ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

GLOBAL PITCH RANGE AND THE PRODUCTION OF LOW TONES IN ENGLISH INTONATION

Donna Erickson[a], Kiyoshi Honda[b], Hiroyuki Hirai[b], Mary E. Beckman[c], Seiji Niimi[d]

[a]Department of Speech & Hearing Sciences, Ohio State University, Columbus, OH 43210-1002 USA
[b]ATR Human Information Processing Research Laboratories, Kyoto, Japan
[c]Department of Linguistics, Ohio State University, Columbus, OH 43210-1298 USA
[d]Research Institute of Logopedics & Phoniatrics, University of Tokyo, Tokyo, Japan

ABSTRACT

In addition to the well-studied high (H) targets, such as the H* peak accent in many declarative sentence, English intonation contours also include many low (L) tones, such as the L* target on the most stressed syllable in a canonical yes-no question contour. To understand better the behavior of such tone targets, we recorded three native speakers of English as they produced sentences containing various L types in soft, normal, and loud voice. For two speakers, we recorded activity from the infrahyoid strap muscles (SH), and for the third we recorded subglottal pressure (Ps). Results showed consistent variation in minimum fundamental frequency (F0) values associated with the L tone, and also consistent variation across the global pitch ranges for the three voice effort levels. Within a given pitch range, the level of associated SH activity was inversely correlated with the F0 value, but the correlation overall was less good because there is no overall variation in the SH level corresponding to the global range variation. Correlation improved when SH values for loud tokens were adjusted downward and those for soft tokens adjusted upward. The pattern of Ps values for the third speaker suggests an explanation. There were large differences in Ps value across the pitch ranges, against which speakers may adjust their SH activity to maintain a constant target F0.

I. INTRODUCTION

Fundamental frequency (F0) values of H tones (local targets **high** in the pitch range) are predicted easily by modeling a simple interaction between local tonal commands and the more global specification of pitch range values associated with degree of phrasal prominence or overall vocal effort. For example, Liberman and Pierrehumbert [1] showed that, in a common English "downstepping" contour, each successive H-toned accent target is proportionately lowered relative to the preceding pitch accent, and this proportion is constant over different global pitch ranges for soft, normal, and loud voice. By contrast, much less is known about L tones (targets **low** in the pitch range.) In English intonation, these include the L* nuclear pitch accent on the most stressed syllable in the "yes-no question contour" and the L- phrasal tone in the "declarative contour" that simultaneously marks the end of the phrase and the fall from the peak accent.

In a previous paper [2], we examined the behavior of such L tones in a corpus of sentences produced in three voice levels by a native speaker of American English. We found that different L types showed consistent differences in local F0 value, just as different H tones do. For example, the L- phrase accent marking a medial phrase boundary in one sentence type was consistently higher than the L* nuclear pitch accent in another type. Moreover, these values shifted with the shift in overall pitch range associated with the global voice-effort level; the F0 value for the L* was higher in the larger overall pitch range of the loud utterances, and lower in the smaller pitch range of the soft utterances. In the same experiment, we also analyzed levels of associated electromyographic activity in the infrahyoid strap muscles (SH). Within a given overall pitch range, the SH level was inversely correlated with the tone's F0 value. However, the correlation overall was less good until we subtracted a constant from the SH values for loud voice productions and added the same constant to the SH values for soft voice productions. We speculated that this constant reflects adjustments by the speaker to maintain the target F0 level against probably different subglottal pressures associated with softer or louder speaking voice.

This paper reports on a new experiment designed to extend and explain these earlier findings. We reproduced the experiment by recording F0 values and associated SH activity levels for two more speakers, and we re-recorded our original speaker with the same corpus, this time measuring her subglottal pressure (Ps) rather than strap muscle activity. The results for our two new speakers replicated those of our original study, and the new physiological measure for our original speaker supports the interpretation offered earlier for the nonlinear relationship between SH and F0 across the different voice levels.

II. METHOD

Our original speaker, MB, is the fourth author. We had her produce the same corpus of utterance types as in our original experiment on SH activity, this time using a digital pressure monitor (Camino Labs model 420) to record subglottal pressure from a transducer mounted at the tip of a catheter designed to measure arterial pressure (Camino Labs model 110-4). We placed the pressure transducer below the glottis by inserting the catheter through the nose and velopharyngeal port to pass down the pharynx. To remove the low-amplitude periodic component due to vocal fold vibration, we smoothed the Ps signal by calculating a moving average over a 20 ms triangular window.

