Final Lowering Effect in Questions and Statements of Chinese Mandarin Based on a Large-scale Natural Dialogue Corpus Analysis

Wei Lai [1, 2], *Ya Li* [2], *Hao Che* [2], *Shanfeng Liu* [2], *Jianhua Tao*[2], *Xiaoying Xu* [1, 2]

[1] School of Chinese Language and Literature, Beijing Normal University, Beijing, China
[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing, China

laiwei_0508@126.com, {yli,hche,sfliu,jhtao}@nlpr.ia.ac.cn, xuxiaoying2000@bnu.edu.cn

Abstract

To support text-to-speech with detailed prosody rules and to generate natural prosody, the paper studied the pitch variation near the end of sentences based on a Chinese Mandarin natural dialogue corpus. An additional lowering effect on the last prosodic word was found in both questions and statements, and proved to be independent of tone influence. Nevertheless, this effect, which is referred to as final lowering in other languages, was claimed to be absent in Chinese by some previous experimental studies. Such a contradiction is very likely to be caused by the difference between experimental speech versus natural speech. Based on this observation, the paper proposed a combination of the two methods in intonation studies, in which experimental speech served as an entry point to develop new topics, while natural speech served as a necessary extension to revise and apply prosody rules.

Index Terms: intonation, questions, final lowering, spontaneous speech

1. Introduction

In many languages, pitch within an utterance tends to drift down, especially in statements. There are many factors accounting for the pitch downtrend. One is the declination effect, which refers to the general tendency of pitch declination over the course of an utterance [4][6][9][12][14]. Another is final lowering, indicating an additional lowering effect near the end of the sentence [5][7]. Besides, there is also a downstep effect, which means that pitch lowering can be triggered by former syllables that carry low tones/accents [1]. Experimental studies of Mandarin showed the presence of declination and downstep [16][17][18], but the absence of final lowering [12]. In Chinese Mandarin, the downstep effect is supposed to be triggered by both low tone T3 and neutral tone T5 [12][16][17]. When it comes to questions, the downtrend is considered suppressed [14].

There have already been a considerable number of intonation studies on questions of Chinese Mandarin. Features of questions in Mandarin were put forward as: a) trends of top/base-lines, Shen, Jiong adjusted Gårding's grid model and suggested a gentle fall on the top-line and a slight rise on base-lines for questions [3][10]; b) starting point, Shen, Xiaonan suggested that all types of questions begin at a register higher than statements [11]; c) boundary tone, Lin, Maocan adopted AM theory and underlined the role of boundary tone in the distinction of interrogative and declarative moods [8]; d) phrase curve and strength, Yuan, Jiahong thought interrogation to be expressed by an overall higher phrase curve and higher strength on final syllables in Chinese [19].

However, the above conclusions cannot fully satisfy the needs of speech synthesis. For one thing, these studies mainly focused on the general intonation trend of questions [10][11][19], but neglected the concrete details of pitch variation within sentences. Is pitch variation distributed evenly over the whole sentence, or is it mainly realized by certain parts of the sentence? Such questions still remain disputable. For another, in order to generate natural prosody, conclusions from experimental works should be checked and revised in natural speech. Since Chinese Mandarin is a tone language, previous intonation research tends to adopt designed stimuli to arrange tone types [12][16][17][18][19][20], while research on spontaneous speech is relatively few. To make up for these inadequacies, we particularly studied the pitch variation near the end of questions and statements by using a large-scale natural dialogue corpus, with the purpose of providing speech synthesis with more detailed prosody rules, and thereby making contributions to generating natural prosody.

2. Corpus

Our research is based on a large-scale Q&A conversation corpus. The corpus contains more than 4000 sentences, i.e., more than 2000 turns of questions and answers, selected from interview programs. A wide range of topics are involved and sentences with less normal style were rewritten. These conversations were transliterated and then re-read by a trained male speaker in a professional recording studio to make sure that F0 values of different sentences come from the same speaker and are comparable. During the recording process, the speaker was required to maintain a natural speaking style without act or exaggeration. The corpus contains an annotation of four-level prosodic units, i.e., syllables, prosodic words, prosodic phrases and intonation phrases, which were manually checked by the first author.

