e]</td>
<td>68.0**</td>
<td>143.2*</td>
<td>244.4*</td>
<td>0.9</td>
<td>1.9**</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4**</td>
</tr>
<tr>
<td>m</td>
<td>19</td>
<td>[e]</td>
<td>55.5**</td>
<td>115.9*</td>
<td>166.1</td>
<td>0.9</td>
<td>2.9*</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4**</td>
</tr>
</tbody>
</table>

Table 2. ‘Difference’ values, with measures from unemphasized vowels subtracted from emphasized vowels.

Negative formant values indicate lower formants for emphasized vowels; negative x-position values indicate more fronting for emphasized vowels; and negative y-position values indicate more lowering for emphasized vowels.

The results show that there are significant differences between emphasized and unemphasized vowels for the speakers as a group in terms of their acoustic and articulatory characteristics. Peak F0 and jaw-y (jaw opening) are the two measures that consistently show a significant difference between emphasized and unemphasized vowels. It is interesting to note that jaw opening increases for emphasized vowels, regardless of the vowel height, as was also reported for /ai/ and /ae/ [4,5,6].

However, the only significant difference in tongue dorsum movement for all the speakers is for T2-y, which is lower for the low vowel (for both male and female speakers) and for T3-y, which is higher for the high vowel (only for female speakers). Note that significantly changed (raised) tongue dorsum (T3) for female speakers of emphasized high vowels is associated with significantly changed (raised) F2 for these speakers. Also, significantly changed (lowered) tongue dorsum (T2) for all speakers of emphasized low vowels is associated with significantly changed (raised) F1 for all speakers (and increased F2 for male speakers.)

The relation between changes in supraglottal articulation to changes in vocal tract formant frequencies has been expounded by a number of researchers [e.g. 10,11], and more recently, specifically with regard to production of emphasized high and low vowels [4].

However, it is not immediately clear what the connection is between increased jaw opening and increased F0 for emphasized vowels. Increased F0 is a function of increased laryngeal muscle tension, resulting in stretched vocal folds. If anything, we would expect to find lowered F0 with a more open jaw. The anatomical connection with the hyoid bone to both the jaw and the larynx is such that lowering the jaw may tend to push the hyoid back, which, all other things being equal, may allow the larynx to lower and tilt in such a way as to shorten the vocal folds, allowing F0 to be lower [9].

In order to examine the relation between F0 and articulation, a correlation analysis was done. A Pearson product-moment correlation coefficient between F0 and all available articulation measurements (i.e., x- and y-values of Jaw, T2, and T3) was calculated. Each coefficient was computed from eight data points, because a speaker repeated emphasized and unemphasized utterances four times each (unless there were missed values). As a result, we had 492 coefficient values (42 speakers * 6 articulatory measures * 2 vowels – 12 combinations with missed values), in which 130 coefficients were statistically significant at least at alpha = .05 level.

Looking at this data, we see a clear tendency for speakers who showed significant correlation between F0 and jaw-x or jaw-y to also show significant correlation between F0 and tongue position. Conversely, speakers who did not show significant correlation between F0 and jaw position showed non-significant correlation between F0 and tongue position. Moreover, significant correlation was found most frequently between F0 and jaw-y. Of the 45 speakers, 18 speakers and 17 speakers showed significance higher than alpha= .05 level for /i/ and /ae/, respectively.

An ANOVA was then done, with peak F0 as the dependent variable and Emphasis (0 or 1) and F0CORRS (high or low) as the independent variables. Two ANOVA were run separately for male and female subjects. The results in terms of Least Square Means (LSM) and Standard Errors (SE) are shown in Tables 3 and 4 below. The mean F0 value includes both /ae/ and /i/.