The two new speakers were DE, another female speaker of American English who is also the first author, and NC, a male speaker of British English. We recorded SH activity for DE using hooked-wire electrodes placed subcutaneously in the muscle fiber [3]. For NC we used a column of electrodes placed on the surface of the throat in the region where the infrahyoid strap muscles run, the same method as for MB in our original experiment. For more details on this recording technique, and for a general description of the smoothing and other processing of the SH data, see [1].

Fig. 1 lists the original corpus of sentence types, along with a description of the intonation pattern produced by MB, and the discourse contexts we provided in trying to elicit the same contours from our two new speakers. DE consistently produced the desired contour except in

1. Nuclear L* pitch accent, in the canonical yes-no question contour:
What would you like for lunch?
D'you have a lean mini-noodle dish?
L* L* H- H%

2. Nuclear L*+H scooped pitch accent, 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%

3. Nuclear L+H* rising peak accent, in the canonical contrastive emphasis 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%

4. Sentence-medial L- phrase accent marking the boundary between two intermediate phrases:
Do you have any pasta less fattening than fettuccine Alfredo?
We have a lean, mini-noodle dish.
H* L- H* L- L%

5. Pre-nuclear H+L* pitch accent, in a contour indicating an obvious pragmatic inference (here of resignation):
Edward, you know we're not supposed to eat meat at lunchtime.
Oh, all right. We'll have the lean mini-noodle with beans.
L* H* H+L* H+L* L- L%

Fig. 1. Discourse contexts and intended intonation contours for the five sentence types. The target sentence is the second one, and the target L tone is underlined in the transcription of the contour. The theory of English intonation assumed in this paper [4] models the relationship between intonation contours and associated text in terms of ***intonational phrases****, each of which contains one or more* ***intermediate phrase****. For each intermediate phrase, the contour contains one or more* ***pitch accents*** *(pitch events associated with intonationally prominent syllables) followed by a H- or L-* ***phrase accent*** *(a demarcative tone that fills the region between the last pitch accent and the phrase edge). The full intonational phrase is demarcated also by a final H% or L%* ***boundary tone****.*

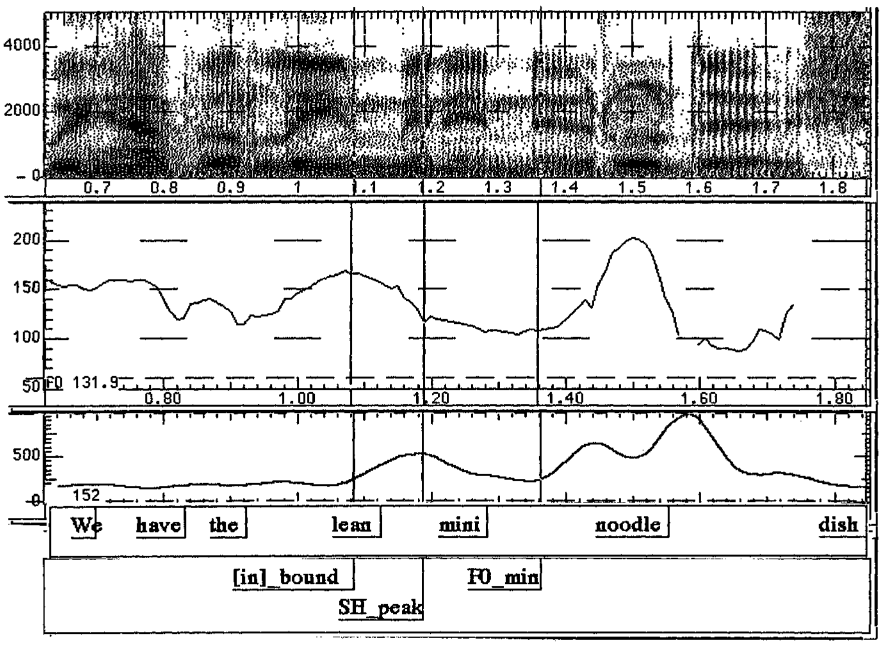

Fig. 2. Spectrogram, F0 contour, and SH activity for a representative token of sentence type 2 produced in normal voice by speaker NC.

