THE ROLE OF PITCH RESET IN PERCEPTION AT DISCOURSE BOUNDARIES

Hsin-Yi Lin & Janice Fon

Graduate Institute of Linguistics, National Taiwan University, Taiwan
d96142001@ntu.edu.tw; jfon@ntu.edu.tw

ABSTRACT

The present study is targeted to investigate how pitch reset as a boundary correlate functions in indicating structural information in discourse for listeners. Two experiments were conducted: one was a boundary detection test, and the other was a boundary hierarchy rating test. Results of both experiments indicated that listeners indeed rely on the presence and the strength of pitch reset in decoding hierarchical segmentation in discourse.

Keywords: pitch reset, boundary correlate, discourse boundary, discourse hierarchy, prosody

1. INTRODUCTION

In human communication, discourse structure plays an important role in providing speakers with a mechanism to arrange information flow. One of the channels for speakers to communicate this structural information across to their listeners is through prosody, and studies concerning this issue have found that several prosodic cues, such as final lengthening, pause, amplitude, and pitch reset, are used by speakers in this regard [3, 4, 8].

Among all the above cues, pitch reset, the resetting of declined pitch to the original or higher level when a new unit starts, attracts much attention, as on the one hand, it displays a close relationship with discourse structure [4, 8], while on the other, its realization is prone to be influenced by many other factors, such as affect (e.g., anger or joy) [2], or sentence type (e.g., declarative or interrogative) [1]. For Mandarin Chinese, such a task is even more complicated, because as a tone language, pitch variations from lexical tones are so densely distributed, which may result in interactive influences [12]. Therefore, there are still a lot remaining to be done regarding this issue.

The close relationship between pitch reset and discourse structure can be viewed from two aspects. The first is its concurrence with discourse boundaries [4, 11]. It has been observed that when two adjacent utterances do not shift in topic, their pitch contours tend to be the same or on the same declining track [4, 5]. Such a melodic continuity can be broken by pitch reset for speakers to mark shift of topic or segmentation in discourse. The second aspect is its ability to reflect hierarchical status of a boundary. Many studies have indicated that the amount of pitch reset is bigger across discourse segments than within [6, 8, 11], suggesting that as boundaries become bigger, the amount of pitch reset is larger too, making it a nice indicator in reflecting boundary hierarchy.

As established in the above studies, speakers use pitch reset, among other cues, to encode segmentation and hierarchical structure in discourse. Therefore, the present study aims itself on investigating whether listeners also use this cue to decode discourse structure in speech. As a matter of fact, in some of previous studies, pitch reset has been found to be used by listeners as a marker of syntactic breaks [10] or of prosodic hierarchy [9]. While these studies demonstrated the disjuncture effect generated by pitch reset, it is still unclear whether and how this cue is used by listeners in the domain of discourse. Swerts et al. [11] examined similar questions by asking subjects to listen to and label boundaries in spoken corpora, and the results showed that listeners were indeed facilitated by prosodic cues, including pitch reset, in recognizing discourse hierarchy. The focus of their study, however, was on searching a neat method to label the structure. As a consequence, the effect of pitch reset they found could be from subjects' judgment after incorporating information from many other sources, including semantics, syntax, and other prosodic parameters. In view of this, the present study employed a more controlled paradigm, ruling out possible confounding factors, in order to have a clearer concept on the role of pitch reset in perception at discourse boundaries.

2. EXPERIMENT 1

2.1. Method

This experiment was to examine whether different degrees of pitch reset affect the detectability of boundaries to listeners.

2.1.1. Subjects

Thirty native speakers of Taiwan Mandarin, aged from 18 to 25 years old, were recruited as subjects. None had ever suffered from language disorders or listening disabilities.

2.1.2. Stimuli

Stimuli were utterances of 18 syllables, which were composed of three nonsense syllables, *bu*, *di*, and *ga* spoken by a native female speaker. The synthesized boundaries were placed on the 7th, 9th, or 11th syllable within an 18-syllable utterance, which divided the utterance into two sub-stretches, Sentence 1 and Sentence 2, as shown in Fig. 1. Two sets of registers were generated: the lower register started with 210 Hz in Sentence 1, and the higher with 220 Hz. Right after Sentence 1, without any pause, followed Sentence 2, and the pitch height of Sentence 2 was either 10, 20, or 30 Hz higher than Sentence 1. Loudness for each syllable was adjusted to be auditorily similar, and duration for each syllable was set at 200 ms. To summarize, there were three levels of length for Sentence 1 (7-, 9-, 11-syllables long), two levels of pitch register (lower and higher), and three levels of pitch reset (10, 20, 30 Hz reset), which added up to 18 conditions. For each condition, five utterances were generated, resulting in 18 × 5 = 90 stimuli in total. Details could be referred to [7].

