On the Role of Intonation in the Organization of Mandarin Chinese Speech Prosody

Chiu-yu Tseng

Institute of Linguistics
Academia Sinica, Taipei, Taiwan
cytling@sinica.edu.tw

Abstract

This paper reports 3 perception experiments on intonation groups and the role of phrasal intonation in the organization of speech prosody. The goal is to help unlimited TTS achieve better naturalness. Experiments were also designed to compliment previous extensive analyses of speech data. Using the PRAAT software and removing segmental information humming experiments of extracted intonation groups ending in interrogative and declarative intonations in both complete and edited forms were used. Results showed that (1.)phrasal or sentential intonation contour is less significant for Mandarin, (2.) yes-no questions with utterance question particles are characterized by a rising pitch on the final syllable only, (3) the general higher register exhibited in yes-no questions without utterance final question particles is not the most salient cue for intonation, (4.) utterance final lengthening appears to be a salient perceptual cue for intonation identification, (5.) speech units larger than single sentences deserve more attention.

1. Introduction

We note that in spoken Mandarin Chinese, instead of complete or complicated sentences, native speakers tend to speak in a sequence of phrases. These larger-than-sentence utterance groups, loosely governed by semantics and are currently under investigation, are perceptually identifiable larger units that bear prosody characteristics. So far such identification, consistent across listeners, can be characterized on the basis of perceived boundaries as functional units of prosody. These results have been analyzed and reported in earlier investigations [1, 2]. The fundamental frequency (F0) patterns and phrasal intonation can also be characterized with respect to their positions in a prosodic group (PG) [3]. Therefore, my current attempt is to see how these phrasal intonations function perceptually in PG final position, whether they are consistently identifiable, and how these intonations interact with the planning of speech prosody units larger than utterances and/or sentences.

Three perception tests were performed to test the following hypotheses, namely, (1.) phrasal intonations exist in Mandarin, but do not play as much a role as they do in non-tonal languages. (2.) Question particles play a more significant role in Mandarin Chinese. (3.) Overall intonation patterns lose their characteristics when the final syllable is removed, thereby also showing the less important role of overall intonations. I will present the three experiments below.

2. Perception Experiments

Three auditory perception experiments were conducted to test the hypothesis.

2.1. Experiment 1

2.1.1. Methodology
10 PG samples of male microphone read speech were chosen from a speech database of 599 read paragraphs collected in sound proof rooms. These PG samples ranged from 8 to 24 characters/syllables (or approximately 1 to 6 secs) in duration. All of the chosen PGs end in yes-no questions without a phrase-final 1-syllable question particle. Among the 10 speech samples, 5 ended in 2-syllable prosodic words; the other 5 in 3-syllable prosodic words. Backward editing of these PGs was performed, removing the last one, last two and last three syllables of the PG respectively. A total of 40 PGs were generated. Using the PRAAT software, the segmental information of these 40 PGs were removed and then replaced by humming whereas the overall F0 patterns were extracted and retained. A total of 40 humming tokens were created to serve as stimuli of Experiment 1. Four repetitions of the tokens were randomized, making up a total of 160 test tokens of the experiments.
2.1.2. Subjects. 4 subjects, 1 male and 3 female, participated in Experiment 1. All of the subjects were college educated native speakers of Mandarin Chinese spoken in Taiwan with no hearing impairment.
2.1.3. Procedures. Identification perception tests were administered in sound proof rooms over headsets. Each subject received different randomization results. Subjects were asked to identify if they heard yes-no question intonation.

2.2. Experiment 2

2.2.1. Methodology
For Experiment 2, 20 PG samples of male microphone read speech from the same speech data base were chosen for Experiment 2. 10 of the PG samples were the same samples from Experiment 1, namely, PGs end in yes-no questions without a phrase-final 1-syllable question particle. Another 10 PGs were samples that ended in declarative phrases. These declarative-ending PG samples ranged from 10 to 24 characters/syllables (or approximately 2.5 to 6 secs) in duration. Among the 10 declarative speech samples, 8 ended in 2-syllable prosodic words; 2 in 3-syllable prosodic words. The same backward editing of these PGs was performed, removing the last one, last two and last three syllables of the PG respectively. A total of 80 PGs were generated. Using the PRAAT software to remove segmental information but retaining overall pitch patterns, a total of 80 humming tokens was created to serve as stimuli of Experiment 2. Four

repetitions of the tokens were randomized, making up a total of 320 test tokens of the experiments.

2.2.2. Subjects. The same 4 subjects participated in Experiment 2 on a different day.

2.2.3. Procedures. The same Identification perception tests were administered in sound proof rooms over headsets. Each subject received different randomization results. Subjects were asked to identify if they heard declarative intonation.

