PHYSIOLOGICAL CORRELATES OF GLOBAL AND LOCAL PITCH RANGE VARIATION IN THE PRODUCTION OF HIGH TONES IN ENGLISH

Mary E. Beckman[a,b], Donna Erickson[c,d], Kiyoshi Honda[d], Hiroyuki Hirai[d], and Seiji Niimi[e]
[a]Ohio State University, Linguistics, Columbus, Ohio, USA
[b]ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan
[c]Ohio State University, Speech & Hearing Sciences, Columbus, Ohio, USA
[d]ATR Human Information Processing Research Laboratories, Kyoto, Japan
[e]Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan

ABSTRACT

Subglottal pressure and cricothyroid muscle activity level were measured to explore the physiological bases of the systematic differences in measured F0 peak value that previous studies have shown for high tones in English pitch accents. There was a strong correspondence between EMG level and F0 variation associated with pitch accent type, but subglottal pressure differences were associated primarily with the overall loudness of the utterance.

INTRODUCTION

The modeling of F0 patterns in large sets of utterances elicited under controlled conditions of variation in intonational function and related prosodic structure has made [+Hightone] ("H") one of the best understood phonetic features. Such modeling studies have shown that the relative F0 target values associated with H tones are sensitive to differences in the tones' paradigmatic intonational categories, their position in the sentence or larger discourse segment, and the prominence of the associated word or phrase. Moreover, these relationships are preserved across global variation such as the overall F0 increase associated with a louder voice. For example, Liberman & Pierrehumbert [1] studied nuclear H* pitch accents in two successive intonational phrases in an English declarative sentence produced many times in each of ten different degrees of overall vocal effort and in the context of two dialogues which contrasted the relative informational prominence of the accented words. When the accented word in the second phrase was the more prominent "answer" focus, the target F0 peak for its H* pitch accent was slightly higher than that for the first, but when it was the weaker "background" focus (which repeated information in the context), its peak was substantially lower. These relationships could be modeled by two effects — a "final lowering" of the second peak and a proportional raising of the "answer" peak — both expressed as constant distances above a "reference line" for the variation in the utterance's global pitch range as the speaker increased or decreased overall loudness. Analogous patterns have been found in other similar studies of English and other languages (e.g. [2, 3, 4])

This paper examines the physiological bases of such fundamental frequency relationships for English. We recorded subglottal pressure and electromyographic activity level from the cricothyroid muscle as a talker produced a sentence in three different intonation contours in which we could examine F0 peaks associated with H tones exemplifying prosodic contrasts that have been well-studied in the experimental literature. The talker produced dozens of tokens of each intonation type in each of three overall vocal effort levels. The results suggest that the talker adjusted cricothyroid activity level to control the variation associated with simple versus rising nuclear pitch accents, but the variation in peak height associated with global pitch range seems to originate primarily in the differences in subglottal pressure associated with soft versus normal versus loud voice.

METHOD

The data were originally recorded as part of our ongoing examination of L(ow) tones in English intonation [5, 6]. We re-examined the three intonation contours listed in Table 1, which had two F0 peaks for pitch accents varying in

accent type and in associated qualitative stress level. Contours (2) and (3) have the same stress pattern, and differ only in the nuclear accent type, whereas contour (1) has two nuclear pitch accents in successive phrases. (See [7] for our analysis of the qualitatively stronger and weaker stress levels corresponding to nuclear versus prenuclear accentual position within an intonational phrase, and [8] for the inventory of phrase tone and accent types and their meanings.) The talker was the first author, a female native speaker of American English, who could produce the contours consistently many times at each of soft, normal, and loud voice.

Table 1. Discourse contexts and intonation contours, with target tones underlined.

1. Two phrases demarcated by a L- phrase accent, with H* nuclear accent in each.

```
Do you have any pasta dishes
less fattening than fettucine
Alfredo?
We have a lean, mini-noodle
                          dish.
     H* L-        H*     L-L%
```

2. Single intonational phrase, with L*+H type in nuclear accent position, in the canonical contrastive contour.

```
Do you have any bean dishes
other than this couscous thing?
We have a lean mini-noodle with
                         beans.
     H*      L+H*        L-L%
```

3. Single intonational phrase, with scooped L*+H type in nuclear accent position, in a contour indicating pragmatic uncertainty.

```
Do all of your rice dishes have
this fatty meat sauce?
We have a lean mini-noodle dish.
     H*        L*+H      L-H%
```

Subglottal pressure (Ps) was recorded from a transducer mounted at the tip of a catheter inserted through the nose, and cricothyroid muscle activity level (CT) was recorded in a separate session, using subcutaneous hooked-wire electrodes. We analyzed about 20 tokens per cell for the Ps dataset and about 24 for the CT dataset. (See [5] for more details concerning recording methods, smoothing, and so on.) There was always a clear peak in the F0 contour and in the smoothed Ps contour that we could measure for each of the two target H tones. The smoothed CT data, however, often showed a series of peaks during the F0 rise, so instead of choosing any single maximum value, we instead took a measure of the average value over the F0 rise, from the minimum in the [h] in *have* (for the first H tone) or from the minimum at the preceding L target (for the second H tone).

