

Comparison of Declarative and Interrogative Intonation in Chinese

Jiahong Yuan, Chilin Shih*, Greg P. Kochanski*

Department of Linguistics, Cornell University

*Bell Labs, Lucent Technologies

jy55@cornell.edu

cls@research.bell-labs.com

gpk@research.bell-labs.com

Abstract

We model the differences between declarative and interrogative intonation in Chinese with Stem-ML, an intonation description language combined with an algorithm for translating tags into quantitative prosody. Our study shows that the diverse surface patterns can be accounted for by two consistent gestures: 1. Interrogative intonation has a higher phrase curve than declarative intonation; 2. Sentence final syllables have more careful intonation and wider pitch swings in interrogative sentences. Phrase curves of the two intonation types tend to be parallel and boundary tones are not necessary for modeling the differences between the two intonation types in Chinese.

1. Introduction

In English as well as many other non-tonal languages interrogative intonation has a rising end whereas declarative intonation has a falling end. This phenomenon has been widely studied and was generalized as the *Strong Universalist Hypothesis* [1], according to which pitch rising indicates a question and pitch falling indicates a statement.

In Chinese, however, the difference between declarative and interrogative intonation is much more complicated because of tone and intonation interaction. For example, interrogative intonation with a final rising tone has a rising end, which is similar to English, whereas that with a final falling tone often has a falling end (as shown in Figure 1).

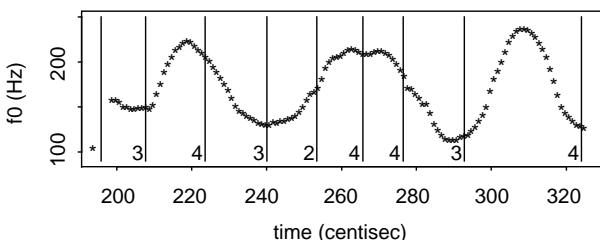


Figure 1: Interrogative intonation in Chinese can have a falling tail. Li3-bai4-wu3 luo2-yan4 yao4 mai3 lu4. "On Friday Luo-Yan

wants to buy a deer." Vertical lines mark syllable boundaries. Numbers indicate tones.

The difference between declarative and interrogative intonation has attracted much attention in Chinese intonation study. De Francis claims that the whole pitch level of the interrogative is higher than that of the declarative [2]. Disagreeing with De Francis, Tsao argues that the whole pitch level has no difference between the two intonation types and interrogative intonation in Chinese is 'a matter of stress' [3]. Gårding models Chinese intonation with 'grids', which qualitatively mark a time-varying pitch range. Lexical tones then fit into that range [4, 5]. In Gårding's model, the two intonation types have different grids. Shen J. proposes that the top line and the base line of a pitch contour are independent in the prosodic system of Chinese [6, 7]. For interrogative intonation the top line falls gradually whereas the base line undulates slightly and ends at a much higher point (compared to declarative intonation). Shen X. investigates the difference between the two intonation types by comparing their pitch values at four points: starting point, highest peak, lowest trough, and ending point [8]. Her conclusion is that interrogative intonation begins at a register higher than declarative, although it may end with either a high or low key.

The studies on Chinese intonation reviewed above draw conclusions from auditory impressions and/or instrumental analyses. In this paper we adopt a new methodology: building detailed mathematical models of F_0 by way of Stem-ML (Soft Template Mark-up Language) [9, 10]. Stem-ML consists of a set of tags that control intonation and an algorithm for generating F_0 curves from the tags. Tones are treated as soft templates, which can be modified due to the interaction between the neighboring tones, the interaction of tones with other components of prosody. The resulting intonation is an optimal compromise between physiological constraints (the F_0 curve must be continuous and smooth over short time scales) and communication constraints (the F_0 curve should match the intended shape). Previous studies on Chinese modeling with Stem-ML can be found in [11, 12].

We built different models as hypotheses to explain the differences between declarative and interrogative intonation. We trained the models with a least-squares fitting procedure and evaluated the models by how well they fit. The procedure and results follow.

2. Corpus

We designed a corpus of 8 pairs of sentences. The two sentences in each pair are identical except that one ends with a period, which indicates a declarative intonation, and the other with a question mark, indicating an interrogative intonation. The criteria for making sentences include: 1. The sentences are natural in both their meaning and tonal sequence. 2. Unvoiced fricatives and affricates are avoided, in order to decrease the segmental effects on F_0 curves. 3. Various tonal combinations and syntactic structures are covered. 4. No question word or structure is used. All the sentences are eight syllables long.