2.3. Experiment 3

2.3.1. Methodology

For Experiment 3, 30 PG samples of male microphone read speech from the same speech data base were chosen for Experiment 3. 10 more PG samples were added to the samples chosen for Experiment 2. That is, in addition to 10 PGs ended in yes-no questions without a phrase-final 1-syllable question particle and 10 PGs ended in declarative phrases, another 10 PGs of yes-no questions with a phrase-final 1-syllable question particle were chosen. These 10 question-ending PG samples ranged from 9 to 23 characters/syllables (or approximately 2.2 to 5.7 secs) in duration. Two of these 10 yes-no questions ended in 2-syllable prosodic words; three in 3-syllable prosodic words. The same backward editing of these PGs was performed, removing the last one, last two and last three syllables of the PG respectively. A total of 120 PGs were generated. Using the PRAAT software to remove segmental information but retaining overall pitch patterns, a total of 120 humming tokens was created to serve as stimuli of Experiment 2. Four repetitions of the tokens were randomized, making up a total of 480 test tokens of the experiments.

2.3.2. Subjects. The same 4 subjects participated in Experiment 3 on a different day.

23.3. Procedures. The same Identification perception tests were administered in sound proof rooms over headsets. Each subject received different randomization results. Subjects were asked to identify if they heard declarative intonation.

2.3. Results. The following tables and figures summarize results of the above 3 perceptual identification experiments. Note that correct identification is defined as follows: For both yes-no questions with and without utterance final question particle, only the complete utterance intonation is defined as question intonation. In other words, all edited tokens were treated as declarative intonation.

Table 1 show the results of Experiment 1, i.e., perceptual identification of humming of yes-no question intonation without question particle. Figure 1 plotted the same results.

Ss	A	B	C	D
S1	44.4%	25.0%	17.1%	35.9%
S2	78.9%	44.7%	14.3%	35.9%
S3	70.3%	35.9%	16.2%	45.9%
S4	70.3%	42.1%	35.1%	63.2%
Avg	65.4%	36.3%	18.8%	47.1%

Table 1: correct identification rates of yes-no questions without question particles.
A: Tokens of full PG
B: Tokens without last syllable
C: Tokens without last 2 syllables
D: Tokens without last 3 syllables

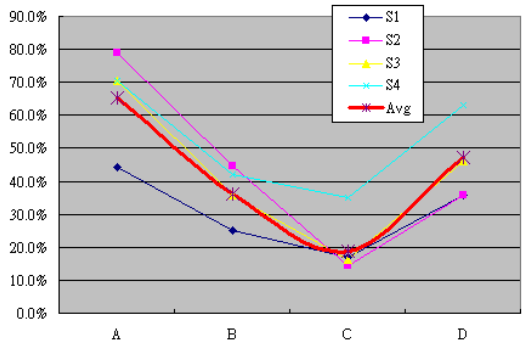

Figure 1: correct identification rates of yes-no questions without question particles.

Table 2 show the results of Experiment 2, i.e., perceptual identification of humming of declarative intonation as well as yes-no question intonation without question particle. Figure 2 plotted the same results.

Ss	A	B	C	D
S1	75.0%	66.3%	60.0%	66.3%
S2	65.0%	61.3%	36.3%	47.5%
S3	60.0%	61.3%	50.0%	60.0%
S4	58.8%	56.3%	46.3%	58.8%
Avg	64.7%	61.3%	48.1%	58.1%

Table 2: correct identification of declarative vs. yes-no questions without question particles

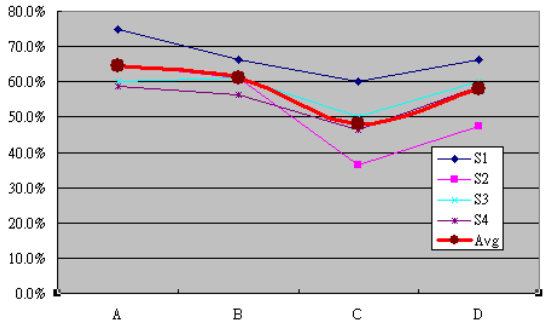

Figure 2: correct identification of declarative vs. yes-no questions without question particles

Table 3 show the results of Experiment 3, i.e., perceptual identification of humming of declarative intonation, yes-no question intonation without question particle, and yes-no question with question particle. Figure 3 plotted the same results.