RESULTS AND DISCUSSION

Figure 1 plots the value for the second F0 peak as a function of that for the first F0 peak in the Ps dataset. The datapoints for contour (1) cluster just below the x=y line, indicating that the second peak is lowered somewhat relative to the first. Compared to the analogous plots in [1], the lowering is substantially less than it would be if the second phrase were subordinated to the first in an answer-background focus sequence, suggesting that the data in Figure 1 show a pure "final lowering" effect. The datapoints for contours (2) and (3), by contrast, show a second peak that is substantially higher than the first. (The scatterplot for the CT dataset is virtually the same except for a slightly smaller dynamic range over the three voice effort levels.)

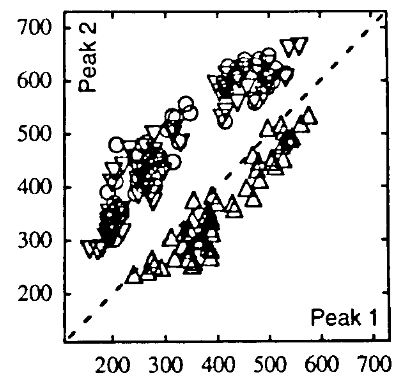

Figure 1. F0 value in second peak as a function of that in first peak for contours (1) Δ, (2) O, and (3) ∇. Values for both axes in Hz.

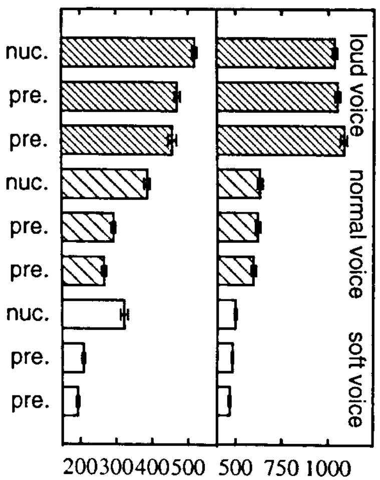

Figure 2. Means for F0 peak (left, Hz) and Ps peak (right, arbitrary units) for the first pitch accent in the Ps dataset, averaged over vocal effort level and over contour, with types (1), (2), and (3) in descending order along the y-axis.

Figures 2-3 give mean values for these F0 peaks and for the associated Ps peaks. They show that the difference among contours in Figure 1 is a function both of the second peaks being higher in contours (2) and (3) than in contour (1), and of the first peaks being lower. The higher second peaks in contours (2) and (3) suggests an inherently greater tonal prominence for the rising accents than for the single-tone accent. This is in keeping with the accents' meanings. Both L+H* and L*+H differ from simple H* in explicitly contrasting the focused discourse entity to other values in the presupposition set [8]. The lower first peaks in types (2) and (3) cannot be attributed to accent type, since all three contours have H* here, but it is in keeping with the different associated stress levels. A prenuclear accent is less prominent than a nuclear accent.

The corresponding Ps values in the right-hand panels of the figures show differences in peak Ps means that correspond well to the differences in the F0 means within a loudness level, differences that we have related to the

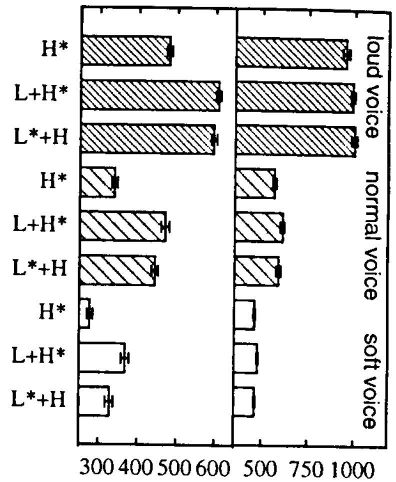

Figure 3. Means for F0 peak (left) and Ps peak (right) for the second pitch accent, arranged as in Figure 2.

tonal prominence inherent to the accent type (Figure 4) or to the associated stress level of that prosodic position (Figure 3). However, the correspondence is not completely consistent. In the loud-voice productions, the prenuclear H* of contour (3) has a higher mean Ps value than the nuclear H* of type (1). Moreover, the variation among accent types and stress levels is very small by comparison to the enormous increase in Ps in going from soft to normal to loud voice.