Figure 1: A schematic example of a stimulus. The syllables in the gray area compose Sentence 1, which is 9-syllable long, 220 Hz in pitch; those in the white area compose Sentence 2, which is 240 Hz in pitch, making a 20 Hz reset between the two sentences.

Syllable	ga	bu	ga	de	bu	di	ga	bu	di	ga	di	ga	bu	di	ga	bu	di	ga
Pitch (Hz)	*220*	*220*	*220*	*220*	*220*	*220*	*220*	*220*	*220*	*240*	*240*	*240*	*240*	*240*	*240*	*240*	*240*	*240*

2.1.3. Equipment

E-prime 2.0 was used to present stimuli with a randomized order, a Serial Response Box was used to record reaction time, and a pair of Sony MDR7502 earphones was used for subjects to listen to stimuli.

2.1.4. Procedure

Subjects were seated in front of a computer in a quiet room wearing earphones. They were told that in each trial, there were two adjacent meaningless sentences presented after a reminder bell, and they needed to press the button on the reaction box as soon as they heard Sentence 2 started. Reaction time (RT) was taken from the starting point of Sentence 2 to the time point subjects reacted. They were reminded to concentrate on the task by watching a small fixation point on the screen. A practice session of five trials was provided before the real experiment began. The whole experiment lasted about 20 minutes, with one short break inserted every five minutes.

2.2. Results of experiment 1

A three-way Register (2) × Reset (3) × Length (3) repeated measures design was carried out on RT. Results showed that all factors had main effects. For Register, as shown in Fig. 2, subjects' boundary detection was significantly faster in the higher register than in the lower one [$F(1, 46) = 4.857$, $p < .05$, $\hat{\eta}^2 = .10$]. As for the factor of Reset, shown in Fig. 3, subjects detected boundaries faster as the amount of reset was bigger [$F(2, 92) = 29.311$, $p < .01$, $\hat{\eta}^2 = .39$], and post-hoc analyses with Bonferroni adjustments were made, showing that all paired comparisons were significant ($p < .01$), but the difference between 20 Hz and 30 Hz was not as strong ($p < .05$). Length also had a main effect [$F(2, 92) = 49.463$, $p < .01$, $\hat{\eta}^2 = .52$], and the post-hoc analyses showed that subjects had shorter reaction time when sentences were longer ($p < .01$), as shown in Fig. 4.

Figure 2: Subjects' mean RT at different levels of register.

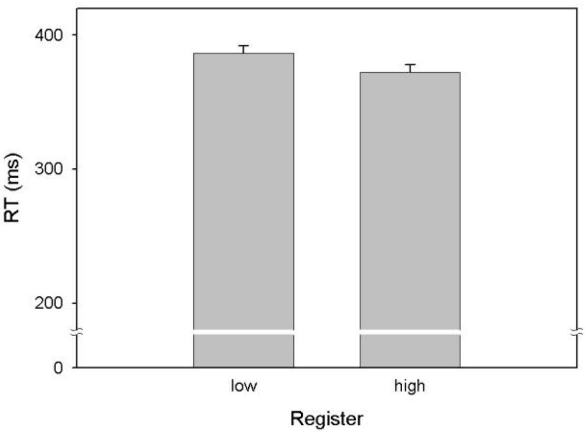


Figure 3: Subjects' mean RT at different levels of pitch reset.

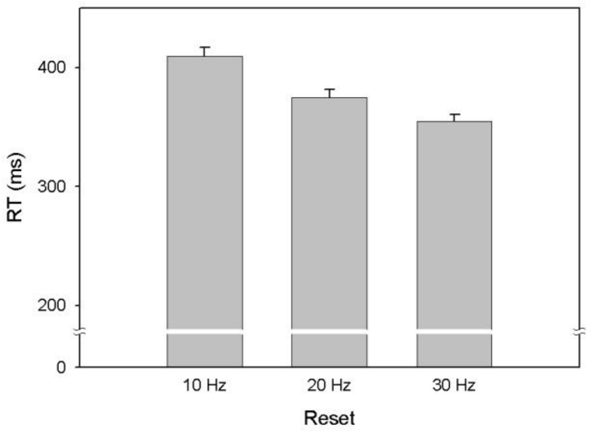
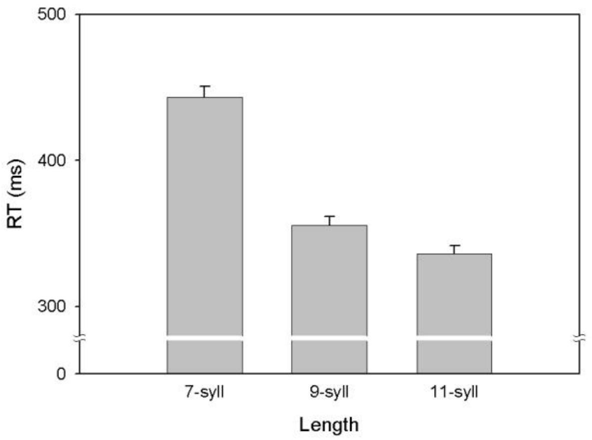