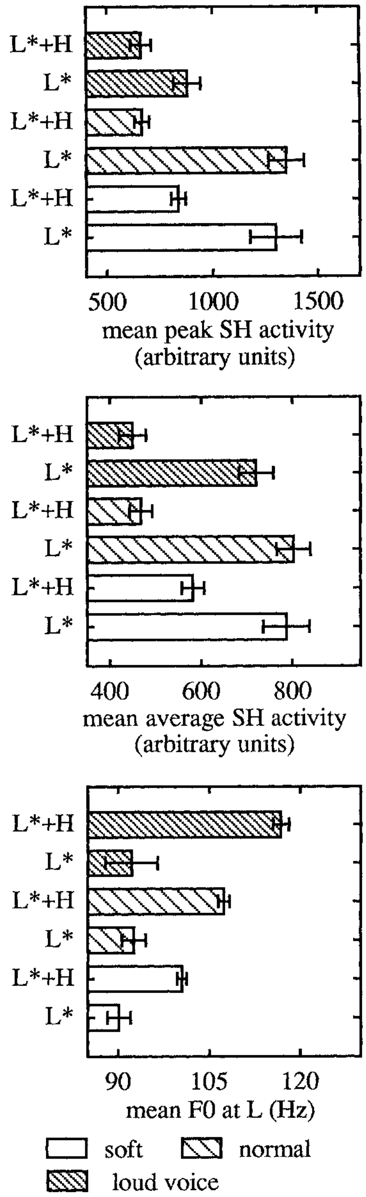

Fig. 3. Mean values for F0 minimum (top), peak SH (middle), and average SH (bottom), for each L type and voice level. Error bars show standard error. Speaker NC.

sentence type 4, where in about 30% of the tokens she produced a single phrase with nuclear H* pitch accent only on *lean*. Thus we have fewer tokens of this type than of the target tones in types 1, 2, 3, and 5. NC was less practiced at producing the targeted intonation contours on demand. For him we have no tokens of H+L* or L+H*, and twice the targeted number of tokens of L*+H.

As in our original experiment, we asked the speakers to produce multiple tokens of each sentence in three different self-selected levels of vocal effort (soft, normal, and loud voice) so that we could examine the interaction of tonal specification with global pitch range. Fig. 2 shows a sample display of an utterance produced by NC in normal voice. The spectrogram at the top and the two windows showing the F0 and SH activity level traces are time aligned to the bottom window, which shows the labels that we made in order to measure the relevant F0 and SH events. We measured the minimum F0 value for the targeted L tone of the L*+H pitch accent ("F0_min"), and the value of the peak associated muscle activity ("SH_peak"). We also measured the average SH activity over an interval before between the offset of the vowel [i] in *lean* ("[in]_bound") and the minimum F0. We devised this "average SH" as an alternative measure of EMG activity, because our original speaker, and speaker DE also, differed from NC in not having a single clear large peak in this interval. Instead their productions showed a succession of several peaks, like the two peaks for the following L- in NC's utterance in Fig. 2, which we interpret as activity for the L tone modulated by small perturbations for jaw-related activity. By comparing NC's SH peak values to his SH average values, we thus can assess the reliability of the latter as a measure of activity associated with the target L tone. For MB's new data, we measured the F0 at the same minimum value, and we measured the subglottal pressure level at the associated minimum in the Ps trace.

III. RESULTS

Fig. 3 shows the mean values for the three measures taken for speaker NC, averaged over the tokens of each target tone type at each voice effort level. The mean peak SH values show the same relationship between the two tone types as his mean average SH values. Both sets of SH values also show the same inverse relationship to the mean F0 value that we saw in our original experiment, supporting our choice of the average SH measure for speakers MB in that experiment and DE in the current experiment.

Fig. 4 shows the analogous mean values for the two measures taken for DE. This figure and Fig. 3 both show patterns essentially identical to those for MB in our original experiment. Both new speakers have noticeable differences among the different tone types, with the L* target of type 1 yielding very low mean values and DE's H+L* target in type 5 yielding high mean values, the same relationship as for MB. Moreover, mean values for the average SH level closely mirror the differences among the mean values for the minimum F0 at any given range, with the L* target showing the highest SH level and H+L* the lowest. Also, the F0 values show a consistent shift with overall voice effort, but this shift is not reflected in the SH values. That is, the local minimum F0 values for the L targets increase with the increase in overall pitch range from soft to normal to loud voice, but the related SH values do not show the converse decrease.

