Declination in Mandarin

Chilin Shih

Bell Labs – Lucent Technologies
700 Mountain Avenue, Murray Hill, NJ 07974, USA
{cls}@research.bell-labs.com

ABSTRACT

In a sequence of high level tones, downstep is by definition absent and the tonal contribution to F0 movement is minimal. The surface F0 contour therefore is a close approximation of the declination slope, or the declarative sentence intonation. This paper reports experimental results of Mandarin high tones, and concludes that there is a clear declination effect stronger in magnitude than previously reported, which is best modeled as an exponential decay. Futhermore, declination interacts with sentence length and focus.

1. INTRODUCTION

An observable F0 pattern is a combination of many factors. A major problem in intonation study is that there is no unique solution in decomposing the observed complex F0 pattern into individual effects. To study one effect one often needs to make strong assumptions about the others, which turns out to be a major cause of debates in the intonation literatures. This paper proposes an experimental design which uses Mandarin high level tones to study the declination effect. Some of the factors affecting intonation are naturally absent in these sentences, thus the experiment provides a unique opportunity to observe the declination effect more directly. Below is a brief summary of the major factors affecting intonation contours:

- Declination: A global downtrend referring to the tendency of F0 to decline over the course of an utterance [8, 9, 15, 1, 2, 5, 14].
- Downstep: A lowering effect that are triggered by a low (L) accent or a L tone, resulting in a step-like function in F0 contours [7, 10, 12, 11].
- Final Lowering: Additional lowering effect near the end of a declarative sentence [7, 10].
- Accents and tones: Accents and tones create local F0 excursions. Each accent and tone is realized with a prominence setting that reflects the meaning and the structure of the sentence, and the speaker's rendition of the sentence.
- Segmental effects: A lot of the observed F0 movements are caused by segmental effects. Voiceless fricative and aspirated stops raise F0, while sonorants typically lower F0. Low vowels have intrinsically lower F0 than high vowels [6, 13].
- Intonation type: Sentence intonation such as declarative, exclamation or question intonation may interact differently with any of the aforementioned effects and with accents and tones [15, 3]. Many models treat declination as the declarative intonation.

There are two usages of the term "declination", which arise from different ways of decomposing intonation contours. The broad definition of declination refers to all of the observed downtrend in a sentence [8], and the narrow definition of declination refers to the global, gradual downtrend, calculated as the residue downtrend when the downstep effect is factored out [7, 10]. Liberman and Pierrehumbert [7] made the strong claim that after downstep effect if factored out, there is no evidence of declination in English. This position is softened in [10] where both declination and downstep are incorporated in the modeling of Japanese intonation.

This paper explores some properties of declination as defined by the narrow definition using data from Mandarin. Mandarin offers a case where the declination effect can be observed more directly: First, in a sequence of high (H) level tones (tone 1), low (L) tonal targets are absent phonologically and phonetically, so by definition there will be no downstep effect. Secondly, the tone shape is level, which means that the contribution from the accent and the tone to the observed F0 movement is minimal. There are still a few factors at work, of which final lowering and segmental effects are relatively easy to control, and at least some of the local prominence effect can be averaged out or smoothed out. The observed downtrend on the surface will then be very close to the narrow definition of declination.

2. EXPERIMENT DESIGN

This experiment investigates the pattern of declination in Mandarin and the possible interaction of declination with sentence length, final lowering and prominence.

The experiment consists of 640 sentences: 10 test sentences, 2 focus conditions, 2 final conditions, 4 repetitions and 4 speakers (two females and two males, two from northern China and two from Taiwan). Each of the 10 test sentences starts with a two-syllable sentence frame in a low tone (tone 3) and a rising tone (tone 2) *Lao3 Wang2* "Old Wang". Tone 3 has a low target (L) and tone 2 rises from low to high (LH), so the frame provides a reference to the pitch range. Sequences of high-level tones (tone 1, H) ranging from 2 to 11 syllable long follow the frame, the sentences share the theme of *Lao3 Wang2* cooking a winter melon in various ways. Sonorant and unaspirated consonant are used as much as possible to minimize consonantal

Plain, final	Lao3-Wang2 zheng1 dong1-gua1
Plain, non-final	Lao3-Wang2 zheng1 dong1-gua1 le0
Focus, final	(Shei2 zheng1 gua1?) Lao3-Wang2 zheng1 dong1-gua1
Focus, non-final	(Shei2 zheng1 gua1?) Lao3-Wang2 zheng1 dong1-gua1 le0

Table 1: Test Sentences

effects on F0.

