COMBINING COMMAND-RESPOND MODEL AND ToBI IN ANALYZING SECOND LANGUAGE ENGLISH SENTENCE PROSODY

Jinghua Zhang[a] *& Yinghao Li*[b]

[a]Department of English Language, Yanbian University, Yanji, China;
[b]Department of Chinese Language and Literature, Peking University, Beijing, China
janggyunghwa@yahoo.com; leeyoungho@pku.edu.cn

ABSTRACT

The paper combines the Fujisaki command-response model and the ToBI convention to analyze the English sentence prosody produced by American native speakers and Korean and Chinese English learners. The model parameters were extracted by *FujiParaEditor* for nine short sentences produced by three groups of speakers. The ToBI-tanscribed results were compared with the model parameters. Results show that model-based approach is a valid complement for judging the pitch and break events in both native and nonnative speech. With the aid of model-generated parameters it is found that both groups of English learners produce shorter prosodic phrase in English sentences but with different prosodic strategies, which is indicative of native language transfer. The Korean subjects invariably produce H-phrasal accent but the Chinese subjects use L-accent in the sentence-medial position. The Chinese subjects accent all content words and sentence-final function words. But the F0 contour produced by Korean subjects shows a narrow pitch range. Both groups have difficulty in acquiring the canonical yes-no question contour (L* H-H%). The majority use L-H% tone for a yes-no question.

Keywords: English sentence prosody, command-respond model, ToBI, English learner

1. INTRODUCTION

Second language prosody remains the least ploughed area for the complexity of the F0 contour [2]. The previous studies found that the intonational errors made by English learners ranges from the local pitch event to the overall pitch contour movement, represented by a narrower pitch range, smaller declination rate, placement of prominence, incorrect pitch on unstressed syllables, and improper prosodic parsing strategies.

The previous studies relied heavily on perceptual judgment or direct comparison of acoustic cues. Preconception might be resulted from the first method while the second neglects the logarithmic nature of human hearing faculty. The command-response model is possible to tackle the two problems [5]. The advantage of the model is that the pitch events and phrasing can be generated and is shown to have close correlation with the pitch events and break index in the ToBI system [10]. To date, the model has been applied in the prosodic analysis for Korean, English and the Standard Chinese [3, 4, 6]. It has been also applied in analyzing second language prosody [7, 9].

The paper first analyzes the model parameters of the English declaratives and yes-no questions produced by American native speakers and Korean and Chinese English learners, aiming at comparing the results of the model-based approach and that in the perceptually-based approach. Second, the command-response model and ToBI convention are combined to analyze the sentence prosody of Korean and Chinese English learners.

Table 1: Sentence corpus.

No.	sentence
1	Did you have a nice trip?
2	Computers are fun.
3	Will you pay by credit card?
4	Did you break your leg?
5	It's going to rain.
6	What can I do for you?
7	Have you ever visited the White House?
8	I will work with them.
9	I got back from a long vacation.

2. METHOD

2.1. Speech corpus

The speech corpus was comprised of nine short sentences, encompassing 4 simple statements, 1 wh-question and 4 yes-no questions (Table 1). The number of syllables in each sentences ranges from

5 to 10. The maximum syllable number for word in the sentences was three with the stress either on the first and second syllable.

2.2. Speakers and recording

Five Korean and four Chinese female under-graduate English majors and two female American English teachers were recruited in the experiment. The average age for two groups of English learners (NNS) was 23.7. All subjects had received the English pronunciation training for a semester one or two years before the experiment. The two native speakers (NS) came from the Midwestern states of the U.S. with the average age of 26.8.

The recording was made in a sound-treated phonetics lab. The subjects were advised to read at their normal speed with normal voice volume after familiarizing the material.

2.3. Measurements

The pitch contours were extracted by PRAAT. The vocal fry and the spurious movement caused by voiceless consonants were deleted. The adapted pitch contours was first labeled following ToBI convention [1]; then the smoothed F0 contour was submitted to the *FujiParaEditor* [10]. Because the Fb was speaker-individual constant, the free-run Fb of the nine sentences read by one subject was averaged for further use of analyzing phrase command and accent command. For the phrase command, the initiation of a new phrase command corresponded to the major syntactic boundaries and indicated by a phrase resetting. The phrase command that was indicative of the re-phrasing was adjusted when it was missed in the analysis. The last step was the setting of the accent command. The local pitch hump was required to align with the stressed syllable or word. The alpha, beta and gamma were set at 3, 20, and 0.9 respectively. The pitch range was also measured for each utterance and then averaged across the nine sentences for each subjects.