The 16 sentences were randomized and displayed one by one on a computer screen to the subject, a male speaker from China who speaks standard Mandarin and was not aware of the purpose of the study. The subject was asked to speak the sentences he saw on the screen in a sound proof recording room. The procedure was repeated three times, with a different randomization of the sentences at each time.

We assume that after being familiar with the sentences, the subject will speak them more naturally and pay more attention to the intonation of his utterance. Therefore, only the last repetition of the recording is used. We extracted F_0 curves of the sentences with ESPS/Waves [13] and then corrected F_0 calculation errors (0.5% of the data) by hand. Finally, the syllabic boundaries as well as the tone category of each syllable were manually labeled.

3. Procedure

The experiment consists of two steps. The first is to *study* differences between the two intonations, and the second is to *verify* the results from the first step. Both of the two steps follow the same procedure. First, we build a model by inserting adjustable parameters into Stem-ML tags. Second, we input the model into an automatic optimizer, which calculates the best value for each parameter by comparing the original F_0 curves and the curves generated by Stem-ML. Finally, we compare the best-fit parameters of the two intonation types.

4. Results (1): study differences

We built a model to study this difference. It makes four main assumptions: 1. All statements in the corpus share one phrase curve and all questions share another. Both of the phrase curves are straight lines defined by two

points, at the beginning and end of a sentence. 2. Words which have the same syntactic structure, occupy the same sentence position and are in the same intonation type have the same strength value. This causes 91% of the syllables to share strength values with at least one other syllable. 3. All tones, including those at the beginning and ends of utterances are generated from the same set of four lexically specified templates. 4. The templates expand (in pitch) as the strength increases. The model follows references [12,14] closely.

The best fit of this model is excellent, with only 9.4 Hz of RMS frequency difference between the original and the generated F_0 curves on the whole corpus. Figure 2 shows the fitting results for two pairs of sentences. The filled circles represent the natural F_0 and the solid lines represent the calculated F_0 .

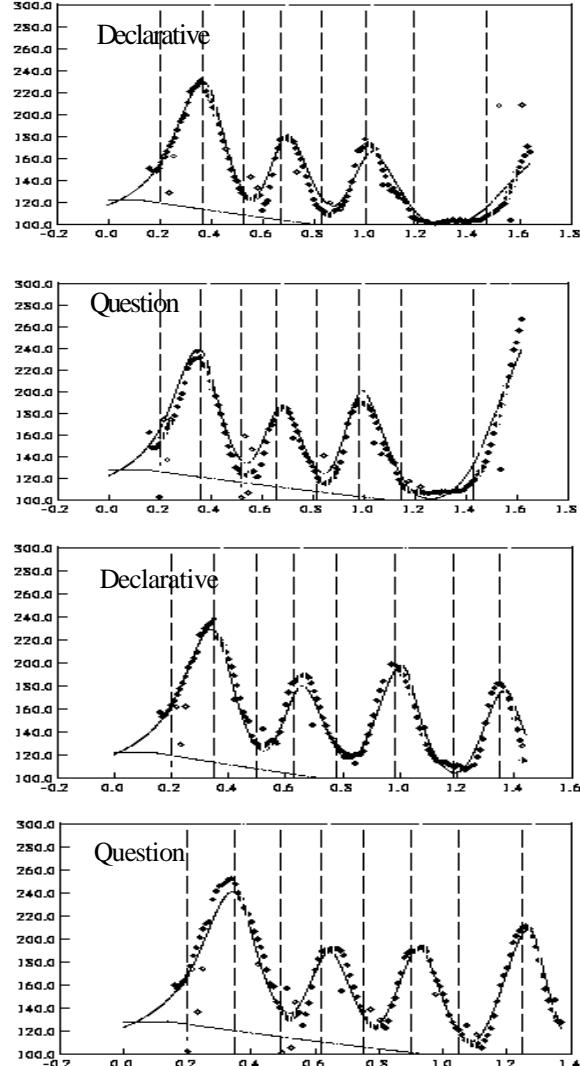


Figure 2: Natural (filled circles) and generated (solid line) intonation curves of two pairs (declarative vs. interrogative) of sentences. The upper pair ends in a rising tone: Luo2yan4 li3bai4wu3 yao4 mai3 yang2. "Luoyan wants to buy sheep on Friday." The lower pair ends in a falling tone: Luo2yan4 li3bai4wu3 mai4 ye3lu4. "Luoyan sells wild deer on Friday."

Dashed lines mark syllable centers. Frequencies are in Hz and time in seconds.

Figure 3 shows the best-fit phrase curves of the two intonation types. The solid line represents the phrase curve of the declarative intonation while the dashed line represents that of the interrogative intonation. We can see that the interrogative has an overall higher phrase curve than the declarative.

