TR－H－023

The production of low tones in English intonation

Donna Erickson, Kiyoshi Honda
Hiroyuki Hirai, Mary E. Beckman

1993. 8. 30

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

The production of low tones in English intonation

* Donna Erickson[a], Kiyoshi Honda[b], Hiroyuki Hirai[b], and Mary E. Beckman[c]

Abstract

This paper examines the relationship between fundamental frequency target and sternohyoid activity in low tones of four different pitch accents and of a phrase boundary in English intonation contours produced at three levels of overall vocal effort. Minimum F0 values for the low targets differed as a function of paradigmatically contrasting tone type and as a function of voice effort level. SH activity level also varied as a function of tone type, in inverse relationship to the F0 value. However, it did not show the same simple relationship to variation in F0 value as a function of overall vocal effort, suggesting a shift in the baseline value due perhaps to concomitant changes in subglottal pressure or to jaw lowering for segmental effect.

* [a]Department of Speech & Hearing, Ohio State University, OH, USA. (110 Pressey Hall, 1070 Carmack Road, Columbus, OH, 43210 USA)
[b]ATR Human Information Processing Research Laboratories, Kyoto, Japan
[c]Department of Linguistics, Ohio State University, OH, USA.

1. Introduction

A useful strategy in the cross-linguistic investigation of intonation is to model fundamental frequency contours as the realization of local tonal commands which interact with the specification of longer-range pitch values for prominence and the like (e.g. Bruce, 1982; Liberman & Pierrehumbert, 1984; Shih 1988; van den Berg, Gussenhoven, & Rietveld 1992). These models can use mathematically simple functions for predicting the values of H tones (i.e., local targets **high** in the pitch range), because relationships among different H targets are observed to be proportionally constant across different overall pitch ranges. For instance, in downstepping sequences in English, each subsequent downstepped H is proportionally lower in the pitch range than the preceding H, and this proportion is constant across reduced, normal, and expanded overall pitch ranges (Liberman & Pierrehumbert, 1984). By contrast, modeling of L tones (targets **low** in the pitch range) is considerably more difficult. In English intonation, such tones include the L% boundary tone at the end of the "declarative sentence" contour and the L* pitch accent on the most stressed syllable in the "yes-no question" contour. There are apparent contradictions in the literature on the scaling of these L tones. Several studies claim that the L% tones at the ends of turn-final "declarative" contours involve a speaker-specific value that remains unchanged across variation in overall pitch range (e.g., Boyce & Menn 1979, Liberman and Pierrehumbert 1984). Other studies, however, claim that L tones do vary with changes in pitch range. Liberman & Pierrehumbert give an example suggesting that an expansion of the pitch range for emphatic prominence can make L* nuclear accents be even lower, whereas Pierrehumbert (1989) shows targets for L* nuclear accents rising with pitch range expansion for increased vocal effort.

We wonder whether these difficulties in modeling L tones might stem from the more complicated interactions between the physiological control mechanisms at work in local pitch lowering and those at work in overall pitch range manipulation. That is, modeling

H tones across different overall pitch ranges may seem easier only because the primary mechanism for producing H tone targets (the elongation of the vocal folds through cricothyroid contraction) may interact more simply with changes in subglottal pressure, which is one likely physiological mechanism for the variation in overall pitch range at different levels of vocal effort. The primary mechanism for producing L tones is not well understood. Previous studies show that the infrahyoid strap muscles are active during L tone production in such diverse languages as Thai (Erickson, 1993), the Osaka and Kumamoto dialects of Japanese (e.g. Kori, Sugito, Hirose, & Niimi, 1990; Simada, Niimi, & Hirose, 1991; Kiritani, Hirose, Maekawa, Sato, 1992), Mandarin Chinese (e.g., Hallé, Niimi, Imaizumi, & Hirose, 1990), and Swedish (Gårding, Fujimura, & Hirose, 1970). A plausible mechanism is suggested by Honda, Hirai, and Kusakawa (1993): contracting the infrahyoid muscles lowers the larynx, which rotates the cricoid cartilage downward and forward around the bend in the cervical vertebra, thus reducing the length of the vocal folds. Whatever the mechanism, we know that the infrahyoid muscles are active also during the production of nuclear L* and utterance-final boundary L% tones in English (Atkinson, 1978). In this paper, therefore, we examine fundamental frequency values and infrahyoid strap muscle activity levels during the production of contrasting L tones in English. The specific questions to be investigated are four. First and second, are there paradigmatic differences among L tones in English intonation contours comparable to the differences among H tones, and if so, is the same mechanism involved in all these low tones? Specifically is the SH involved in all, or to the same degree in all?, Third, do the L tones vary across differences in overall pitch range, and if so, how? Fourth, if there is variation in F0 level across pitch range, is this also reflected in the laryngeal control?