3. RESULTS

3.1. Model parameters

As shown in Figure 1, the NS invariably produced one phrase command (PC) in all sentences. The NNS produced one PC when the sentences were short (No.1, 2, 5, 6), but more PCs when the sentences were longer (No. 3, 7, 9).

A one-way ANOVA result shows that the number of PC is significantly different across the three groups ($p<0.01$); the result of the Turkey's Post-Hoc test shows that the number of PC produced by NS is significant less than that by Korean ($p<0.05$) and Chinese ($p<0.01$), but no difference is found between the latter two groups.

Figure 1: Means of PC number for nine sentences across three groups. (NS is American natives; KL is Korean learners, and CL the Chinese learners)

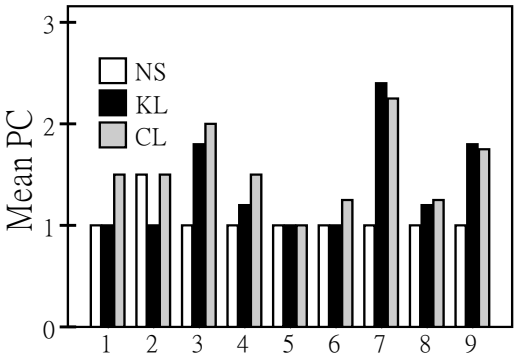
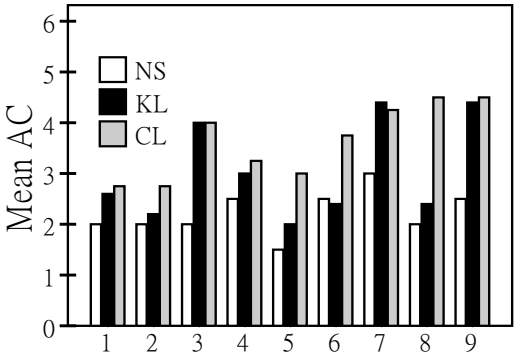


Figure 2 shows the means of accent command (AC) numbers produced by three groups. In general both the NNS groups produced more ACs than NS groups, and Chinese group produced a little more ACs than Korean group.

Figure 2: Means of AC numbers for nine sentences across three groups.



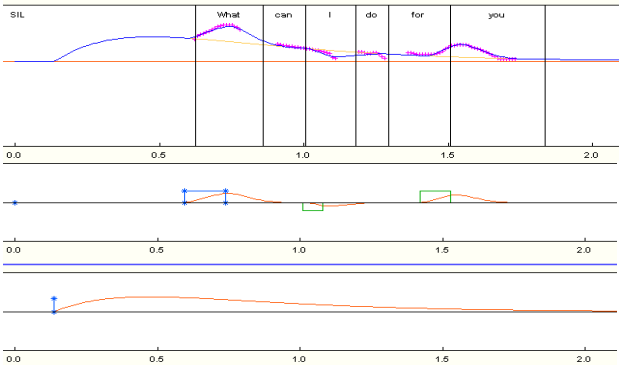
A one-way ANOVA result shows that the number of AC produced by three groups is significantly different ($p<0.001$). A series of sentence-wise one-way ANOVA results show that no significant difference is found across the three groups in shorter sentences (No.1, 2, 4, 6). But in longer sentences (No.3, 5, 7, 8, 9) significant difference is found. In No.3, 7 and 9, the NS produced significantly less AC than the NNS. In No.5, AC number falls in the order of NS<Korean<Chinese ($p<0.01$). In No.8, Chinese learners produced significantly more AC than the other two groups.

The free-run Fb differs significant between NS and NNS ($p<.05$), but it remains ambiguous in linguistic sense. The difference in pitch range tends to convey more linguistic information: the pitch range for Korean learners is significantly lower than that of NS and marginally from that of the Chinese group.