All the questions in the corpus can be divided into 6 types according to syntax structure and pragmatic function. There are 1405 wh- questions, 184 v-neg-v questions (a kind of particular Chinese question syntax), 101 alternative questions, 382 yes-no questions, 114 tag questions, and 2179 statements. Yes-no questions can be further divided into 41 unmarked yes-no questions and 341 particle yes-no questions (including interrogative particles "吧"/ba/ and "吗"/ma/). Since the contrast focus of alternative questions and the tag utterance of tag questions will make the situation much more complicated, they are not discussed in this paper.

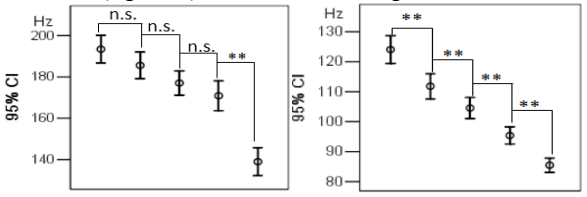
3. Experiments and Results

The present experiment was designed to study the F0 variations between prosodic words near the end of the sentences. Peaks and valleys of the last five prosodic words in each type of sentences were extracted. Sentences made of less than five prosodic words were removed from the samples. T5 in Chinese is a neutral tone, and the syllables of T5 are always

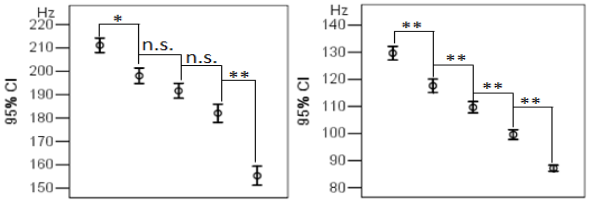
10.21437/SpeechProsody.2014-118

unstressed. Therefore, only sentences that end up with full lexical tones (T1-T4) were selected for wh- questions, v-neg-v questions, unmarked questions and statements.

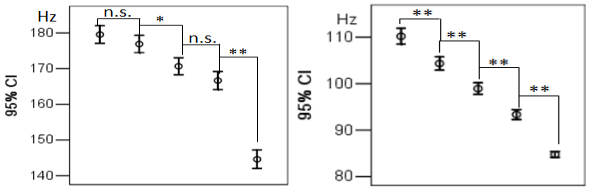
Error bars in Fig. 2 directly describe the pitch variations of the last five prosodic words in different types of sentences. Further ANOVA post-hoc test was done to compare pitch of adjacent prosodic words. Hereinafter, * stands for a significant difference (sig<.05); ** stands for an extremely significant difference (sig<0.01); n.s. stands for no significance.



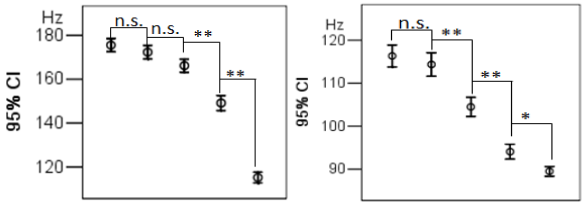
(a) The Topline (left) &Baseline (right) in V-neg-v Q



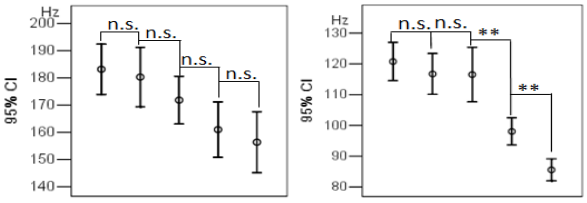
(b) The Topline (left) &Baseline (right) in Wh- Q



(c) The Topline (left) &Baseline (right) in Statement



(d) The Topline (left) &Baseline (right) in Particle Q



(e) The Topline (left) &Baseline (right) in Unmarked Q

Figure 2. *Pitch Variation of the Last Five Prosodic Words in Five Types of Sentences*

As shown in Fig. 2, pitch falls on top-lines are not evenly distributed over the whole sentence. An addition lowering effect appears at the end of all types of sentences except unmarked question, which is usually referred to as final lowering. Unlike top-lines, base-lines are encoded by a general pitch fall throughout the selected parts in most types of sentences. Therefore, the final lowering effect is supposed to be realized mainly by top-lines.

