DOWNDRIFT AND PITCH RANGE IN CHICHEWA INTONATION

Scott Myers
University of Texas at Austin, USA

ABSTRACT

In many languages, questions have a higher pitch range and less steep downdrift than in statements. It is also known that the pitch range is higher in louder speech. In this study, the effects of these two factors on f0 were studied in Chichewa, a Bantu tone language of Malawi. Three speakers produced three sentences both as statements and as questions, at three levels of loudness. F0 peaks corresponding to four high tones per sentence were measured. The general trends above for questions vs. statements and louder vs. softer speech were replicated in these data. There was also a significant interaction between the two factors. Multiple regression models were used to predict the values of later f0 peaks on the basis of the first f0 peak, position, and sentence type. Notably, the best models were based on a linear decline in f0, augmented by a boost for the phrase-initial peak.

1. DOWNDRIFT AND PITCH RANGE

Questions are often distinguished from statements in languages by a phrase-final f0 excursion - generally upward in questions and downward in statements [1]. But they are also often distinguished in terms of the f0 realization of nonfinal tones. For example, downdrift (the tendency for each successive high tone in a phrase to be realized at a lower f0 value than the previous one) is significantly reduced in questions in a number of languages [2, 3, 4]. Questions are also distinguished from statements by having a higher pitch range [2, 4].

In a study of f0 for six speakers of Chichewa [5], a Bantu language of Malawi, questions were found to be distinguished from statements in all these ways. Questions generally had a sharp rise in f0 on the final syllable, while statements generally had a sharp fall. Downdrift was significantly less in questions than in statements, and the pitch range in questions was higher. Quantitative models were provided that predicted the f0 values of noninitial high tones on the basis of the f0 value of the initial high tone, and whether the utterance was a statement or a question.

However, in that study, speakers all selected their own loudness level. This resulted in a rather small range of moderate f0 values, thus raising the question of whether such quantitative models would work for a broader sample of an individual's full range of f0 values. Moreover, Liberman and Pierrehumbert have shown that in English louder speech is produced at a higher pitch range than quieter speech [6]. It is not known how the effect on pitch range of the question vs. statement distinction interacts with the effect of varying loudness.

A new study of intonation in Chichewa was undertaken, in which loudness was systematically varied in both questions and statements. The interaction of loudness with sentence type (question or statement) was investigated. Quantitative models of the results were produced.

2. METHODS

Three sentences of Chichewa were used, each with four lexical high tones, differing just in the length of the string between the two medial peaks.

(1) a. Mlónda ámayámba kunyénya.
watchman begins goof-off
"The watchman begins to goof off."

b. Mlónda ámayenéra kunyénya.
watchman must goof-off
"The watchman must goof off."

c. Mlónda ámalepherétsa kunyénya.
watchman prevents goof-off
"The watchman prevents (someone) from goofing off."

The sentences were printed on sheets of paper together with instructions as to (a) sentence type (statement or question), and (b) loudness (soft, normal, or loud). Speakers were instructed that the *loud* condition should be as if shouting to a person in another room, *normal* as if speaking to a person across the table, and *soft* as if speaking as quietly as possible (without whispering) to a person right next to them.

Three native speakers of Chichewa produced the sentences: SM (a male from Nkhotakota in Central Malawi), DJ (a male from Blantyre in Southern Malawi), and CJ (a female from Mzuzu in Northern Malawi). Each speaker produced 20 repetitions of each sentence in each sentence type and loudness, yielding 120 tokens of each sentence for each speaker (20 repetitions * 3 loudness conditions * 2 sentence types). Thus each speaker produced a total of 360 utterances for this study. The utterances were recorded on a Sony DAT tape recorder in a sound-treated booth.

The recordings were redigitized and f0 tracks produced using PLIB ("Phonology Laboratory in a Box"), a PC-based software package designed by Anthony Woodbury. Measurements were made of the local f0 peaks corresponding to each of the four high tones in each sentence. Values for the local f0 lowpoints between the peaks were not included in the study, since it was found that these values were transitional values dependent on the time elapsed between peaks [7]. There are thus no lexical low tone targets in Chichewa, just trough transitions between high tone targets, as with English H* accents [8].

Due to recording problems, three utterances had to be excluded from SM's database, leaving 357 tokens. In the loud condition, f0 for the female speaker CJ frequently went above

500 Hz, which is beyond the reliable pitch tracking threshold for PLIB. The entire Loud condition was excluded from her dataset, leaving only the Normal and Soft conditions (240 tokens). Speaker DJ was represented by the complete set of 360 tokens.