2. Methods

We recorded 3 American English speakers as they produced a corpus of five sentences contrasting various L tones. They produced 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 the tonal specification with global pitch range. Here we report data only for the one speaker who produced the most tokens — 16 of each type at each of the voice effort levels. We simultaneously recorded strap muscle activity using a noninvasive method developed by Yoshida, Honda, & Kakita (1993). A column of nine surface electrodes is attached along the surface of the subject's throat, and the voltage is measured across each pair of adjacent electrodes, giving 8 recording channels reflecting activity from the superficial supra- and infrahyoid strap muscles. For the speaker whose results we discuss in this paper, we observed that the lowest three channels showed similar patterns and were most consistently active during regions of low fundamental frequency. We therefore averaged over these three channels to get what we assume is a sternohyoid (SH) signal. Before averaging the three channels, we rectified the data, and after averaging we smoothed twice with a 70 ms rectangular window centered around the target sample.

Fig. 1 shows a spectrogram, F0 contour, and SH trace for a representative token of each of the sentence types. The target is always an L tone occurring somewhere in the *lean mini-noodle*. We chose the vowels in the syllables around the L to be high, so as to reduce segmental effects on the SH from jaw lowering, and the consonants to be mostly sonorants, so as to reduce segmental perturbations of the F0 contour. The F0 minimum that we measured for the target is marked with an arrow. The figure also shows the background dialogues that we used to cue the different intonation contours. The contour in sentence type 1, for example, is the familiar "yes-no question" intonation, which puts a nuclear L* on *noodle*. There is also a prenuclear L* on *lean*, but it was not difficult to locate the right minimum F0 corresponding to the target L*. In sentence type 2, the target L is again the tone on the nuclear-accented syllable in *noodle*, but here the pitch accent is not a simple L*, but a scooped bitonal rising pitch accent. Here the F0 minimum was even easier to pick out because it occurred after the fall from a prenuclear H* on the *lean*. In sentence type 3 also the target L is part of a bitonal rising pitch accent, but this time it is not the L but the following H that is associated to the accented syllable in *noodle*, in a common "contrastive emphasis" pattern. The minimum F0 value that we measured for the target L tone is again after the fall from the H* peak on the preceding accented syllable *lean*, but it is somewhat earlier than in type 2, because now the rise into the following H tone puts the peak, rather than the valley, on the nuclear-accented syllable. For sentence type 4, the target is a L- phrase tone. That is, this sentence, unlike the others, is broken into two minor intonation phrases, and the L target marks the final boundary of the first phrase. The tone falls between two H* accent peaks, and thus is easy to pick out even though it is not associated with any particular syllable. In sentence type 5, as in types 1 through 3, the L is for an accent on *noodle*, but here it is not the nuclear accent, and it is a tone usually much higher in the pitch range. This H+L* accent is characterized by a fall onto the accented syllable. The following H tone for the nuclear accent on *beans* is then downstepped to nearly the same level, so that this contour does not show the pronounced dip in F0 characteristic of the other types. However, it was possible to consistently find a local minimum in the accented syllable at the end of the fall from the prenuclear H* on *lean*.

In addition to the F0 minimum, we also measured SH activity level for the target L tones. Given what we know about the relationship between strap muscles and pitch lowering in shorter utterances in other languages (e.g. Erickson, 1993), we would expect a high level of SH activity some time before the L tone, and when there is a preceding H tone (as in types other than sentence 1) we might expect to see a gradual increase in SH activity over this interval where F0 is falling from the preceding H tone into the target L. In the examples of types 2, 3, and 4 in Fig. 1, the SH does show high activity in this interval, but there is also a noticeable pulse-like pattern. (These pulses seem to correspond with the segmental gestures. Measurement of the timing of the SH peaks relative to releases of the consonants during this falling interval in these three sentence types shows there is a regular relation between the SH peaks and the releases.) Because of this pulse-like behavior, we could not identify any single peak for each target low tone to measure peak SH activity level. Instead, we averaged over all SH activity in some relevant interval prior to the F0 minimum by integrating over the area under the SH curve during the interval and dividing by the length of the interval. The cursors in Fig. 1 demarcate this SH interval. The criteria for choosing demarcation points necessarily differed somewhat for different sentence types. For types 2, 3 and 4, where the target L is surrounded by H tones, the interval began at the onset of the first SH peak after the beginning of the rise into the preceding H* tone, and it ended at the offset of the last SH peak before the rise into the following H tone. For sentence types 1, the SH curve typically showed a series of two or three closely spaced peaks, reminiscent of the peaks in the interval for types 2, 3, and 4, and we used the onset and offset of these peaks to define the SH interval. For sentence type 5, by contrast, there were no particularly striking SH peaks associated with the target L tone, and we could rely only on the F0 pattern. The beginning of the interval was the end of the F0 rise into the preceding H* accent, and the end was the time point where we measured the F0 value for the target L.