3.1. Top-lines: Presence of Final Lowering

Final lowering can be obviously detected in wh- questions, v-neg-v questions and statements. Pitch drops much more rapidly when it comes to the last prosodic word (around 30~40 Hz) than the former ones (around 10 Hz). Since sentences with neutral tone endings were removed from the sample, the above result was not influenced by the final unstressed syllables.

In particle questions, unstressed final particles were not removed and possibly made a contribution to the final pitch drop. However, a rapid fall also takes place between the last 2nd prosodic word and the last 3rd one. Since there is no evidence suggesting that neutral tone would affect its former syllables, the last second drop might be due to the effect of final lowering.

The presence of final lowering in v-neg-v questions and particle questions is supported by statistic results, for extremely significant differences are detected between the last (and last 2nd) pairs of prosodic words on top-lines. As for wh-questions and statements, there are also significant differences in other positions. Perhaps that is because pitch from natural speech varies more unstably due to the complication of tone. Although an extra pitch drop near the end of the sentence is observed in error bars of the two types of sentences, it has not yet been completely confirmed by statistical data. Thus, a supporting experiment is needed to eliminate tone differences and to narrow pitch gaps.

3.2. Base-lines: Evenly Distributed Pitch Falls

Fig. 2 shows two types of base-line encodings. One is a general decliantion throughout the selected part of sentences, which goes for wh-questions, v-neg-v questions and statements. The other is an additional final drop near the end of the sentence, which goes for yes-no questions. In unmarked yes-no questions, the drop takes place in the last two prosodic words. Coincidentally, in particle yes-no questions, if the last T5 particles were excluded, the extra drop also occurs on the last two prosodic words.

Statistical results on base-lines make a good fit with pitch variations in error bars. The difference between each pair of adjacent prosodic words is significant in wh- questions, v-neg-v questions and statements, suggesting a general smooth fall on base-lines. For unmarked questions and particle questions, the pitch differences stay insignificant except for the last two or three prosodic words. But in general, pitch falls on base-lines are distributed more evenly compared to top-lines.

3.3. Unmarked Q: Suppression of Final Lowering by Boundary Tone Effect

Final lowering is absent in unmarked questions, but the absence can be explained. Unlike other kinds of questions, unmarked questions do not depend on interrogative marks (such as wh- words, interrogative particles or syntactic means) to convey interrogative mood. Instead, boundary tone effect on the last syllable, which brings bring higher pitch as well as steeper pitch curse, is claimed to play an important part in the expression of interrogation [8]. According to our observations, final lowering in unmarked questions is partly set off by pitch rise caused by the boundary tone effect on the last syllable.

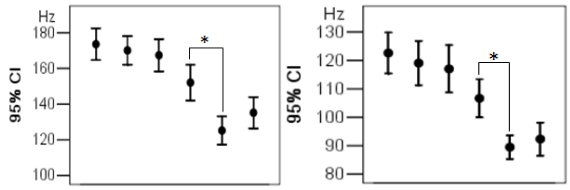


Figure 3: *Pitch Variation of the Last Six Syllables on Top-lines (left) and Base-lines (right) in Unmarked Q*

In Fig. 3, significant differences can be detected between the last 2nd &3rd syllables on both lines while in Fig. 2*(e)*, the differences becomes insignificant. This change is partly due to the pitch increase of the last syllable on top-lines. By contrast, the final pitch is not raised much on base-lines, so the significance is preserved. It is supposed that in unmarked questions, the pitch ascension of the last syllable on top-lines narrows the pitch gap between the last two prosodic words and suppresses the occurrence of final pitch drop to some extent.

3.4. The Basic Unit of Final Lowering: PW vs. SYL

In the light of Herman's data, the scope of final lowering is up to the last three syllables [5]. Also, we found that different syllables are influenced by different sentence types. According to Fig. 4, in wh- questions, pitch difference between the last 2nd&3rd syllables reaches its maximum; in v-neg-v questions, the biggest pitch falls take place between the last 1st&2nd and 3rd&4th syllables; in statements, the last three syllables are all influenced and share an even pitch fall. The conclusion contains much randomness and will cause extra troubles in the process of application.