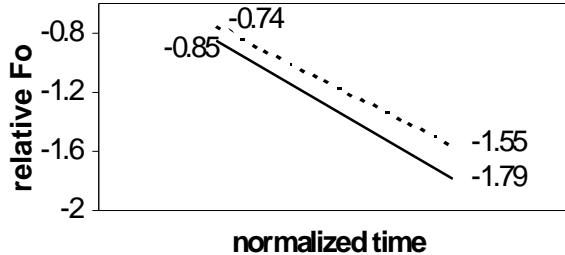


Figure 3: Phrase curves of the interrogative intonation (dashed line) and the declarative intonation (solid line). The interrogative intonation has higher phrase curve.

Figure 4 shows the differences of strength values between the interrogative sentences and the declarative sentences (mean and standard error of the mean over the eight pairs), plotted by syllable positions. It is shown clearly that the strengths on sentence final syllables are higher in the interrogative than in the declarative. The increased strengths at the end imply tighter adherence to the ideal tone shapes and larger pitch excursions.

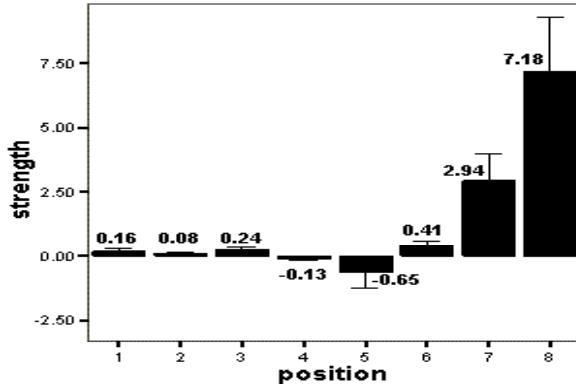


Figure 4: Differences of syllable strengths between the interrogative and the declarative intonation, plotted by sentence positions. The bars with the numbers indicate the mean of the differences and the error bars indicate the standard error of the mean.

5. Results (2): verify differences

The model in Section 4 assumes that phrase curves are straight lines. In Stem-ML the phrase curve is used to model the global trend of the fundamental frequency over the course of an utterance, similar to the concept of reference line or baseline/topline in literature. Reference line or baseline/topline has been treated either as a straight line [3, 4, 15], which is consistent with the

model in section 4, or as a curve which can have a falling or rising tail [5, 6, 7]. Considering alternative representations of phrase curves, we ask whether the results in Section 4 might change if we allow more flexibility in the phrase curve model, especially the one that the strengths on sentence final syllables are higher in the interrogative.

Another possible limitation of the model in Section 4 is the lack of a boundary tone [16]. First, boundary tone is widely accepted for many languages and has been used to describe Chinese intonation [17]. Second, it is reasonable to assume that the difference between interrogative and declarative intonation is that the former has a higher target at the end than the latter. One might hope that a higher boundary tone might replace the higher strengths in the final syllables of interrogative sentences.

We built three models to check the above two possibilities: one for phrase curve (model P) and two for boundary tone (models B1 and B2). Model P defines phrase curves by three points instead of two. It also assumes that the position of the middle point is sentence specific. Both model B1 and B2 assume that there is a boundary tone for declarative sentences and a different one for interrogative sentences. In model B1 the boundary tone is located at the very end of a sentence whereas in model B2 it is located at the center of the last syllable.

Stem-ML treats a boundary tone just like any other tone. They are placed at boundaries, and allow the model more freedom to match the data near the ends of utterances. The resulting intonation will be a compromise between the lexical tone shape and the shape of the boundary tone, weighted by their strengths. To the extent that the boundary tone can provide a more consistent representation of the data, the optimizer will use it, give it a larger strength, and reduce the strength of the lexical tones.

The results from model P are shown in Figure 5 and Table 1. Figure 5 shows the optimal phrase curves of the two intonation types for each pair. Table 1 lists the mean difference of strength values between the interrogative and the declarative in each syllable position.

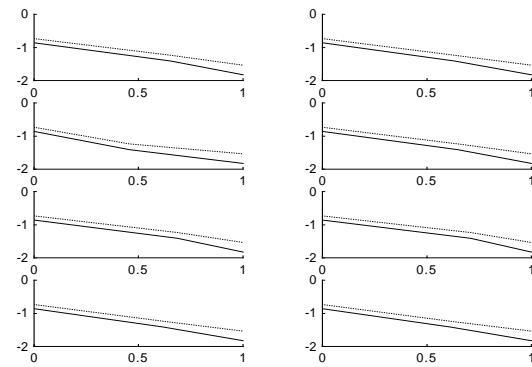


Figure 5: (Model P) Optimal phrase curves of the two intonation types of each pair. The x-axis represents normalized time and the y-axis represents relative F_0

Position	1	2	3	4	5	6	7	8
Mean strength diff.	.04	.04	.2	-.2	-.6	.5	3.0	7.2

Table 1: (Model P) Mean differences of strength.