3. Results and Discussion

Figure 2 shows the mean results for the different L tones in each of the three speaking ranges — low, normal, and high. Looking first at the mean F0 values in the left-hand panel, we see that the L types differ; at each of the three pitch ranges, the L* single-tone accent and L* of the bitonal L*+H accent have the lowest F0, then the L- phrasal tone and the leading L in the L+H* accent, which in turn are lower than the L* of the H+L* pitch accent. The different overall pitch ranges themselves also differ, reproducing Pierrehumbert's (1989) findings; for all five tone types, the F0 value decreases in going from normal voice to the soft voice, and increases in going from the normal voice to the loud voice.

Focusing next on the mean results for the average SH activity in the figure's right-hand panel, we see that the SH activity reflects the paradigmatic differences among the five tone types; within each overall pitch range, the mean value varies, in inverse order to the ranking of the F0 minimum values among the different tones. However, mean SH activity does not show the corresponding difference among the three pitch ranges; values for the L tones in the low range are not greater than those for the same L tones in the normal range, and the values for the normal range in turn are not greater than those for high range. Instead, there is roughly the same amount of SH activity for each of the ranges, despite the clear differences in F0.

Fig. 3 shows the relationship between minimum F0 and average SH activity in more detail by plotting the trend over the tokens individually. It is a scatterplot, with three different y-axis ranges for the tokens produced at the different overall pitch ranges. The figures show a clear inverse relationship between SH and F0; the average SH is higher for tokens with lower F0, once we have shifted the axes appropriately to reflect the different F0 ranges at different overall voice effort levels. The shift in the y-axis ranges from group to group also brings out the exponential nature of this curve. That is, the lower the F0 goes, the more drastically the SH increases. We interpret this result 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 chose the shifting y-axis ranges in the figure to reflect this apparent baseline shift.

The source of this baseline shift is not clear, although there are several plausible explanations. We know that changes in vocal effort involve not only changes in overall F0, but also changes in overall subglottal pressure. For example, greater vocal effort surely produces greater subglottal pressure from the increased volume of air being ejected from the lungs. Also, greater vocal effort involves greater jaw opening into the vowel, as Schulman (1989) for instance has shown. Given these other concomitants of vocal effort, we might say that the greater-than-expected SH activity in the high voice range is because the jaw is lowering more for loud speech and the SH is involved in jaw opening gestures. Alternatively, we might say that the speaker uses the SH more 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 the lesser jaw opening into softer vowels, or an adjustment by the speaker to a decreased volume of airflow from the lungs.

In summary, these preliminary results for one speaker show that there are paradigmatic differences among L tones in English intonation contours which are comparable to the differences among H tones. They also show that the SH is involved in producing these differences; the lower the tone, the greater the SH activity. Moreover, the L tones show a consistent pattern of F0 variation across the different levels of vocal effort; the F0 values of the tones in the normal range are higher than those for the tones in the low range, and are lower than for those in the high range. This variation in F0 level across overall pitch range is not reflected in the mean results for SH activity. However, the token-to-token relationship suggests a shift in baseline SH level, which could reflect concomitant jaw height differences, or active compensation for subglottal pressure differences associated with the different effort levels, or some combination of these two (or possibly other as yet unknown mechanisms). Further experiments are underway to tease out at least these two potential explanations by measuring SH again, along with subglottal pressure and jaw movement.

Acknowledgments

The EMG recordings were done in the ATR Human Information Processing Research Laboratories, Kyoto, Japan. In addition to the support of ATR, the work was supported in part by the National Science Foundation under grant no. IRI-8858109 to Mary Beckman.

References

Atkinson, J. E. (1978) Correlation analysis of the physiological factors controlling fundamental voice frequency. *Journal of the Acoustical Society of America*, **63**, 211-222.

van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992) Downstep in Dutch: implications for a model. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (G.J. Docherty & D.R. Ladd, editors), pp. 335-359. Cambridge: Cambridge University Press.