3. RESULTS

An overview of the results can be seen in Figures 1-3, which show the mean values for each high tone position in each of the six conditions of the experiment. Questions are indicated by solid lines, and statements by dashed lines.

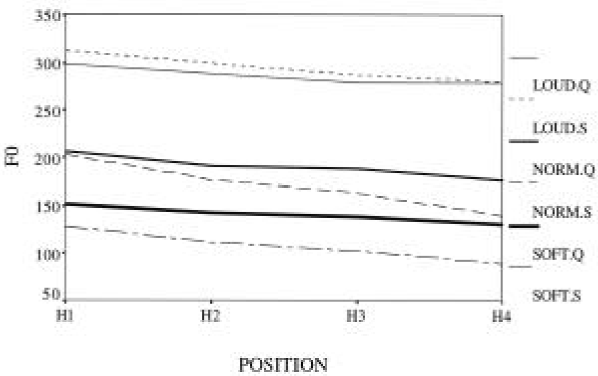

Fig. 1. Mean f0 valuesby position, Speaker SM.

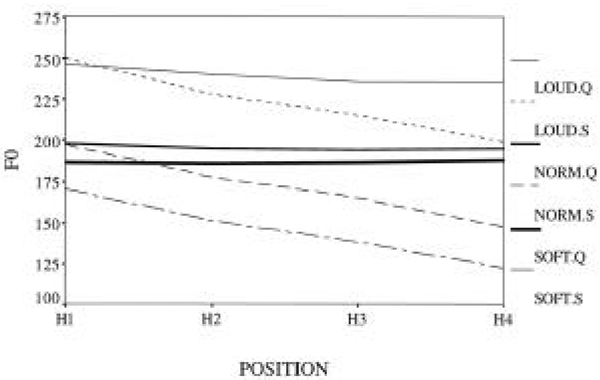

Fig. 2. Mean f0 values by position, Speaker DJ.

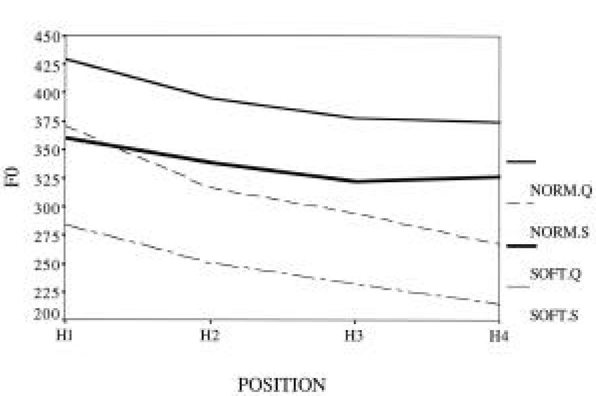

Fig. 3. Mean f0 values by position, Speaker CJ.

Several trends are evident in these graphs. First, the solid lines representing questions slope downward less than the dashed lines marking statements, indicating less downdrift in questions than in statements (cf. [5]). Second, the mean values are higher for louder speech than for softer speech, as found by Liberman and Pierrehumbert in English [6]. Third, the means are consistently higher in questions than in statements of the same loudness level.

One quantitative measure of the difference in downdrift between questions and statements is the quotient of the f0 value of a noninitial high tone H_n divided by the f0 value of the preceding high tone H_{n-1}. A value close to one would indicate that successive high tones were roughly even in f0 value, and the farther the quotient is below one, the greater the downdrift between successive tones.

For all three speakers, H_n/H_{n-1} was significantly higher in questions than in statements. The means are given in Table 1.

Speaker	Question	Statement	t-test
a. SM	.96	.91	t (1012.50) = -13.1, p < .001
b. DJ	.99	.91	t (954.9) = -35.8, p < .001
c. CJ	.96	.91	t (684.8) = -16.4, p < .001

Table 1: Mean H_n / H_{n-1} by sentence type.

The last column gives the results of t-tests for independent samples of uneven variance, showing that the difference in downdrift between question and statement is significant ($p < .05$) for all three speakers.

One measure of the effect of sentence type and loudness on the f0 level of high tones would be f0 levels of all high tones in the various conditions. But since downdrift is greater in statements than in questions, this effect alone would, all else being equal, lead to lower mean f0 values in statements. To factor out the effect of downdrift, then, only values for the first H in the sentence (H1) were compared. For each speaker, a 3 * 2 ANOVA was performed, with the f0 value of H1 as the dependent variable, and the factors Loudness (Soft, Normal, Loud) and Sentence Type (Question, Statement).