The test sentences have tone 1 at the utterance final position. A non-final condition is created by adding a sentence final perfective particle *le* to the test sentences so that the last tone 1 syllable is no longer utterance final. If there is a final lowering effect, the last tone 1 syllable in the sentences without *le* should be lower than the ones in the sentences with *le*.

These 20 sentences were presented in two ways to the speakers: unadorned, plain sentences intended to solicit unmarked reading style, and sentence with a leading question *Shei2 zheng1 gua1?* "Who is cooking the melon?", intended to solicit a narrow focus reading with the prominence landing on the frame *Lao3 Wang2*. The four test conditions are shown in Table 1 with one of the test sentence *Lao3-Wang2 zheng1 dong1-gua1* "Old Wang steams the winter melon".

Three of Speaker C's sentences were read wrong and were discarded, resulting in a total of 637 sentences and 4139 tone 1 syllables in the database. One F0/time measurement is taken from each syllable: the lowest point of tone 3, the highest point of tone 2, and 50 msec after the vowel onset for tone 1.

3. RESULTS AND DISCUSSIONS

We would like to find out what factors have an effect on the scaling of F0 values, and to apply the findings to intonation modeling. Preliminary inspection of the data suggests that the location of word boundary and the location of the verb have some effects, so these two factors were also coded in addition to the three factors in the experimental design. The five factors were used in the ANOVA analyses to model the difference in F0 value between adjacent tone 1 syllables: sentence length (10 levels, from 2 to 11 tone 1 syllables), focus (2 levels, with or without narrow focus), final (2 levels, with or without final *le*), word (2 levels, whether the two syllables in question straddle a word boundary), verb (3 levels, the second syllable is a verb, the first syllable is a verb, or neither is a verb). The results are given in Table 2.

Figure 1 plotted the averaged F0 trajectory of the four speakers, including the the two frame syllables at position 1 (tone 3, L) and position 2 (tone 2, H). Tone 1 syllables (H) occur from position 3 and on. The two final conditions are plotted separately: the final condition without *le* in solid lines and the non-final condition with *le* in dotted lines. The two populations matche closely within each speaker. Not surprisingly, the difference between the final

Speaker A, Female from Beijing

	Df	Sum of Sq	Mean Sq	F	Pr(F)
length	9	750.89	83.43	1.79	0.07
final	1	17.48	17.48	0.38	0.54
focus	1	10.62	10.62	0.23	0.63
word	1	113.96	113.96	2.45	0.12
verb	2	3633.28	1816.64	38.98	0.00
Resid.	865	40315.07	46.61		

Speaker B, Female from Taiwan

	Df	Sum of Sq	Mean Sq	F	Pr(F)
length	9	4408.15	489.79	4.83	0.00
final	1	489.67	489.67	4.83	0.03
focus	1	32564.18	32564.18	321.09	0.00
word	1	939.20	939.20	9.26	0.002
verb	2	4559.81	2279.91	22.48	0.00
Resid.	865	87727.46	101.42		

Speaker C, Male from Taiwan

	Df	Sum of Sq	Mean Sq	F	Pr(F)
length	9	908.18	100.91	1.96	0.04
final	1	55.25	55.25	1.07	0.30
focus	1	75.26	75.26	1.46	0.23
word	1	312.69	312.69	6.08	0.014
verb	2	9672.35	4836.18	94.06	0.00
Resid.	847	43551.47	51.42		