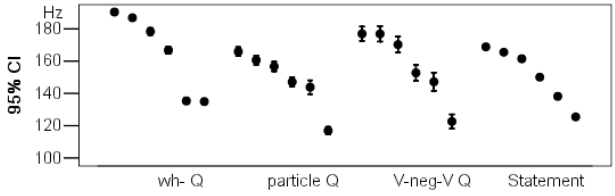


Figure 4. *Final Lowering Carried by Different Syllables on Top-lines of Different Sentences*

Once we changed the basic unit of final lowering from syllable into prosodic word, conclusions of higher consistency were obtained. As illustrated in Fig. 2, final lowering affects exclusively the last prosodic word in all sentences. This conclusion can reflect the prosody rules more precisely, and is more adaptable to intonation modeling and speech synthesis.

4. Supporting Experiment

The complication of tone combination in prosodic words brings about many troubles to our research. First, the pitch information of tone and intonation is hard to be decomposed. Second, many other tone-related effects will take place, such as downstep and neutral tone effect (also considered as a kind of downstep in some works) respectively triggered by low tone T3 and neutral tone T5. Besides, pitch difference between prosodic words could be enlarged. The above factors make the identification of final lowering more difficult.

To make up for such disadvantages, a supporting experiment was conducted to control tone effects. Pitch of the last six syllables was extracted and compared. Section 4.1 aims to discard tone difference by comparing syllables tone by tone. Section 4.2 is designed to evaluate whether our result is influenced by downstep and neutral tone effect. Since final lowering is mainly realized on top-lines according to Section 3, only maximal pitch of syllables is discussed in this section.

4.1. Elimination of Tone Difference

Tone-by-tone comparisons were done to neighboring syllables in order to rule out the differences of tone. Five pairs of syllables (last 1st&last 2nd~last 5th&last 6th) × 5 tones =25 times of comparison were done.

Table 1. *Syllable Pairs with Significant Pitch Difference*

Syllable Comparison	Last 6&5	Last 5&4	Last 4&3	Last 3&2	Last 2&1
Wh-	T2	T2	T1,T4	T1,T2 T4,T5	T1,T2 ,T4
Statement			T1,T4	T1,T2 T4, T5	T1,T3 T4,
V-neg-v				T2,T4	T3
Particle			T1,T3	T2	
Unmarked	n.s.				

According to Table 1, very few significant pitch falls are detected between the former syllables (last6&5, last5&4): only T2 syllables make a difference in wh- questions. Significant differences mainly occur between the last three pairs of syllables. Among them, the highest significance rate lies between the last second pairs, namely between the last 2nd&3rd syllables. To make things clearer, the rate of significant lowering between each pair of adjacent syllables is worked out.

$$Lowering\ Rate = \frac{Numbers\ of\ Significant\ Lowerings}{Five\ Times\ of\ Comparison} \quad (1)$$

Table 2. *Lowering Rate between Each Pair of Syllables*

Syllable Comparison	Last 6&5	Last 5&4	Last 4&3	Last 3&2	Last 2&1
Wh-	0.2	0.2	0.4	0.8	0.6
Statement	-	-	0.4	0.8	0.6
V-neg-v	-	-	-	0.4	0.2
Particle	-	-	0.4	0.2	-
Unmarked	-	-	-	-	-

According to Table 2, significant pitch lowering intensively occurs between the last three pairs of syllables, which approximately corresponds to the last prosodic words. In wh- quesitons, rapid pitch falls exist throught the whole sentence, but they are still most likely to appear between the last 2nd pairs of syllables, with a lowering rate of 0.8. The result proves that the occurrence of final lowering detected in this paper is independent of tone combinations.

4.2. Elimination of Downstep Effect

Since syllables of T3 and T5 mixed in the prosodic words may also result in pitch drops, in this section, sentences involving such syllables were excluded from the sample. Wh- questions and statements were chosen to be tested in this section because of their big quantity (1405 and 2179 respectively).