Figure 4: Subjects' mean RT in different sentence lengths.



3. EXPERIMENT 2

3.1. Method

This experiment was to examine whether different degrees of pitch reset affect subjects' judgment on boundary hierarchy.

3.1.1. Subjects

Fifteen subjects were recruited according to the same criteria as in Experiment 1.

3.1.2. Stimuli

Stimuli were utterances of 12 syllables, which were composed of the same three non-sense syllables as those in Experiment 1. A boundary was set after the 7th syllable for each stimulus, dividing each string into Sentence 1 and Sentence 2. As in Experiment 1, two sets of register were generated, the lower one starts from 210 Hz in Sentence 1, and the higher from 220 Hz. Pitch of Sentence 2 was either 0 Hz, 15 Hz, 30 Hz, or 45 Hz higher than that of Sentence 1, resulting in four levels of reset. There were Register (2) × Reset (4) = 8 conditions, for each of which, six utterances were generated, totaling 8 × 6 = 48 stimuli.

3.1.3. Equipment

Equipment was the same as in Experiment 1, except that subjects responded with a keyboard instead of a response box.

3.1.4. Procedure

Subjects were told that for each trial they were to hear two meaningless synthesized sentences, and meanwhile, the syllable content of the sentences would be presented on the screen with an arrow pointing to the space in between the 7th and the 8th syllables, indicating the place of boundary. Right after the audio stimuli and the visual aid, a question slide would be prompted asking them how big they thought the boundary was, and they answered by pressing a corresponding key, where "0" represented no boundary, "1" small boundary, "2" a medium boundary, "3" a medium-to-big boundary, and "4" a very big boundary. Afterwards, another slide was shown, asking how confident they were with their answer for the last question, with 0 being the lowest, and 4 the highest. Careful accounts on boundary hierarchy and two practice trials were given before the experiment started. The whole procedure took about 15 minutes, with a short break every 5 minutes.

3.2. Results of experiment 2

Two two-way Register (2) × Reset (4) repeated measures designs were carried out on subjects' boundary rating and confidence rating, respectively. For boundary rating, results showed that only Reset had a main effect [$F(3, 255) = 22.73$, $p < .01$, $\hat{\eta}^2 = .21$]. Post-hoc tests with Bonferroni adjustments showed that all pairwise comparisons were significant, except between 30 Hz and 45 Hz. As shown in Fig. 5, subjects' rating on boundary sizes became bigger as reset was larger. As to confidence rating, Reset was also the only one to have a main effect [$F(3, 261) = 4.83$, $p < .01$, $\hat{\eta}^2 = .05$], and the post-hoc tests showed that significant difference existed between 0 Hz vs. 15 Hz, and 0 Hz vs. 45 Hz . As shown in Fig. 6, in general, subjects' confidence decreases as the amount of reset increases.

Figure 5: Subjects' boundary rating at different levels of reset.