American English and Korean are claimed to have no negative accent commands, but Chinese does have negative accent commands. In [7] it was found that the negative accent command appeared in the Japanese sentences produced by Chinese learners. In this experiment negative commands appear at sentence-medial positions in sentences produced by three Chinese learners. An example is shown in Figure 3.

In short, the model offers quantitative measures for improper rephrasing and accent placement strategies for NNS speakers compared with NS speakers. The free-run Fb and pitch range differences may correspond to the perceptual judgment that the NNS is incapable of skillful voicing manipulation in producing target language prosody. The language-specific interference in producing second language prosody may also be quantitatively analyzed.

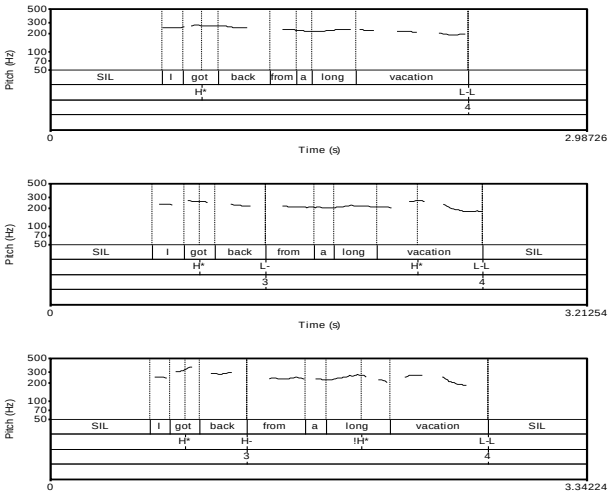
Figure 3: The sentence pitch contour (Top of the diagram) and the parameters (AC in the middle and PC in the bottom) generated by *FujiParaEditor*. The AC value for "I" is negative, indicating a fast fall of pitch contour.



3.2. Model-generated parameters and the ToBI transcribed results

The intonational pattern of a simple statement is characterized by (H*) H* L-L% contour [1]. In Figure 4, NNS produced more pitch accents than NS and with larger AC amplitude (Table 2). For NNS an additional break index was transcribed in the sentence-medial position, corresponding to the initiation of new PC. What's more, Korean subjects produced H- and Chinese L-if sentence-medial breaks do occur, which is indicative of different prosodic strategies by two groups of learners.

Figure 4: ToBI transcription for a simple statement produced by a native (top), a Chinese (middle), and a Korean subject (bottom).



Difference is also found in pitch accent placement for three groups. For example, in No.2 sentence, the Chinese subjects accented all syllables in '*com-put-ers*' whereas the NS only accented the middle syllable. The Korean subjects tend to initiates a rather low AC at the beginning of the word and then the F0 contour declines slowly after the accented syllable. This result shows that both NNS groups have difficulty in producing the most prominent syllable in a multi-syllable word.

Table 2: AC amplitude for each accented syllables of three groups.

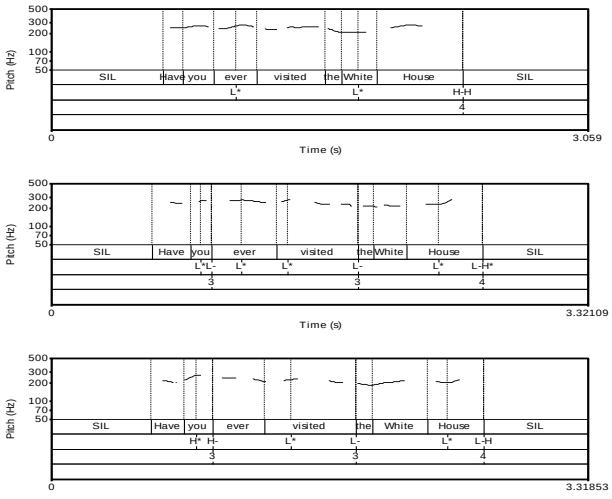
	I	got	back	from	a	long	vacation
NS	0	.19	0	0	0	.10	.27
C	0	.29	.19	.10	0	.17	.18
K	0	.34	.19	.07	0	.27	.20

It is found that the Chinese learners consistently place accents on the sentence-final function words. For example, "*you*" in No.6, and "*them*" in No.8 sentence, most of them were labeled with H* or !H*. However, the Korean subjects rarely placed accent on the sentence-final function words.