Boyce, S. & Menn, L. (1979) Peaks vary, endpoints don't: implications for linguistic theory. In *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistic Society.*

Bruce, G. (1982) Developing the Swedish intonation model. *Working Papers, Department of Linguistics, University of Lund*, **22**: 51-116.

Erickson, D. (1993) Laryngeal muscle activity in connection with Thai tones. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, **27**.

Gårding, E., Fujimura, O., and Hirose, H. (1970) Laryngeal control of Swedish word tones. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, **4**, 45-54.

Hallé, P.A., Niimi, S., Imaizumi, S. & Hirose, H. (1990) Modern standard Chinese 4 tones: EMG and Acoustic patterns revisited. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, **24**, 41-58.

Honda, K., Hirai, H., & Kusakawa, N. (1993) Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, **27**.

Kiritani, S., Hirose, H., Maekawa, K., Sato, T. (1992) Electromyographic studies on the production of pitch contour in accentless dialects in Japanese. *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, pp. 783-785.

Kori, S., Sugito, M., Hirose, H., Niimi, S. (1990) Participation of the sternohyoid muscle in pitch lowering: Evidence from Osaka Japanese. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, **24**, 65-75.

Liberman, M.Y., & Pierrehumbert, J.B. (1984) Intonational invariance under changes in pitch range and length. In *Language Sound Structure: Studies in Phonology Presented to Morris Halle* (M. Aronoff & R.T. Oerhler, editors), pp. 157-233. Cambridge, MA: MIT Press.

Pierrehumbert, J.B. (1989) A preliminary study of the consequences of intonation for the voice source. *STL-QPSR*, **4**, 23-36.

Shih, C.-L. (1988) Tone and intonation in Mandarin, *Working Papers of the Cornell Phonetics Laboratory*, **3**, 83-109.

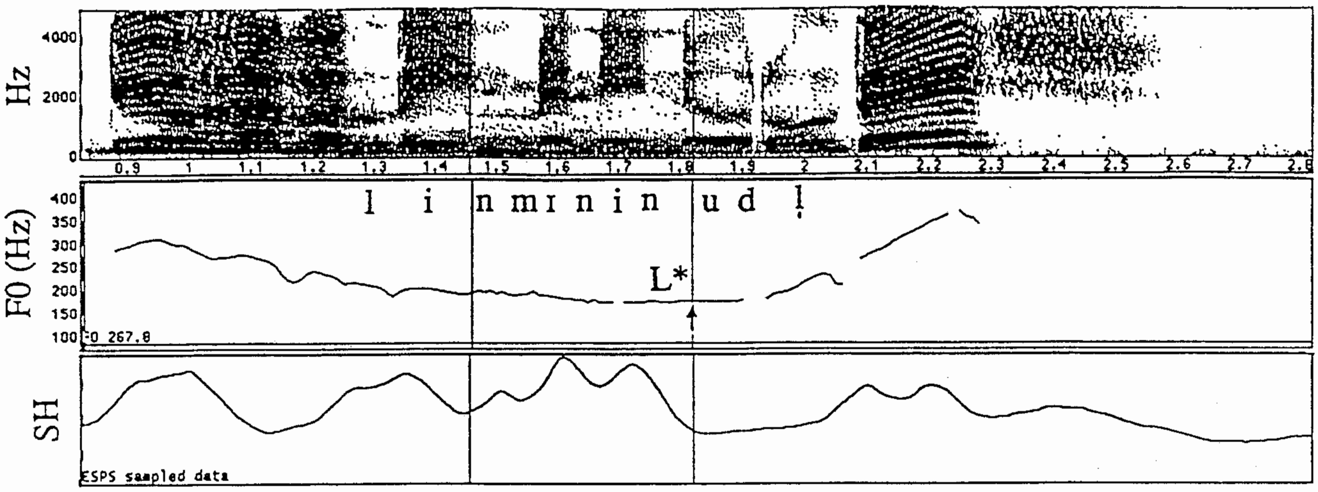
Schulman, R. (1989) Articulatory dynamics of loud and normal speech. *Journal of the Acoustic Society of America*, **85**, 295-312..

Simada, Z.B., Niimi, S., Hirose, H. (1991) On the timing of the sternohyoid muscle activity associated with accent in the Kinki dialect. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* , **25**, 39-45.