Speaker D, Male from Tianjin

	Df	Sum of Sq	Mean Sq	F	Pr(F)
leng	9	5654.57	628.29	9.35	0.00
final	1	62.57	62.57	0.93	0.34
focus	1	3299.44	3299.44	49.12	0.00
word	1	6932.94	6932.94	103.22	0.00
verb	2	4838.29	2419.15	36.02	0.00
Resid.	865	58098.01	67.17		

Table 2: Results of ANOVA analysis

and non-final conditions are not significant in the ANOVA analysis. Furthermore, when just the last tone 1 syllables are compared, the differences are still not significant at the $p < 0.001$ level for all four speakers. So in subsequent analysis these two conditions are collapsed.

There are a few interesting aspects shown in Figure 1: all four speakers show a clear decline in tone 1 values. The last values at position 13, although phonologically equivalent to those at position 3, lie either below or close to the mid points between the initial L and the highest H. The rate of decline is faster in the beginning and slows down as the sentence progresses. The pattern of declination is reminiscent of the downstep equation proposed in [7], where each successive peak is modeled as a fraction of the previous peak plus a constant. The pattern is incompatible with the time-constant declination in [10, 4] where declination is modeled as a fixed amount of F0 drop over the same duration of time. The magnitude of pitch decline we observe here is also larger than the 10Hz/second drop proposed in [10]. A linear least-squares model is used to estimate the F0 value of a given syllable P_i from the F0 value of the preceding

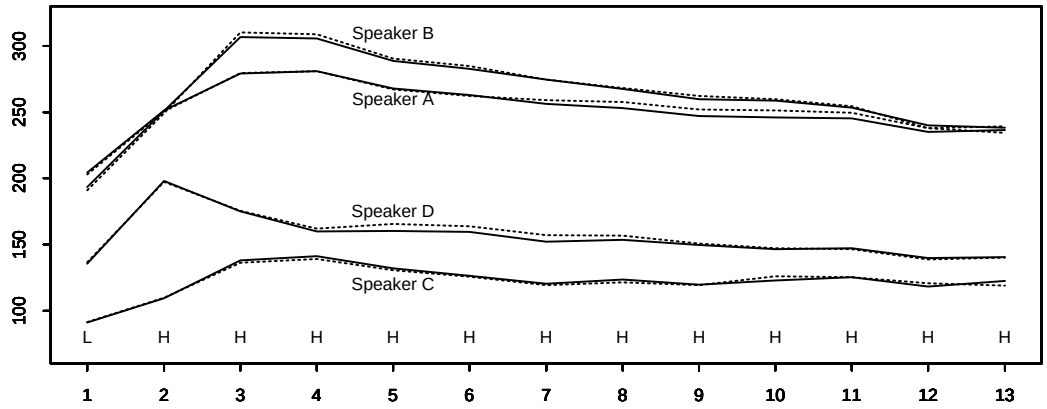

Figure 1: Averaged F0 trajectories of sentences with and without *le*

Speaker	μ	α	Correlation
A	23.52	0.89	0.82
B	39.36	0.83	0.84
C	24.10	0.79	0.56
D	25.94	0.80	0.63

Table 3: Summary of the Linear Least Squares Models

syllable P_{i-1}. The constant (intercept) μ and the coefficient α for each speaker are given in Table 3, together with the correlation of the model.

$$P_i = \mu + \alpha P_{i-1}$$

The models presented above can be improved if more factors are incorporated. Sentence length has a clear effect on the scaling of Tone 1's. Figure 2 plots the plain sentences of speaker A, sentences with different length are plotted separately. In shorter sentences the initial tone 1's are lower and the final tone 1's are higher, while in longer sentences the initial tone 1's are higher and the final tone 1's are lower. The low tone of the sentence initial tone 3 are slightly lower in short sentences, but the fluctuation is not as dramatic as in the high tone range.

All four speakers lengthened the frame syllables *Lao3 Wang2* in the focus condition, suggesting that all speakers indeed followed the cues and put emphasis on *Lao3 Wang2*, but only two speakers (B and D) raised the pitch range in addition to lengthening in their rendition of prominence.