<table>
<thead>
<tr>
<th>F0CORRS</th>
<th>LS Mean</th>
<th>SE</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>119.5 Hz</td>
<td>4.03</td>
<td>48</td>
</tr>
<tr>
<td>Low</td>
<td>141.2 Hz</td>
<td>3.49</td>
<td>64</td>
</tr>
</tbody>
</table>

Table 3. ANOVA with F0 as dependent variable, and Emphasis and F0CORRS as independent variables for male speakers.
<table>
<thead>
<tr>
<th>F0CORR$</th>
<th>LS Mean</th>
<th>SE</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>215.4 Hz</td>
<td>6.96</td>
<td>61</td>
</tr>
<tr>
<td>Low</td>
<td>243.7 Hz</td>
<td>4.21</td>
<td>167</td>
</tr>
</tbody>
</table>

Table 4. ANOVA with F0 as dependent variable, and Emphasis and F0CORR$ as independent variables for female speakers.

These results show the clear tendency that subjects who showed high significance in F0-jaw correlation tend to show lower peak F0 values, and conversely, subjects who showed no significance tend to have higher peak F0. A possible interpretation for this is that there are two different strategies for manifestation of emphasis. Some people like to use both articulation and F0 in the manifestation of emphasis (high correlation subjects) and these people do not raise F0 very high because emphasis is also manifested by spectral information (i.e., articulation). On the other hand, there are people who rely more upon F0 than articulation (low correlation). The peak F0 of this latter group is higher than the previous group because F0 is the principal device of emphasis manifestation for the subjects in the latter group.

Based on this hypothesis, we analyzed the difference of jaw position between the “high” and “low” F0-jaw opening correlation groups. The results for the male speakers are shown in Figure 2. These graphs show clearly the tendency that high-correlation subjects (denoted as “High” in the graphs) show greater difference of jaw opening gesture between emphasized (1) and non-emphasized (0) utterances. The same tendency was observed for female speakers, but the results are not shown here.

These results suggest the following: (1) Some speakers show high correlation between peak F0 value and jaw opening (high-correlation subjects), and others do not show any correlation at all (low-correlation subjects). (2) The peak F0 value of emphasized utterances uttered by low-correlation speakers is statistically significantly higher than that of high-correlation subjects. (3) Difference of articulatory gesture due to emphasis, as measured by the extreme (maximum opening) value of Jy, is greater in high-correlation speakers than in low-correlation speakers.

Our tentative conclusion is that there at least are two different strategies of emphasis realization. Some speakers (low-correlation speakers) rely more upon F0 than articulation, and some others (high-correlation) rely more on articulation.

This conclusion makes sense in terms of the physiology of the vocal tract, and the anatomical connection of the jaw to the larynx via the hyoid bone, as mentioned above.

4. SUMMARY

The findings from this study suggest that both increased F0 and increased jaw opening are significant characteristics of emphasis in English. Although all speakers tended to show both increased F0 and increased jaw opening, there seems to be an interspeaker difference in fine-tuning for producing emphasis. Those speakers who increased their F0 to about 140 Hz (for men) or 240 Hz (for women) tend to open their jaw for emphasis less than do those speakers who show less of an increase in F0 for emphasized vowels (about 120 Hz for men, and about 215 Hz for women). These speakers show more jaw opening for emphasized vowels.

This finding is consistent with physiological accounts of the interaction between jaw opening and vocal fold tension, and as such suggests a physiological underpinning for the interaction of stress patterning and tonal patterning characteristics of English prosody. It also suggests there may be interspeaker variation in terms of preferences for tonal patterning vs. stress patterning with regard to producing English prosody. These ideas are currently being investigated specifically as they apply to the development of phonological models of articulation. The findings seem especially relevant to the C/D model with its treatment of prosody as the base function of phonetic implementation of spoken English.

5. ACKNOWLEDGEMENTS

We wish to thank Osamu Fujimura for his many helpful discussions with regard to the contents of this paper.

6. REFERENCES

Fig. 1. Data plots for one female speaker (S52). Emphasized vowels indicated by capital letters.

Males: Word="these"

Fig. 2. Results of ANOVA test with the two groups of male speakers (High F0 and Low F0) where Jaw-y is dependent variable, and Emphasis is independent variable.