Yoshida, Y., Honda, K., Kakita, Y. (1993) Noninvasive EMG measurement of laryngeal muscles and physiological mechanisms of prosody control. *ATR Technical Report, TR-A-0168.*

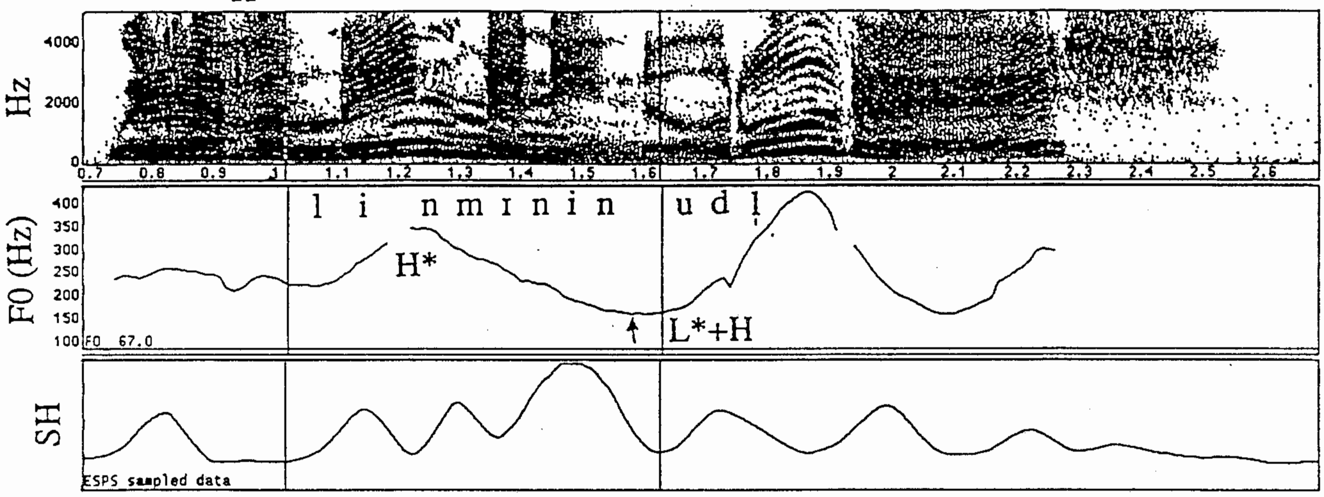
1. What would you like for lunch?
   D'you have a lean mini-noodle dish?

L* L* H- H%



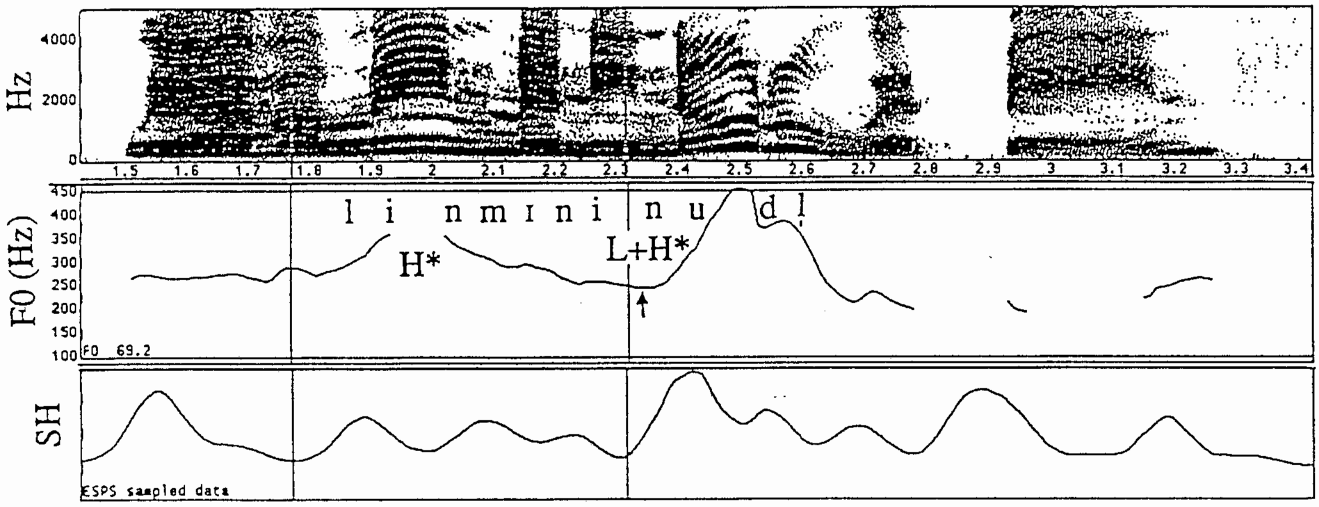
2. Do all of your rice dishes have this fatty meat sauce?
   We have a lean mini-noodle dish.