The canonical yes-no question contour is (L*) L* H-H%. The F0 contour starts to rise on or after the last nuclear pitch accents (L*) [1]. In most cases, the F0 contour of NS maintains flat before the final L* followed by H-H%. However, NNS tend to expand the pitch range of the accented words to achieve the non-final low pitch accent with large AC amplitude. Three out of four

Chinese subjects and three out of the five Korean subjects placed L* on the last word followed by L-H% to end a yes-no question, as shown in Figure 5.

Figure 5: ToBI transcription for a yes-no question produced by a native (top), a Chinese (middle), and a Korean subject (bottom).



4. CONCLUSION

The present paper combines Fujisaki command-response model and ToBI to analyze the English sentence prosody produced by native speaker and non-native English learners. The model-based approach is proved to offer quantitative measures which are likely to help with the perceptual judgment and quantitatively analysis for pitch and break events in complex sentence prosody in second language learners.

When the model-based quantitative approach and the ToBI-based qualitative approach are combined in analyzing the English sentence prosody, it is found that both Korean and Chinese learners have difficulty in realizing the prominent syllable. The Chinese learners accent all content words and sentence-final function words. In the case of multi-syllabic word that happens to be nuclear pitch accent, they are likely to raise the key of the word as a whole. The narrower pitch range produced by Korean subjects makes it hard to determine the accent events. Secondly, both NNS groups have problems in prosodic parsing by producing shorter intonational phrases. The universal constraints for L2 speech planning skills development may have a play [11]. However, the sentence-medial phrasal accent or boundary tone is indicative of native language transfer. Regarding the F0 contour of yes-no question for both NNS groups, it is characterized by L* L-H% pattern, and the pitch accent and the boundary tone are both realized upon the last content word, which is supported by the later initiation of accent command for NNS instead of NS. This tends to be the influence of the native language, for in both languages, the rising boundary tone is placed on the last morpheme of the sentence.

5. ACKNOWLEDGEMENTS

This research was supported by Yanbian University Scientific Research Start-up Fund (Grant YJ-4-2008049).We thank the anonymous reviewer for the comments on the earlier version of the paper.

6. REFERENCES

[1] Beckman, M.E., Elam, G.A. 1997. Guidelines for ToBI labelling (ver.3.0). *http://www.ling.ohio-state.edu/~tobi/*

[2] Chun, D.M. 2002. *Discourse Intonation in L2: From Theory and Research to Practice*. John Benjamins Publishing Company.

[3] Fujisaki, H. 1996. Analysis and modelling of fundamental frequency contours of Korean utterances – A preliminary study. *Phonetics and Linguistics – In Honor of Prof. H.B. Lee* Seoul, 640-657.

[4] Fujisaki, H., 2004. Information, prosody, and modelling-with emphasis on the tonal features of speech. In Fant, G., et al. (eds.), *From Traditional Phonology to Modern Speech Processing*. Beijing: Foreign Language Teaching and Research Press, 111-128.

[5] Fujisaki, H.; Hirose, K., 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan* 5, 233-242.

[6] Fujisaki, H., Wang, C.F., Ohno, S., Gu, W.T. 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command-response model. *Speech Communication* 47, 59-70.

[7] Hirano, H., Gu, W.T., Hirose, K. 2006. Model-based analysis of F0 contours of Japanese sentences uttered by Chinese speakers. *Proc. 7th PCC* Beijing.

[8] McGory, J.T. 1997. *Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers*. Ph.D. dissertation, Ohio State University.

[9] Mixdorff, H. 1996. Foreign accent in intonation patterns – A contrastive study applying a quantitative model of the F0 contour. *Proc. of ICSLP* Philadelphia, 1469-1472.

[10] Mixdorff, H. 2000. A novel approach to the fully automatic extraction of Fujisaki Model parameters. *Proc. of ICAAAP*, Vol. 3, 1281-1284.

[11] Ueyama, M., Jun, S.A. 1998. Focus realization in Japanese English and Korean English intonation. *Japanese/Korean Linguistics* 7, 629-645.