Each speaker showed a strong main effect of loudness, the mean f0 level being higher for louder speech (SM: $F(2, 356) = 1219.5$, $p < .001$; DJ: $F(2, 359) = 531.2$, $p < .001$; CJ: $F(1, 239) = 294.7$, $p < .001$). Two of the three speakers also had a significant main effect of sentence type, with the mean f0 level of H1 higher in questions (SM: ($F(1, 356) = 3.2$, $p = .08$, n.s.; DJ: $F(1, 359) = 6.0$, $p = .02$; CJ: $F(1, 239) = 227.6$, $p < .001$). The effect of sentence type was smaller than that of loudness, enough so that the former effect could get lost in the variation due to the second.

All three speakers had a significant interaction between these two factors (SM: $F(2, 356) = 16.7$, $p < .001$; DJ: $F(2, 359) = 13.2$, $p < .001$; CJ: $F(1, 239) = 3.9$, $p = .05$). For all three, the difference between statement and question was greater in soft speech than in the louder conditions. One interpretation of this effect would be that variation in softer speech is limited by the firm lower limit of a speaker's pitch range, while the upper limit is not approached except in the most extremely emphatic speech. Thus the greater variability of f0 values in the louder conditions overwhelmed the small effect due to the contrast between questions and statements.

4. MODELLING

It has been demonstrated that it is possible to model f0 contours with relatively simple models, predicting later f0 values from earlier ones [5, 6, 8, 9, 10]. Multiple regression techniques were used to derive such models for the Chichewa f0 data.

The dependent variable in these models was the f0 value of each noninitial H (H2-H4). These values were predicted on the basis of three independent variables: (a) *H1* (the f0 value of H1), (b) *type* (1 = statement, 0 = question), and (c) *n* (the position of the tone in the sentence: 1 = H2, 2 = H3, 3 = H4).

The first class of model considered had the form in (2):

(2) $H_n = (a * H1) + ((b + (c * type)) * n) + d$

According to this model, the downtrend from one H to the next is a constant decrement, which is greater for statements than for questions. The best fitting models of this for each of the three speakers are summarized in Tables 2-4.

Speaker	a	b	c	d	R^2
SM	.99	-5.58	-7.43	-4	.96
DJ	.90	0.08	-15.1	17.13	.93
CJ	.76	-4.62	-20.1	65.85	.94

Table 2: Additive models

These models define two downward-sloping straight lines as approximations of the f0 downtrend over the course of the phrase, with the line for statements sloping more steeply than the one for questions by the slope factor *c*. The equation for SM, for example, predicts a constant fall of 5.58 Hz per H tone in questions (= b) and 13.01 Hz per H tone in statements (= b + c). The R^2 values in the last column indicate that these models fit the data quite well, accounting for 93-96% of the variance in H_n.

According to these additive models, one contribution to the downtrend in f0 is the constant fall at each successive high tone, expressed by the coefficients *b* and *c*. But this linear fall is not from the height of H1, but from a proportion of H1 expressed by the coefficient *a*. Thus for speaker DJ, the linear decline starts from 90% of H1, while for speaker CJ it is 76%.

One simple interpretation of this would be that downdrift for our three Chichewa speakers is a simple linear decline, but that the first H1 is boosted above that line: by 1% for SM, 10% for DJ and 24% for CJ. As a result, the line from H1 to H2 in Figs. 1-3 tends to slope more sharply downward than those from H2 to H3 or H3 to H4. A boost for the first peak in a phrase has been found in English by Ladd [11] and in Japanese (right-branching phrases) by Kubuzono [12]. It can be seen as an instance of the same sort of phrase-initial fortition that has been found by Fougeron and Keating in segmental articulatory gestures [13, 14].

Liberman and Pierrehumbert [6] have demonstrated that downdrift in English takes the form of a decaying exponential, with the slope getting progressively less steep over the course of the phrase. Exponential models of the form in (3) were computed for the Chichewa data, and compared to the additive models above.

(3) $H_n = ((a + (b * type))^n * H1) + c$

According to this sort of model, the downtrend from one H to the next is not constant, but is larger initially and smaller later on in the phrase. The fits for these models are summarized in Table 3.

Speaker	a	b	c	R^2
SM	.99	-0.03	-13.8	.95
DJ	.996	-0.08	-3.06	.92
CJ	.97	-0.05	-12.6	.92

Table 3: Exponential models

These models define two downward-sloping curves. The equation for SM, for example, predicts that in a statement H2 will be 96% (= a + b) of H1, while H3 will be 92.2% (= $(a + b)^2$) and H4 will be 88.5% (= $(a + b)^3$). The drop in f0 thus decreases slightly as one proceeds across the sentence: the drop is 4% of H1 going from H1 to H2, then 3.8% from H2 to H3 and 3.7% from H3 to H4.