In our earlier paper describing comparable data for MB, we examined the relationship between the average SH value and the minimum F0 value in more detail by plotting the values token by token. When we did this separately for the three voice effort levels, and then overlaid the plots, we found that a translation of 200 units upward for the soft-voice tokens and 200 units downward for the loud-voice tokens produced an inverse exponential curve neatly describing the data. Fig. 5 shows DE's data plotted in this way, with the same translation of ±200 units along the y-axis. Although the data do not cluster as tightly as in the original experiment, there is the same exponential trend.

Previously, we interpreted the similar result for MB as suggesting that the relationships between SH activity and F0 at the different pitch ranges are indeed part of the same overall function, but that there is some kind of a shift of the "baseline" for the function from one vocal effort level to the next. We speculated that the source of this baseline shift be from several sources. We know, for example, that changes in vocal effort involve not only changes in overall F0, but also changes in overall subglottal pressure. Perhaps the speaker uses the SH more in loud voice productions to achieve the L tone target frequency against the increased subglottal pressure of the louder speech. Conversely, for the low pitch range, the smaller-than-expected level of SH activity may reflect an adjustment by the speaker to a decreased volume of airflow from the lungs.

Our motivation for measuring subglottal pressure in our new recording from speaker MB was to test the plausibility of this explanation of the baseline shift. Fig. 6 shows her mean minimum F0 values and mean Ps values at the associated pressure minimum. The minimum F0 values show the same pattern of variation as in our original experiment and as in Fig. 4 for speaker DE. The subglottal pressure values mimic this pattern exactly. The mean Ps value is low for types with low F0 targets and high for types with high F0 targets. However, by comparison to their effects on F0, the effect of tone type on Ps is dwarfed by the effect of voice effort level. That is, whereas the minimum F0 values overlap at the edges of the different pitch range, the Ps values do not. The L type with the lowest Ps value at normal voice has a higher Ps than the L

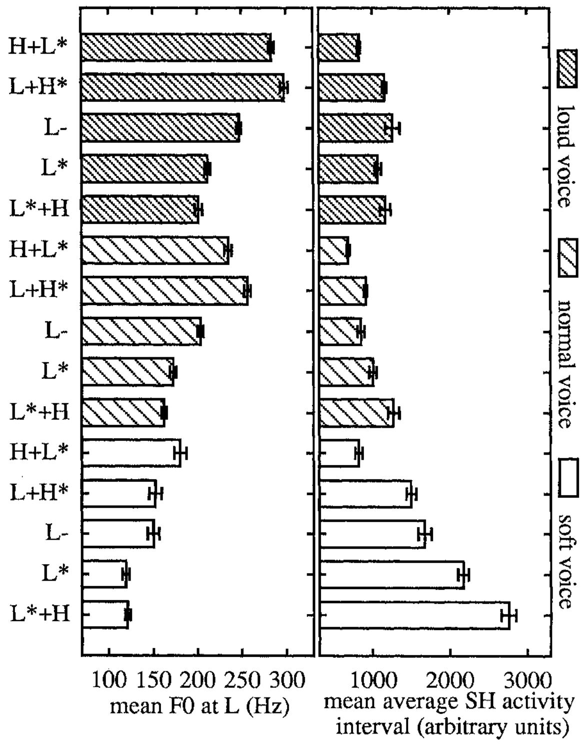

Fig. 4. Mean values for F0 minimum (left) and average SH (right) as in Fig. 3. Speaker DE.

type with highest Ps value at soft voice, and this shift is even larger between the normal and loud voice tokens. This pattern of greater effect of global pitch range on Ps is in keeping with our proposed interpretation of the SH pattern.

IV. CONCLUSION

In summary, these preliminary analyses from our new experiments replicate the results from our earlier study, and lend some support to the tentative interpretation offered there. For all three new recordings, including those for our two new speakers, there are again paradigmatic differences among the F0 values for L tone targets in English intonation contours, comparable to the differences among H tones demonstrated by Liberman and Pierrehumbert [1] and many other studies. Moreover these F0 targets show a consistent pattern of variation across the different voice effort levels, being lower in the decreased global pitch range of soft voice and higher in the increased pitch range of loud voice. As in our original experiment, the paradigmatic variation in F0 values for the different L types within a pitch range is mirrored exactly by the SH values. This was true not just of the average SH measure which we could obtain for all three speakers, but also of the peak SH measure that we could obtain for the male speaker. When the average SH value was plotted against the minimum F0 value token by token for our new female speaker, the scatterplots suggested the same baseline shift in an exponential relationship as for our original speaker's productions. Measurements of subglottal pressure for our original speaker support one possible explanation for this baseline shift — namely that the speaker adjust SH activity to maintain a proportionally constant F0 target value against the dramatically different subglottal pressures characteristic of the different voice effort levels.