Ss	A	B	C	D
S1	66.7%	65.8%	55.0%	61.7%
S2	57.5%	43.3%	32.5%	41.7%
S3	71.7%	43.3%	39.2%	42.5%
S4	70.8%	40.0%	38.3%	43.3%
Avg	66.7%	48.1%	41.3%	47.3%

Table 3: correct identification of declarative utterances vs. yes-no questions vs. yes-no questions with interrogative particles

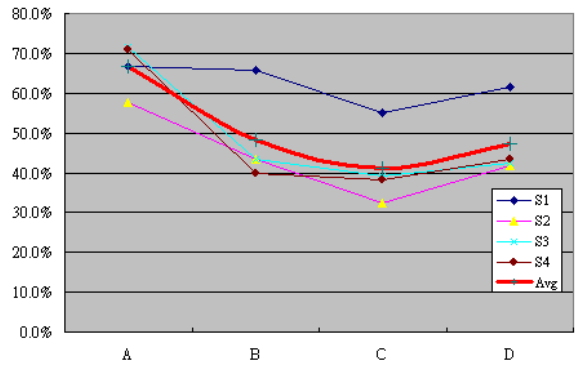

Figure 3: correct identification of declarative utterances vs. yes-no questions vs. yes-no questions with interrogative particles

3. Discussion

The results of the three perception experiments showed that yes-no question in Mandarin Chinese can be carried via two types of intonations, namely, overall higher register for yes-no questions without question particle or the rising of the very last syllable for yes-no questions with question particles at utterance final positions. Note that the kind of rising of F0 pattern is only prominent in the last syllable of a PG that could be 6 seconds in duration. Removal of that one syllable made the PG's ending declarative instead.

Subjects also reported two strategies used for the identification tests. First, a large number of intonations ended abruptly—these were apparently the edited versions--and as a result did not sound like questions. Second, final lengthening was used as an indicator for question intonation. The first strategy suggests that abruptness overrides intonation pattern, and is associated with declarative intonation. The second strategy implies that subjects somehow associated question intonation with a smoother ending, thereby suggesting the significance of utterance final lengthening effect due to the decline of energy.

However, note that identification was worst when two syllables were edited out while subjects' overall performance bounced back when three syllables were edited out. Since the overall numbers of 2-syllable vs. 3-syllable prosodic words were balanced, this may imply that utterance final energy distribution may have some effect on the overall pattern. And the energy distribution is significant for the last 3 syllables of a PG. Further studies on PG final boundaries should shed more light on this.

4. Conclusions

The results confirmed in part of sentence intonation of Mandarin [4, 5, 6] and an earlier study on global question [7]. However, note that these earlier studies used sentences produced in isolation, and thereby focused on sentence intonation pretty much the same way English sentence intonations are investigated in the literature. Much of the attention had been given to how yes-no questions possess overall higher register in declarative sentence. Since no prosody units larger than or higher above sentence level was postulated or discussed, these studies used sentences under 10 syllables almost all the time and implied, perhaps not explicitly, connected speech could be seen as connecting isolated sentences in succession. Researches in TTS have demonstrated that this kind of approach is hardly sufficient for the generation of unlimited text. The present study differs from previous studies in the following sense: (1.) larger units that imply overall planning of speech output must be taken into consideration. (2.) Phrasal intonations of tone languages are not as significant as they are in intonation languages. (3.) The default intonation in Mandarin is the declarative intonation. Questions can be conveyed through linguistic vehicles other than intonation. Humming tests further proved the lesser role of phrasal or sentential intonations in Mandarin Chinese. If this is indeed the case, unlimited TTS definitely requires more knowledge of prosody planning than phrases or sentences in order to achieve naturalness. Concatenating phrases or sentences is simply not adequate. As a result, it is only more obvious that speech prosody should be investigated by including domains and units larger than phrases and sentences.

5. References

[1] C. Tseng, "The prosodic status of breaks in running speech: Examination and evaluation", in *Speech Prosody 2002*, 11-13 April, Aix-en-Provence, France, pp. 667-670, 2002.

[2] C. Tseng and F. Chou, "A prosodic labeling system for Mandarin speech database", *Proceedings of the XIV International Congress of Phonetic Science*, Aug.1-9, 1999, San Francisco, USA, pp2379-2382

[3] C. Tseng, "Towards the organization of Mandarin speech prosody: Units, boundaries and their characteristics", *XIV International Congress of Phonetics Science,* Aug.1-9, 2003, Barcelona, Spain.

[4] Ho, A.-T. Mandarin tones in relation to sentence intonation and grammatical structure", *Journal of Chinese Linguistics,* 4, 1976, pp. 1-13

[5] Shen, J. "Beijinghua shengdiao de yinyu he yudiao (Pitch range of tone and intonation in Beijing dialect, in Chinese)", in Lin T. and Wang L eds. *Beijing Yuyin Shiyanlu* (Working Papers in Experimental Phonetics), Beijing, Beijing University Press, 1985, pp. 73-130

[6] Lin, M-C, "Hanyu yunlyu jiegou han gongneng yudiao (Mandarin prosody organization and functional intonations, in Chinese)", *Report of Phonetic Research* 2002, Phonetics Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences pp. 7-23

[7] Y. Chang, "les indices acoustiques et perceptifs des questions totales en Mandarin parle de Taiwan", Cahiers de Linguistique Asie Orientale, 1998, pp. 51-78