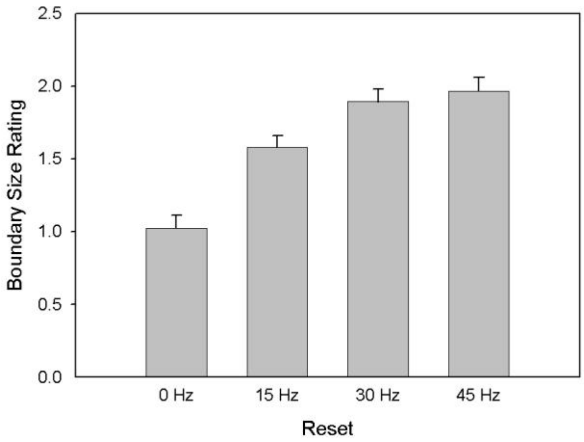
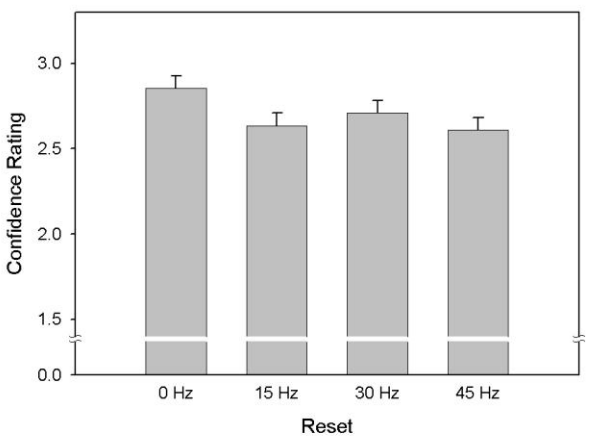


Figure 6: Subjects' confidence rating at different levels of reset.



4. DISCUSSION

This study has revealed a clear effect of pitch reset on listeners' perception at discourse boundaries: in Experiment 1, when a pitch reset is presented within a string, listeners recognize it as a boundary, and their speed of detection increases with the increase of amount of reset, suggesting that bigger pitch reset seems to activate a representation of a bigger discourse boundary. This point is even more confirmed in Experiment 2, showing that boundaries with bigger reset are indeed rated as boundaries of higher hierarchy. As for pitch register, while it had an effect on boundary detection, it did not in boundary size judgment. This implied that listeners used this cue for boundary identification, not for judging boundary hierarchy. The effect of sentence length may be caused by the fact that subjects know the possibility of boundary occurrence increases as a sentence proceeds. Such an expectation may lower the thresholds of information, eliciting shorter RTs.

One interesting phenomenon is that listeners appeared to have more difficulties in differentiating among boundaries of bigger reset. As in Experiment 1, RTs for bigger resets were not as different as for smaller ones. Also, in Experiment 2, boundaries with bigger resets did not show significant differences in their ratings, and listeners' confidence decreases as reset is bigger. Such a lack of fine mapping between reset and big boundaries may be due to the fact that big boundaries are actually more often signaled by long silent pauses, instead of pitch reset [12], which makes it harder for listeners to associate this cue with big boundaries. Further studies are needed about this interaction between temporal and pitch parameters in this regard.

5. REFERENCES

[1] Chuang, Y.Y., Huang, Y.H., Fon, J. 2007. The effect of incredulity and particle on the intonation of yes/no questions in Taiwan mandarin. *Proc. 16th ICPhS* Saarbrücken, 1261-1264.

[2] Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., Maneewongvatana, S. 2006. The roles of pitch contour in differentiating anger and joy in speech. *International Journal of Signal Processing* 3, 129-134.

[3] Fon, J., Johnson, K., Chen, S. (to appear). Durational patterning at discourse boundaries in Mandarin spontaneous speech. *Language and Speech.*

[4] Hirschberg, J. Nakatani, C. 1996. A prosodic analysis of discourse segments in direction-giving monologues. *Proc. 34th Annual Meeting of the Association for Computational Linguistics* Santa Cruz, CA., 286-293.

[5] Kutik, E.J., Cooper, W.E., Boyce, S. 1983. Declination of fundamental frequency in speakers' production of parenthetical and main clauses. *J. Acoust. Soc. Am.* 73, 1731-1738.

[6] Ladd, D.R. 1988. Declination reset and the hierarchical organization of utterances. *J. Acoust. Soc. Am.* 84, 530-544.

[7] Lin, H.Y., Fon, J. 2010. Perception on pitch reset at discourse boundaries. *Proc. 11th InterSpeech* Makuhari, Japan, 1225-1228.

[8] Oliveira, M. 2003. Pitch reset as a cue for narrative segmentation. *Proc. Prosodic Interfaces 2003* Nantes, France, 73-78.

[9] de Pijper, J.R., Sanderman, A.A. 1994. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *J. Acoust. Soc. Am.* 96, 2037-2047.

[10] Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C. 1991. The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* 90, 2956-2970.

[11] Swerts, M., Geluykens, R., Terken, J. 1992. Prosodic correlates of discourse units in spontaneous speech. *Proc. ICSLP-1992* Alberta, Canada, 421-424.

[12] Yang, Y., Wang, B. 2002. Acoustic correlates of hierarchical prosodic boundary in Mandarin. *Proc. Speech Prosody 2002* Aix-en-Provence, France, 707-710.