The R^2 values for the exponential models are slightly less than those for the additive models in Table 2. Moreover, because of the large number of datapoints that were fit, these small differences are significant, according to the t-test comparison between two model fits [15] (SM: $t_{1068} = 10.1$, $p < .01$; DJ: $t_{1077} = 6.0$, $p < .01$; CJ: $t_{717} = 5.4$, $p < .01$). The additive models in Table 2 have a significantly better fit than the exponential models in Table 3.

The residuals of the additive models were also examined in order to determine whether they varied systematically by position, as they would if the true pattern were curvilinear, or if there were a final lowering effect. For two of the speakers, there was no significant difference in the residuals among the three positions H2-H4 (SM: $F(2, 1070) = 1.16$, $p = .31$, n.s.; DJ: $F(2, 1079) = 0.80$, $p = .45$, n.s.). For the third speaker , CJ, the residuals did differ according to position ($F(2, 719) = 4.64$, $p = .01$). Post-hoc Scheffé tests showed that the residuals in H3 were significantly lower than those in H2 and H4, but H2 and H4 were not significantly different from each other. This would be compatible with a very slight curvature in the actual observed downtrends.

4. CONCLUSION

The following conclusions emerge from our study about the speech of these three Chichewa speakers.

(4) Downdrift is less in questions than in statements.

(5) The range of f0 values for high tones is higher in questions than in statements.

(6) The range of f0 values for high tones is higher in louder speech than in softer speech.

(7) The f0 values of noninitial H tones can be predicted with considerable accuracy on the basis of:
(a) the f0 value of the initial H tone,
(b) the position in the sentence of the predicted tone, and
(c) whether the tone is in a question or a statement.

(8) The best-fitting models of the downtrend in f0 over the course of a phrase assumed a linear decay of a constant amount per high tone, coupled with a boost of the initial H in the phrase.

ACKNOWLEDGEMENTS

The research reported here was supported by Grant # SBR-9514481 from the National Science Foundation. I thank Sam Mchombo, and Daniel and Chipo Jamu for patiently serving as subjects. I would also like to thank Dr. Mchombo and Dr. John Ohala for making available the recording facilities of the Phonology Laboratory of the Univeristy of California at Berkeley, Troi Carleton for assisting in the recording, and Sung-a Kim and Graham Horwood for helping with the pitch tracking.

REFERENCES

[1] Bolinger, D. 1978. Intonation across languages. In J. Greenberg (ed.), *Universals of Human Language, Volume 2: Phonology*. Stanford: Stanford University Press, 471-524.
[2] Hadding-Koch, K. and M. Studdert-Kennedy. 1964. An experimental study of some intonation contours. *Phonetica*, 11, 175-185.
[3] Thorsen, N. 1978. An acoustical investigation of Danish intonation. *Journal of Phonetics*, 6, 151-175.
[4] Inkelas, S. and W. Leben. 1990. Where phonology and phonetics intersect: The case of Hausa intonation. In J. Kingston and M. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press, 17-34.
[5] Myers, S. 1996. Boundary tones and the phonetic implementation of tone in Chichewa. *Studies in African Linguistics*, 25, 29-60.
[6] Liberman, M. and J. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle (eds.), *Language Sound Structure*. Cambridge: MIT Press.
[7] Myers, S. In press. Surface underpecitication of tone in Chichewa. *Phonology*, 15.
[8] Pierrehumbert, J. 1980. *The phonetics and phonology of English intonation*. Bloomington: Indiana University Linguistics Club.
[9] Pierrehumbert, J. and M. Beckman. 1988. *Japanese tone structure*. Cambridge: MIT Press.
[10] Prieto, P., C. Shih, and H. Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* , 24, 445-473.
[11] Ladd, D. R. 1988. Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, 84, 530-544.
[12] Kubozono, H. 1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* , 6, 39-68.
[13] Fougeron, C. and P. Keating. 1996. Variations in velic and lingual articulation depending on prosodic position: Results for 2 French speakers.*UCLA Working Papers in Phonetics*, 92, 88-96.
[14] Fougeron, C. and P. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* , 101, 3728-3740.
[15] Cohen, J. and P. Cohen. 1975. *Applied multiple regression / correlation analysis for the behavioral sciences*. New York: Lawrence Erlbaum.