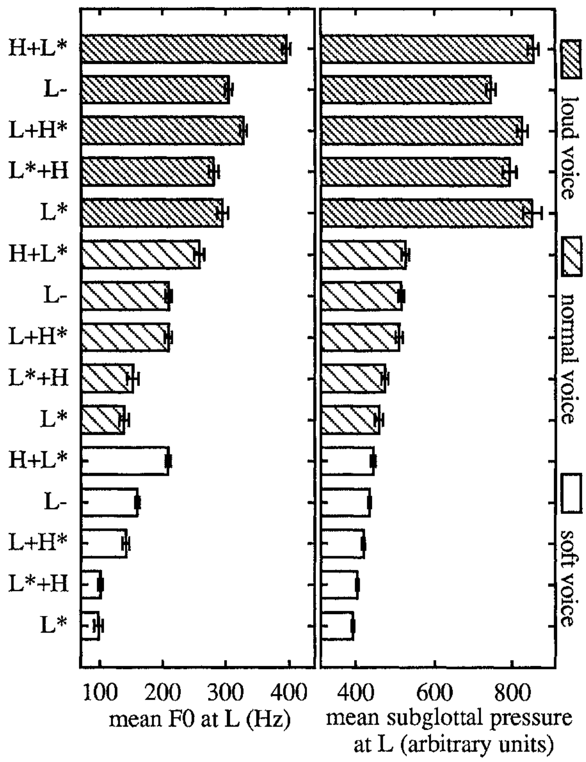

Fig. 6. Mean values for F0 minimum (left) and Ps (right), for each L type and voice level. Error bars for standard error. Speaker MB.

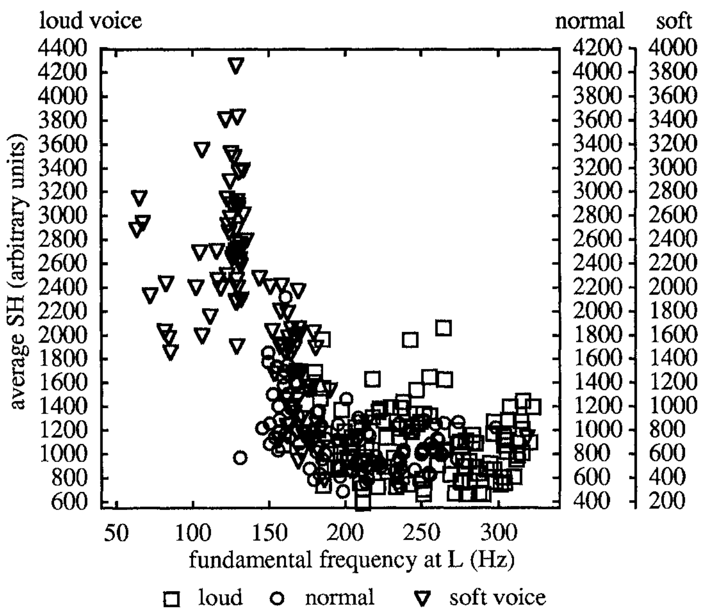

Fig. 5. Average SH values as a function of minimum F0 values for speaker DE.

ACKNOWLEDGMENTS

The surface EMG and subglottal pressure recordings were done in the ATR Human Information Processing Research Laboratories, Kyoto, Japan. The subcutaneous EMG was recorded at RILP. The work reported here was supported in part by the National Science Foundation under grant no. IRI-8858109 to Mary Beckman.

REFERENCES

[1] M. Y. Liberman, J. B. Pierrehumbert, "Intonational invariance under changes in pitch range and length," In M. Aronoff, R. T. Oehrle, eds., *Language Sound Structure: Studies in Phonology Presented to Morris Halle*, pp. 157-233, 1984.

[2] D. Erickson, K. Honda, H. Hirai, M. Beckman, "The production of low tones in English Intonation," *J. Phonetics*, in press.

[3] H. Hirose, T. Gay, M. Shome, "Electrode insertion techniques for laryngeal electromyography," J. Acoust. Soc. Am., Vol. 50, pp. 1449-1450, 1971.

[4] J. Pierrehumbert, J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," In P. R. Cohen, J. Morgan, & M. E. Pollack, eds., *Intentions in Communication*, pp. 271-276, 1990.