Under model P, three points define the phrase curve, so a rising or falling tail can be realized, the strengths on sentence final syllables are still higher in the interrogative than in the declarative. Furthermore, Figure 5 suggests that the phrase curves of the two intonation types tend to be parallel, and not very different from a straight line.

Position	1	2	3	4	5	6	7	8
Mean Str. Diff (B1)	.3	.1	-.04	-.2	-1.1	.3	3.0	7.1
Mean Str. Diff (B2)	.3	.1	-.2	-.2	-1.0	.2	3.0	7.1

Table 2: (Model B1, B2) Mean differences of strength.

We can see from Table 2 that under model B1 and B2, in which the two intonation types can have different boundary tones, the strengths on sentence final syllables are still higher in the interrogative than in the declarative.

The difference of the optimal F_0 values of the boundary tones between the two intonation types is small: 4 Hz out of 300 Hz in model B1 and 18 Hz out of 300 Hz in Model B2. These results suggest that we do not need different boundary tones to account for the difference between declarative and interrogative intonation in Chinese.

6. Conclusion

Our study shows that the difference between declarative and the interrogative intonation in Chinese can be accounted for by two mechanisms: an overall higher phrase curve for the interrogative, and higher strength values of sentence final tones for the interrogative. This result is consistent with a perception study of question intonation [18], where listeners are more likely to interpret higher peak and higher ending pitch as questions, independent of their language background.

Our study also suggests that the phrase curves of the two intonation types tend to be parallel and boundary tones are not necessary for modeling the difference between the two intonation types in Chinese.

7. References

- [1] Ladd, D. R., 1981. On intonational universals. In Myers, T. et al. (Ed.) *The Cognitive Representation of Speech*, Amsterdam: North Holland Publishing.
- [2] DeFrancis, J. F., 1963. *Beginning Chinese*. New heaven: Yale University Press.
- [3] Tsao, W., 1967. Question in Chinese. *Journal of Chinese Language Teachers' Association*, 2, 15-26
- [4] Gårding, E., 1985. Constancy and variation in Standard Chinese tonal patterns. *Lund University Working Papers 28, linguistics-phonetics*, 19-51.
- [5] Gårding, E., 1987. Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica* 44, 13-29.
- [6] Shen, J., 1985. Beijinghua shengdiao de yinyu he yudiao[Pitch range of tone and intonation in Beijing dialect]. In *Beijing Yuyin Shixianlu*, Lin, T.; Wang L. (ed.). Beijing: Beijing University Press.
- [7] Shen, J., 1994. Hanyu yudiao gouzao he yudiao leixing[Intonation structure and intonation types of Chinese]. *Fangyan*, 3, 221-228.
- [8] Shen, X., 1989. *The Prosody of Mandarin Chinese*, University of California Press.
- [9] Kochanski, G. P.; Shih, C., in print. Soft templates for prosody mark-up. *Speech Communications*.
- [10] Shih, C; Kochanski, G. P; Fosler-Lussier, E.; Chan, M.; Yuan, J.. 2001. Implications of prosody modeling for prosody recognition. ISCA Workshop on Prosody in Speech Recognition and Understanding. Red Bank, NJ.
- [11] Shih, C.; Kochanski, G. P. 2000. Chinese tone modeling with Stem-ML. In *ICSLP*, Beijing, 2000.
- [12] Kochanski, G. P.; Shih, C.; Jing, H. 2001. Hierarchical structure and word strength prediction of Mandarin prosody. In *4th ISCA Tutorial and Research Workshop on Speech Synthesis*, Scotland.
- [13] Talkin, D.; Lin, D., 1996. Get_f0 online documentation. ESPS/Waves release 5.31. Entropic Research Laboratory.
- [14] Kochanski, G. P.; Shih, C.. 2001. Automatic modeling of Chinese intonation in continuous speech. *Eurospeech 2001*, pp. 911-914. Aalborg, Denmark.
- [15] 't Hart, J.; Rene Collier; A. Cohen, 1990. *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge University Press.
- [16] Pierrehumbert, J. B., 1980. *The Phonology and Phonetics of English Intonation*, Ph.D. dissertation, MIT.
- [17] Peng, S. et al., 2000. A Pan-Mandarin ToBI. <http://deall.ohio-state.edu/chan.9/MToBI.htm>
- [18] Gussenhoven, Carlos; Chen, Aoju, 2000. Universal and language-specific effects in the perception of question intonation. In *Proceedings of ICLSP 2000*, Beijing.