H* L*+H L- H%

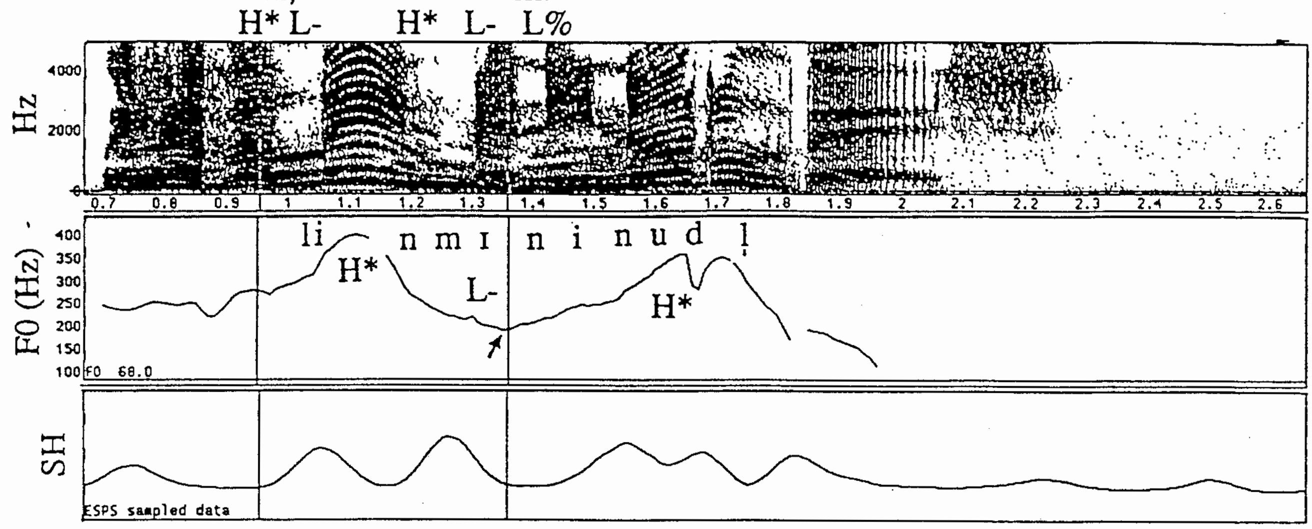


3. Do you have any bean dishes other than this couscous thing?
   We have a lean mini-noodle with beans.

H* L+H* L-L%



4.Do you have any pasta less fattening than fettuccine Alfredo?
We have a lean, mini-noodle dish.



5. Edward, you know we're not supposed to eat meat at lunchtime.
Oh, all right. We'll have the lean mini-noodle with beans.