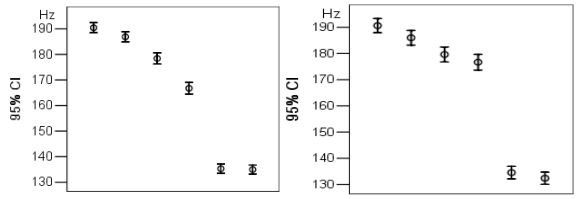


Figure 5. *Pitch Fall Original (left) and after the Exclusion of Sentences with T3 & T5 (right) in Wh- Q*

The left error bar in Fig. 5 shows that final lowering in wh-questions originally occurs between the last 2nd & 3rd syllables. Then, sentences with their last 3rd syllable carrying T3 or T5 are removed. Pitch variation of the qualified sentences was illustrated on the right. It turns out the only change is a pitch rise on the last third syllable, which is due to the delectation of low tones. Nothing happened to the following two syllables. In this case, the abrupt lowering of the last 2nd syllable is not caused by T3 or T5 of the last 3rd syllables in wh- questions.

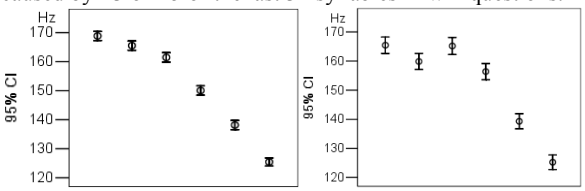


Figure 6. *Pitch Fall Original (left) and after the Exclusion of Sentences with T3 & T5 (right) in Statement*

In statements, significant pitch fall originally occurs at the last 3rd~1st syllables, instead of the last 2nd syllable in wh-questions. Accordingly, the exclusion of sentences was done with more syllables taken into account (the last 4th~2nd syllables). In consequence, a pitch rise appears on the last 4th~2nd syllables due to the exclusion of low tones. The extra pitch fall still occurs in the absence of T3 and T5. Based on the above analyses, we can deduce that our result is not affected by downstep effect triggered by low tones.

5. Discussion

The study brings about a disagreement on the presence vs. the absence of final lowering effect in Chinese Mandarin. Then, a more fundamental problem remaining to be solved is: what leaded to such a disputation? In previous research, designed sentences formed by syllables of the same high tone were used to get a direct observation of intonation, which indicated the absence of final lowering [12]. However, in our study of natural speech, when tone effects were controlled, an additional pitch lowering still appeared on the last prosodic words (or the last two to three syllables). By comparing the two methods, we noticed that the most obvious and substantial factor possibly leading to this disputation is the stimuli/corpus. In fact, the results suggested a potential difference between experimental speech vs. conversational speech, which is instructive to prosody research.

It is reasonable that different methods lead to different conclusions. Sometimes it is hard to label them as absolutely right or wrong, for both materials have their own pros and cons. On the one hand, it is true that designed stimuli can help us get minimal contrast pairs, and high-tone sentences do facilitate our observation of intonation. On the other hand, however, how likely are we to put together syllables of the same tone under natural conditions? Conversely, we tend to combine different tones in our speech intentionally or unintentionally to make it more melodious. That is why experiment stimuli are sometimes accused of being unnatural. Such problem can be fixed by natural corpora, for they contain a large quantity of sentences, extensive topics, diverse tones and flexible syntactic forms. Natural speech has its problems too. Since it involves the information of tone, intonation, stress, prosodic boundary, etc., extra efforts must be made to decompose all kinds of influents and to get what we want.

The experimental control has long been an indispensable part of phonetic research. It serves as an entry point in prosody research, owing to its convenience of simplifying complicated situations and developing new topics. However, we should not be complacent with the existing conclusions from experimental speech. In the next step, these conclusions should be extended into natural speech to get checked, revised or applied. Through this study, we propose a balanced consideration and a combination of the two methods. In our opinion, experimental speech is very essential for developing general topics into specific rules, and then the rules need an additional extension into natural speech for further check-up, revision and application. The combination of experimental control and natural corpus analysis will definitely bring new research topics, novel methods and promising conclusions

6. Conclusion

To support text-to-speech systems with plentiful prosodic details, we analyzed the pitch variation near the end of sentences in a Chinese natural dialogue corpus. A disagreement arises about the presence vs. the absence of final lowering in Chinese, and our results support the presence of final lowering in both questions and statements. Specifically, the effect occurs in wh- questions, v-neg-v questions, particle yes-no questions and statements, but is suppressed by boundary tone in unmarked yes-no questions. Besides, in the light of our data, final lowering is realized mainly on top-lines and affects exclusively the last prosodic word. At the end of the paper, we proposed a combination of experimental speech and natural speech in prosody research, with the former as an entry point of new topics and the latter as an extension of the existing prosodic rules.