Figures 4-5 show the analogous mean values for F0 peak and average CT over the F0 rise for the productions in the CT dataset. The F0 peaks showed the same pattern of effects as in the other dataset, but the CT values showed two different patterns, neither exactly like the pattern of effects for the Ps values. For the first accent peak, the average CT value over the F0 reflected both the F0 variation due to overall pitch range and the variation within a pitch range due to local accentual prominence (although the correspondence to the latter was best at soft voice). For the second peak, however, there was only the variation due to accent type, and virtually no difference in means corresponding to the differences in F0

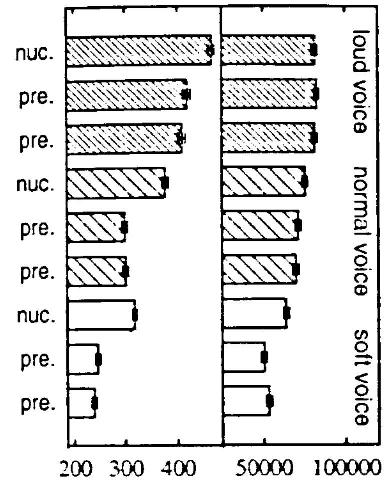

Figure 4. Means for F0 peak (left, Hz) and average CT over the F0 rise into the peak (right, arbitrary units) for the first pitch accent in the CT dataset, contour types as in Figure 2.

values across the three overall vocal effort levels.

Although further experimentation is necessary to understand the pattern of CT activity for the first F0 peak, differences in global pitch range for different overall vocal efforts seems to be related primarily to the different associated subglottal pressures. By contrast, the pattern of CT activity for the second F0 peak suggests that at least some aspects of local pitch control can be differentiated physiologically from the more global effect.

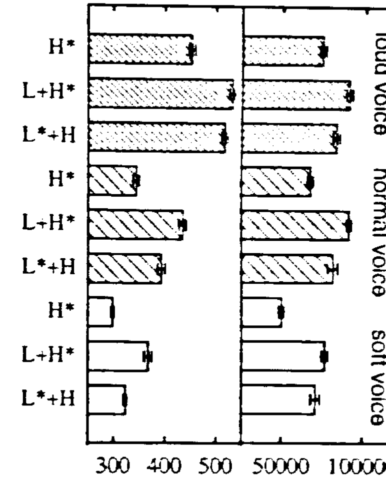

Figure 5. Means for F0 peak (left) and average CT over the F0 rise (right) into the second pitch accent, as in Figure 4.

REFERENCES

[1] Liberman, M., & Pierrehumbert, J. (1984). "Intonational invariance under changes in pitch range and length," In M. Aranoff & R. Oehrle, eds, *Language Sound Structure*, Cambridge, MA: MIT Press.
[2] Bruce, G. (1982). "Developing the Swedish intonational model," *Working Papers, Lund*, no. 22, pp. 51-116.
[3] van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992). "Downstep in Dutch: implications for a model," In G. J. Docherty & D. R. Ladd, eds, *Papers in Laboratory Phonology II: Segment, Gesture, Prosody*, Cambridge, UK: Cambridge University Press.
[4] Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese Tone Structure*, Cambridge, MA: MIT Press.
[5] Erickson, D., Honda, K., Hirai, H., Beckman, M. E., & Niimi, S. (1994). "Global pitch range and the production of low tones in English intonation," *ICSLP'94*, vol. 2, pp. 651-654.
[6] Erickson, D., Honda, K., Hirai, H., & Beckman, M. E. (in press). "The production of low tones in English intonation," To appear in *Journal of Phonetics*.
[7] Beckman, M. E., & Edwards, J. (1994). "Articulatory evidence for differentiating stress categories," In P. A. Keating, ed., *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge, UK: Cambridge University Press.
[8] Pierrehumbert, J. & Hirschberg, J. (1990). "The meaning of intonation contours in the interpretation of discourse," In P. R. Cohen, J. Morgan, & M. E. Pollack, eds, *Intentions in Communication*, pp. 271-311. Cambridge, MA: MIT Press.