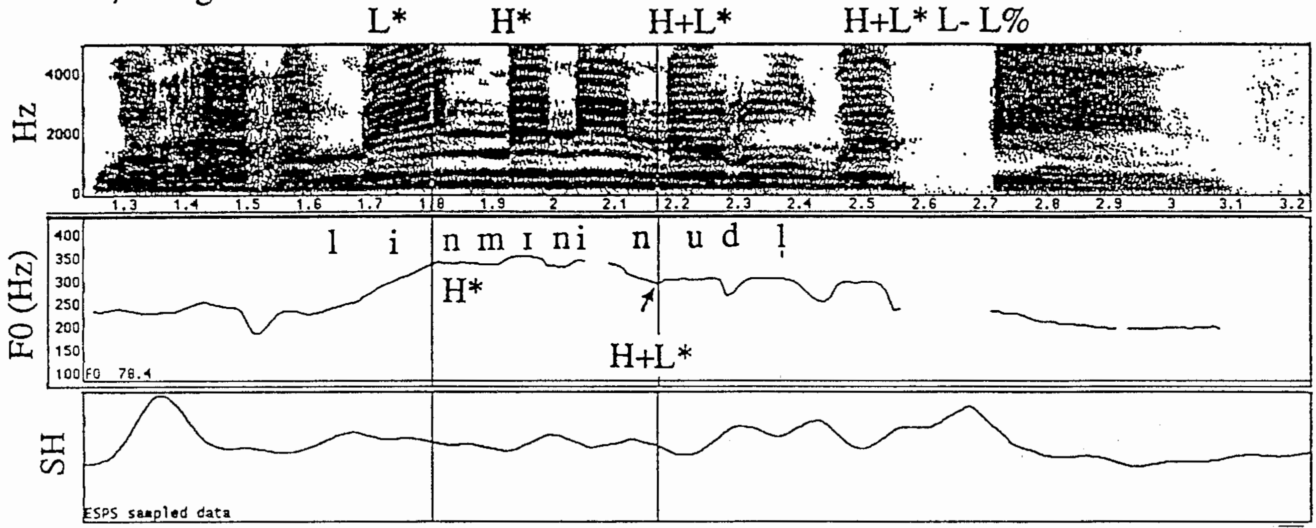


Figure 1. Spectrogram, F0 contour, and SH activity for representative tokens of the five sentence types. The cursors demarcate the target interval for averaging the SH activity, and the arrow points to the minimum F0 value measured for the tone.

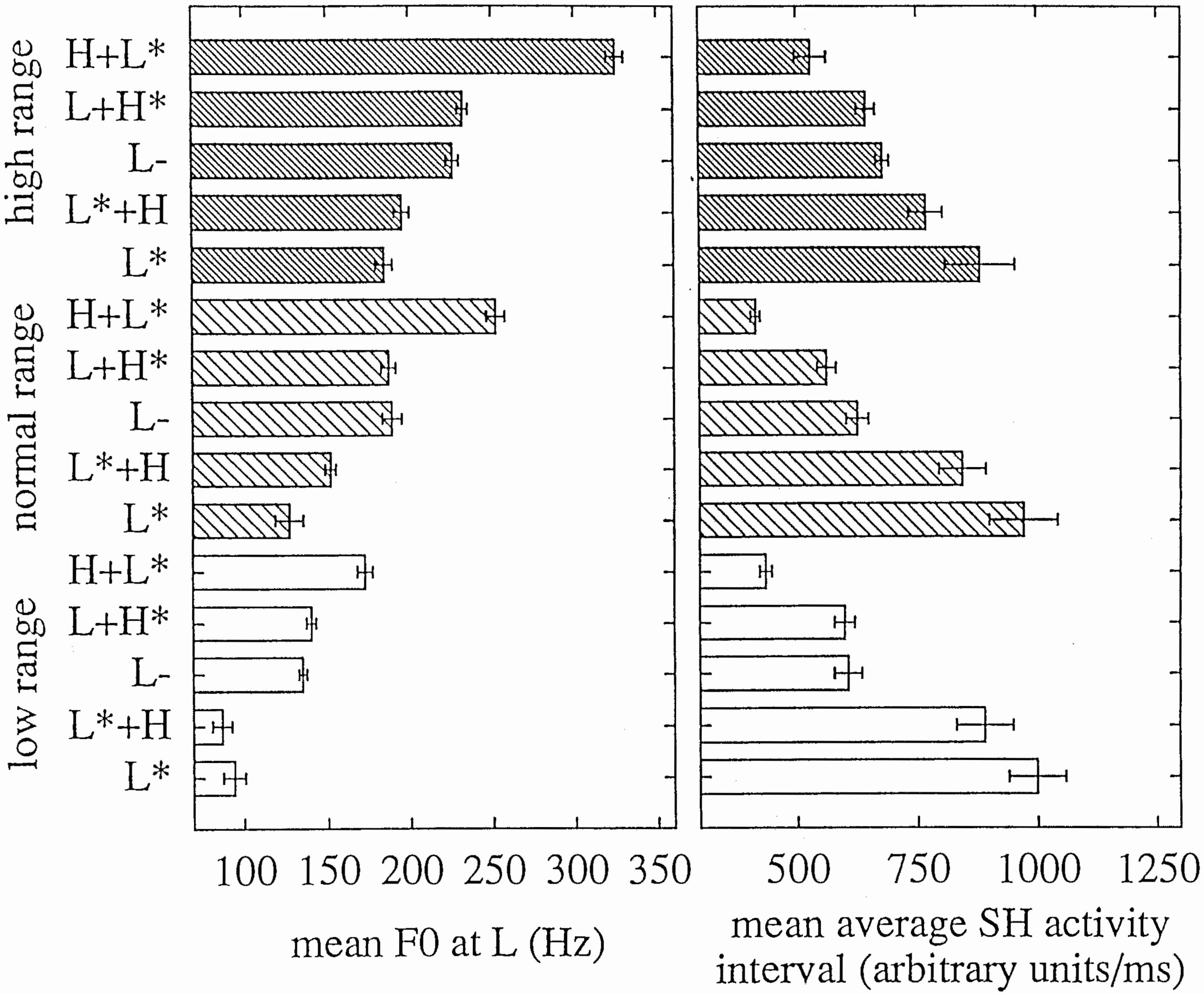


Figure 2. Mean values for F0 minimum and for average SH, pooled over L type and voice effort level. Types within a given voice effort level are arranged in order of increasing mean SH activity, which was consistently 5 > 3 > 4 > 2 > 1.