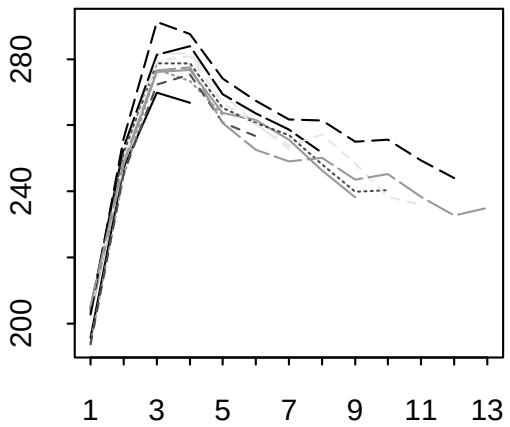

Figure 2: The effect of sentence length

Figure 3 plots the averaged F0 trajectory of the plain sentences (solid lines) and the sentences with focus (dotted lines) for three of the speakers. Speaker A's data is omitted to allow for clear viewing of speaker B's data. Speaker A's trajectories in this case are similar to the ones shown in Figure 1. Again, the two frame syllables are included in the plot.

Speakers A and C did not use F0 to differentiate focus vs. plain reading, and the two populations have nearly identical F0 trajectories. Speakers B and D both raised F0 to signal prominence. Speaker D raised F0 on the word with narrow focus, and the highest F0 value was reached by the end of the frame *Lao3 Wang2*, after that F0 started to decline and in two syllables it reached a level lower than the plain sentence and remained low throughout the sentence. Speaker B showed a similar pattern except that the highest point was on the first tone 1 syllable, one syllable *after* the narrow focus, and it was not until the 10th position, or eight syllables after the narrow focus, that the F0 trajectory crossed over the plain sentence. After the cross over the pitch value remained below the plain sentence until the end of the sentence. This observation is compatible with the claim that narrow focus is realized with expanded pitch range, and after the focus the pitch range is compressed [3]. However, one difference is in how soon the compression of pitch range occurs. Garding's claim is that the compression happens immediately after the narrow focus, which is true if there is a low tone following the focus [12], but in the case of high tones, the compression could occur quite late.

Although there is a significant effect of sentence length on the F0 value of the last syllable in plain sentence, as shown in Figure 2, the effect is weak in the focus condition. Figure 4 plots the focus sentences from speaker B. The ending F0 values in this plot is more similar than those in Figure 2.