In our future work, we will continue to explore the essence of final lowering in Chinese. More linguistic determinants such as accent, boundary tone, and position in the discourse are to be taken into account. Hopefully, the further research would bring out a full discussion on the physiological explication versus the phonological explanation.

7. Acknowledgements

The work was supported by the Beijing Normal University (11-30, 10-12-02), the Fundamental Research Funds for the Central University (2010105565004GK), the State Language Commission 12th Five-Year language research programme (YB125-41), the National Natural Science Foundation of China (NSFC) (No.61273288, No.61233009, No.61203258, No.61305003, No. 61332017, and No.61375027), and partly supported by the Major Program for the National Social Science Fund of China (13&ZD189) and the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM (CSIDM) Programme Office.

References

[1] Beckman, Mary, and Janet Pierrehumbert. "Intonational structure in Japanese and English." Phonology yearbook 3.1 (1986): 5-70.

[2] Cooper, William E., and John M. Sorensen. *Fundamental frequency in sentence production*. New York: Springer-Verlag, 1981.

[3] Gårding, Eva. "Speech act and tonal pattern in Standard Chinese: constancy and variation." *Phonetica* 44.1 (1987): 13-29.

[4] Cohen, Antonie, Rene Collier, and Johan t HART. "Declination: construct or intrinsic feature of speech pitch?" *Phonetica* 39.4-5 (1982): 254-273.

[5] Herman, Rebecca. "Final lowering in Kipare." *Phonology* 13 (1996): 171-196.

[6] Ladd, D. Robert. "Declination: a review and some hypotheses." *Phonology yearbook* 1 (1984): 53-74.

[7] Liberman, Mark, and Pierrehumbert, Janet. "Intonational invariance under changes in pitch range and length." *Language sound structure* 157 (1984): 233.

[8] Lin, Maocan. "Yiwen he chenshu yuqi yu bianjiediao." [Interrogative and Declarative Mood and Boundary Tone]. *Zhongguoyuwen* 4 (2006): 364-376.

[9] Pierrehumbert, Janet. "The perception of fundamental frequency declination." *The Journal of the Acoustical Society of America* 66 (1979): 363.

[10] Shen, Jiong. "Hanyu yudiao gouzao he yudiao leixing." [Intonation structure and intonation types of Chinese]. *Fangyan* 3 (1994): 221-228.

[11] Shen, Xiaonan. *The Prosody of Mandarin Chinese*. Vol. 118. University of California Pr, 1990.

[12] Shih, Chilin. "A declination model of Mandarin Chinese." *Intonation*. Springer Netherlands, 2000. 243-268.

[13] Thorsen, Nina. "Intonation and text in Standard Danish." *Annual Report Institute of Copenhagen* 18 (1984): 185-242.

[14] Umeda, Noriko. "F Declination is situation dependent." *The Journal of the Acoustical Society of America* 68 (1980): S70.

[15] Vaissière, Jacqueline. "Language-independent prosodic features." *Prosody: Models and measurements*. Springer Berlin Heidelberg, 1983. 53-66.

[16] Wang, P., et al. "Putonghua chenshuju zhongde yingao xiaqing he jiangjie." [Declination and Downstep Effect in Declarative Sentences of Chinese Mandarin]. *Zhongguoyuyinxuebao* 3(2012):54-60.

[17] Xu, Yi, and Q. Emily Wang. "What can tone studies tell us about intonation?." *Intonation: Theory, Models and Applications*. 1997.

[18] Xu, Yi. "Principles of tone research." *Proceedings of International Symposium on tonal aspects of languages*. 2006.

[19] Yuan, Jiahong, Chilin Shih, and Greg P. Kochanski. "Comparison of declarative and interrogative intonation in Chinese." *Speech Prosody 2002, International Conference*. 2002.

[20] Yuan, Jiahong. "Perception of Mandarin intonation." Chinese Spoken Language Processing, 2004 International Symposium on. IEEE, 2004.