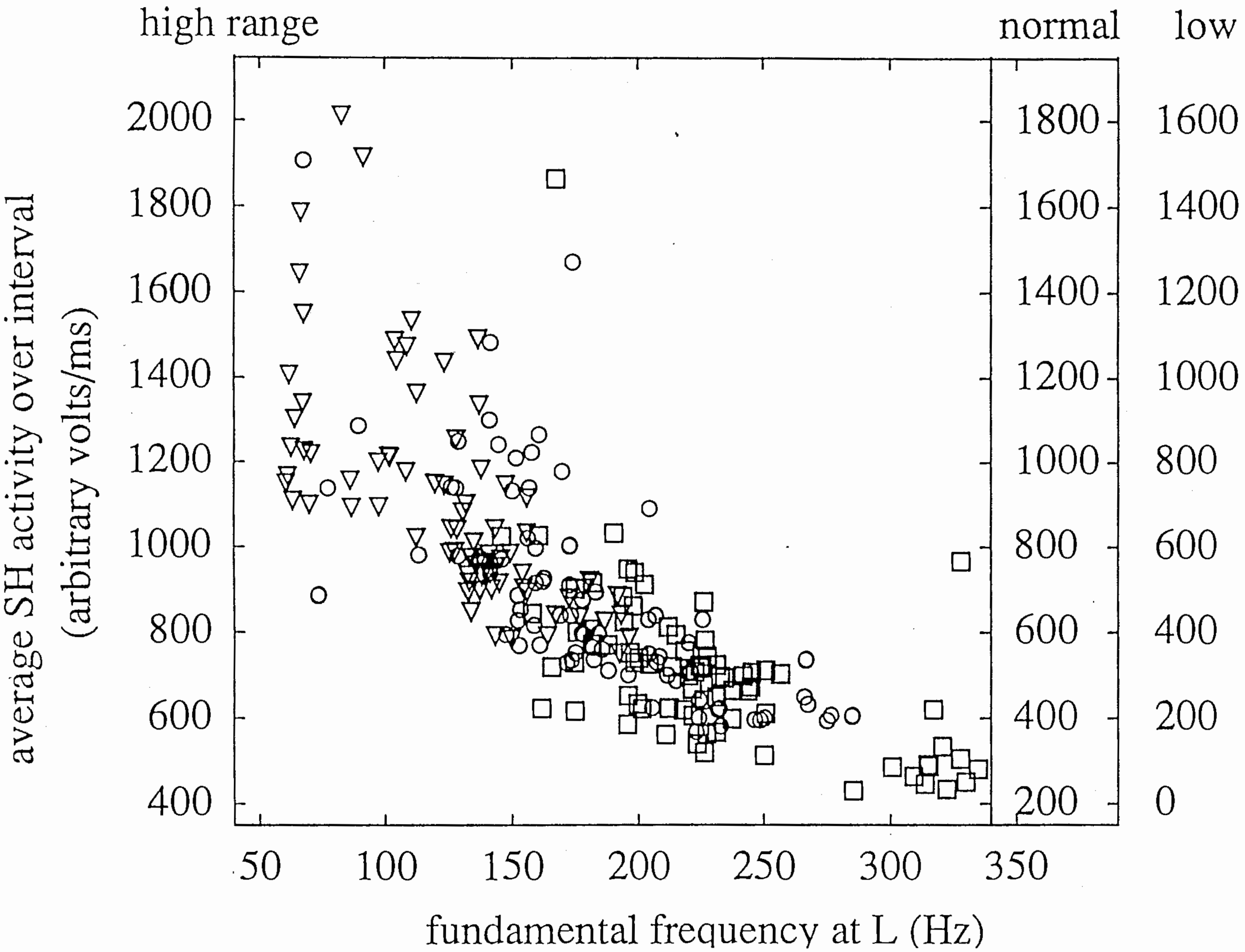


Figure 3. Average SH activity as a function of minimum F0 in all tokens of all sentence types at all levels of voice effort.