Word boundary has a weak effect on F0 scaling but the

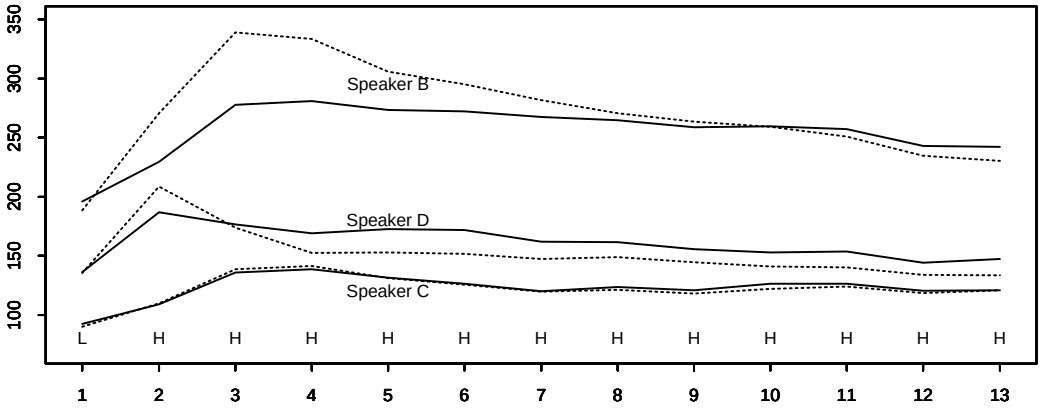

Figure 3: F0 trajectories of sentences with and without focus

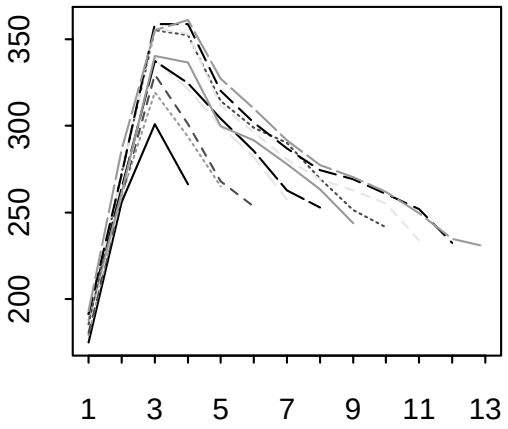

Figure 4: The effect of sentence length

effects are not consistent across speakers: two speakers show more F0 drop across word boundaries and two speakers show more F0 drop within words. The effect of verb is consistent across all four speakers: verbs have lower F0 values than their surrounding syllables, creating a shallow F0 valley. The most plausible explanation of this phenomenon is that verbs are metrically weaker than subject and object nouns in Mandarin. In F0 modeling this effect can be captured by assigning a low default prominence setting to the verbs which are followed by an object noun.

4. CONCLUSION

In this paper we show that Mandarin has a strong declination effect. The F0 decline is more pronounced near the beginning of the utterance and the effect can be modeled as an exponential decay. The slopes of declination are different for short and long sentences. Furthermore, declination interacts with focus. When pitch range is elevated the declination slope is steeper, the post-focus F0 trajectory eventually drops below the level of the controlled, plain population.

5. REFERENCES

1. A. Cohen, Collier R., and 't Hart J. Declination: Construct or intrinsic feature of speech pitch? *Phonetica*, 39:254–273, 1982.
2. Hiroya. Fujisaki. Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage, editor, *The Production of Speech*, pages 39–55. Springer-Verlag, 1983.
3. Eva Gårding. Speech act and tonal pattern in standard Chinese: Constancy and variation. *Phonetica*, 44:13–29, 1987.
4. Charlotte Gooskens and Vincent J. van Heuven. Declination in Dutch and Danish: Global versus local pitch movements in the perceptual characterisation of sentence types. In *Proceedings of ICPhS 95*, volume 2, pages 374–377, Stockholm, 1995.
5. D. R. Ladd. Declination: A review and some hypotheses. *Phonology Yearbook*, 1:53–74, 1984.
6. W. Lea. Segmental and suprasegmental influences on fundamental frequency contours. In L. Hyman, editor, *Consonant Types and Tones*, pages 15–70. University of Southern California, Los Angeles, 1973.
7. Mark Y. Liberman and Janet B. Pierrehumbert. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle, editors, *Language Sound Structure*, pages 157–233. M.I.T. Press, Cambridge, Massachusetts, 1984.
8. Philip Lieberman. *Intonation, Perception, and Language*. MIT Press, Cambridge, Massachusetts, 1967.
9. Janet Pierrehumbert. *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT, 1980.
10. Janet Pierrehumbert and Mary Beckman. *Japanese Tone Structure*. The MIT Press, Cambridge, Massachussetts, 1988.
11. Pilar Prieto, Chilin Shih, and Holly Nibert. Pitch downtrend in Spanish. *Journal of Phonetics*, 24(4):445–473, 1996.
12. Chilin Shih. Tone and intonation in Mandarin. In *Working Papers of the Cornell Phonetics Laboratory, Number 3: Stress, Tone and Intonation*, pages 83–109. Cornell University, 1988.
13. Kim E. Silverman. *The Structure and Processing of Fundamental Frequency Contours*. PhD thesis, University of Cambridge, 1987.
14. H. Strik and L. Boves. Downtrend in f0 and psb. *Journal of Phonetics*, 23:203–220, 1995.
15. N. Thorsen. A study of the perception of sentence intonation – evidence from Danish. *Journal of the Acoustical Society of America*, 67:1014–1030, 1980.